Estimation of latent variables in high-dimensional factor models with weak residual dependence

Xinyao Fan and Harry Joe

Abstract For D-dimensional Gaussian 1-factor, bi-factor, oblique factor models and their copula-based counterparts, proxies have previously been defined to estimate the latent variables for large D. It is shown that the previously-defined proxies are asymptotically consistent and hence robust to some interpretable assumptions of weak residual conditional dependence of observed variables given the latent variables. Proofs make use of techniques of likelihoods with mis-specified models. Some simulation results provide concrete behavior of the effect of weak residual dependence and increasing D.

1 Introduction

Factor models are useful tools for modeling the dependence of large sets of variables when the dependence between the observed variables can be explained by several latent variables. When the variables can be divided into groups, structured factor Gaussian and copula models in [9] and [10] are often good first-order models to explain the dependence of variables. The models assume the conditional independence of observed variables given latent variables. As the dimension increases (the number of observed variables linked to each latent variable increases), it may be reasonable to assume a slight departure from the conditional independence assumption, such as weak conditional dependence given the latent variables. In this article, we refer to this as weak residual dependence.

Xinyao Fan
University of British Columbia, Vancouver, Canada
e-mail: xinyao.fan@stat.ubc.ca

Harry Joe
University of British Columbia, Vancouver, Canada
e-mail: Harry.Joe@ubc.ca

For multivariate dependence modeling, extending Gaussian classical factor models to have (a) flexible non-Gaussian tail behavior via factor copulas and (b) weak conditional dependence given latent variables is important for high-dimensional applications.

For a large dimension D for the total number of variables, proxy variables consist of estimates for latent variables based on conditional expectations given the observed variables, for Gaussian factor models or their copula counterparts. For Gaussian factor models, the proxy variables are the factor scores as given in [8]. Consistency results for these proxy variables are proved in [4], under some non-stringent conditions.

The main purpose of this article is to show that proxy variables, defined based on the conditional independence assumption, are robust to weak residual dependence. Detailed proofs using non-standard asymptotic methods are given for 1-factor and bi-factor Gaussian models, and 1-factor copula models, based on interpretable sufficient conditions for weak residual dependence. By interpreting the sufficient conditions, we can intuit sufficient weak residual dependence conditions that might hold more generally. The adequacy of the proxy estimation method is further demonstrated for 1-factor and bi-factor copula models through simulation studies.

Section 2 provides the definitions and notation for Gaussian and copula-based factor models with residual dependence. Section 3 summarizes the conditional expectation form of proxy variables for the Gaussian and copula models. Section 4 provides some interpretable sufficient conditions on weak residual dependence for the consistency of proxy variables for the case of a known factor model. Section 5 has sufficient conditions on residual dependence for the consistency of the proxies in 1-factor models when parameters are estimated. Section 6 presents a sequential estimation method for practical use when linking copula families are to be estimated from data. Section 7 includes simulation studies to provide additional support for the usefulness of the proxy variables in factor copula models when there is weak residual dependence. Section 8 has the conclusions and explains how results in this article can be applied. The proofs of the main results are mostly lengthy, because of the need to keep track of asymptotic order of many terms, and they are given in Section 9.

Abbreviations that are used include cdf for cumulative distribution function and MLE for maximum likelihood estimator or estimation. Copulas which are cumulative distribution functions with $U(0,1)$ univariate margins are denoted with C and subscripts. The copulas are assumed to be absolutely continuous, and the corresponding densities are denoted with c.

2 Factor models with residual dependence

This section defines the Gaussian and copula factor models with their notation to prepare for the analysis of estimation of latent variables in the subsequent sections. Factor models have parsimonious dependence and there is more chance to depart from the dependence structure as the number of variables increases. As discussed in [7], adding conditional dependence given latent variables may possibly lead to a more interpretable model than adding additional latent variables.

2.1 Gaussian factor models

When there are many variables that are related, the factor models assuming conditional independence might be good first approximations but it is possible that the conditional (residual) dependence exists after accounting for latent variables. In this section, we assume that the observed variables in the Gaussian factor models have been standardized to be standard normal $\mathcal{N}(0,1)$.

Let $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$ be a p-vector of independent latent variables and let $\boldsymbol{Z}_D = (Z_j)$ be a D-vector of observed variables that are each $\mathcal{N}(0,1)$. Let $\boldsymbol{\varepsilon}_D \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_D)$ be a random vector of idiosyncratic noise or specific factors that are independent of $\boldsymbol{W}$, where $\boldsymbol{\Omega}_D$ is a correlation matrix.

The stochastic representation for the Gaussian p-factor model with residual dependence is:

$$\boldsymbol{Z}_D = \boldsymbol{A}_D \boldsymbol{W} + \boldsymbol{\Psi}_D \boldsymbol{\varepsilon}_D \tag{1}$$

where $\boldsymbol{A}_D = (\alpha_{jk})$ is a $D \times p$ loading matrix, $\boldsymbol{\Psi}_D = \mathrm{diag}(\psi_j : j = 1, \ldots, D)$ and the loading parameters $\{\alpha_{jk}\}$ are such that $\psi_j^2 = 1 - \sum_{k=1}^{p} \alpha_{jk}^2 > 0$ for all j. The correlation matrix of $\boldsymbol{Z}_D$ is $\mathrm{Cor}(\boldsymbol{Z}_D) = \boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Gamma}_D$, where $\boldsymbol{\Gamma}_D = \boldsymbol{\Psi}_D \boldsymbol{\Omega}_D \boldsymbol{\Psi}_D$. Note that $\boldsymbol{\Gamma}_D = \mathrm{Cov}(\boldsymbol{Z}_D | \boldsymbol{W})$ is the covariance matrix of $\boldsymbol{Z}$ given $\boldsymbol{W}$. If $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$, then the components Z_j of $\boldsymbol{Z}$ are conditionally independent given $\boldsymbol{W}$ (for the usual p-factor model).

In the remainder of this article, because of non-identifiability to rotation of the loading matrix for $p \geq 2$, we focus on the 1-factor model with residual dependence if the variables are not grouped. If the 1-factor model is not an adequate fit for data, the 1-factor model with residual dependence is an alternative ([11]) to adding additional factors and it might be more interpretable in some applications. For high-dimensional $\boldsymbol{Z}_D$, estimation of latent variables is studied in [4] for the 1-factor model but not the p-factor for $p \geq 2$.

For variables that have been divided into G non-overlapping groups, we will consider structured bi-factor models and then the structured loading matrix (with many 0s in it) is based on $p = G + 1$ latent variables with one global latent variable and G group-based latent variables. We will also consider the oblique factor model for G non-overlapping groups where the structured loading matrix $\boldsymbol{A}$ is $D \times G$ and $\boldsymbol{W}$ consists of dependent latent variables.

Definition 1. (1-factor Gaussian with residual dependence). The stochastic representation, with $\boldsymbol{A}_D = \boldsymbol{\alpha}_D = (\alpha_1, \ldots, \alpha_D)^\top$, is

$$Z_j = \alpha_j W + \psi_j \varepsilon_j, \quad j = 1, \ldots, D, \tag{2}$$

where $W \sim \mathcal{N}(0,1)$, $\boldsymbol{\varepsilon}_D = (\varepsilon_1, \dots, \varepsilon_D)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_D)$ is independent of W, and $\alpha_j \in (-1,1)$ and $\psi_j^2 = 1 - \alpha_j^2$ for all j. Note that $\mathrm{Cov}(\mathbf{Z}_D|W) = \boldsymbol{\Gamma}_D = \boldsymbol{\Psi}_D \boldsymbol{\Omega}_D \boldsymbol{\Psi}_D$ with $\boldsymbol{\Psi}_D = \mathrm{diag}(\psi_1, \dots, \psi_D)$.

Definition 2. (Bi-factor Gaussian with residual dependence).
With G groups, let $D = \sum_{g=1}^G d_g$, with d_g variables from group g. Write the loading matrix as $\boldsymbol{A_D} = [\boldsymbol{a}_{0,D}, \mathrm{diag}(\boldsymbol{b}_{d_1 1}, \dots, \boldsymbol{b}_{d_G G})]$ where $\boldsymbol{a}_{0,D} = (\boldsymbol{b}_{0,d_1 1}^\top, \dots, \boldsymbol{b}_{0,d_G G}^\top)^\top$ and for $g \in \{1, \dots, G\}$, $\boldsymbol{b}_{d_g g} = (\alpha_{1g}, \dots, \alpha_{d_g g})^\top$ and $\boldsymbol{b}_{0,d_g g} = (\alpha_{0,1g}, \dots, \alpha_{0,d_g g})^\top$. The stochastic representation is

$$Z_{jg} = \alpha_{0,jg} W_0 + \alpha_{jg} W_g + \psi_{jg} \varepsilon_{jg}, \quad j \in \{1, \dots, d_g\}, \quad g \in \{1, \dots, G\}, \tag{3}$$

$\boldsymbol{\varepsilon}_g^\top = (\varepsilon_{1g}, \dots, \varepsilon_{jg}, \dots, \varepsilon_{d_g g})$, $\boldsymbol{\varepsilon}_D^\top = (\boldsymbol{\varepsilon}_1^\top, \dots, \boldsymbol{\varepsilon}_G^\top)$ with $\boldsymbol{\varepsilon}_D \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_D)$, W_0 and $\{W_g\}$ are mutually independent latent $\mathcal{N}(0,1)$ random variables independent of $\boldsymbol{\varepsilon}$, and $\alpha_{0,jg}, \alpha_{jg} \in (-1,1)$ such that $\psi_{jg}^2 = 1 - \alpha_{0,jg}^2 - \alpha_{jg}^2 < 1$ for all (j,g).

Definition 3. (Gaussian oblique factor with residual dependence). With G groups, let $D = \sum_{g=1}^G d_g$, with d_g variables from group g. The stochastic representation is

$$Z_{jg} = \alpha_{jg} W_g + \psi_{jg} \varepsilon_{jg}, \quad j \in \{1, \dots, d_g\}, \quad g \in \{1, \dots, G\}, \tag{4}$$

where $\boldsymbol{\varepsilon}_g^\top = (\varepsilon_{1g}, \dots, \varepsilon_{jg}, \dots, \varepsilon_{d_g g})$, $\boldsymbol{\varepsilon}_D^\top = (\boldsymbol{\varepsilon}_1^\top, \dots, \boldsymbol{\varepsilon}_G^\top)$ with $\boldsymbol{\varepsilon}_D \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_D)$, $\boldsymbol{W} = (W_1, \dots, W_G)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_W)$ independent of $\boldsymbol{\varepsilon}$, and $\alpha_{jg} \in (-1,1)$ such that $\psi_{jg}^2 = 1 - \alpha_{jg}^2$ for all (j,g). Here, $\boldsymbol{\Sigma}_W$ is a correlation matrix.

2.2 *Factor copula models*

In this section, the copula counterparts of the Gaussian factor models are given. Loading parameters for the main factor are replaced with bivariate linking copulas of the observed and corresponding latent variable, and loading parameters of other latent variables (e.g., p-factor and bi-factor) are transformed to partial correlations given previous latent variables and then replaced with bivariate linking copulas to represent conditional dependence.

With copulas, we assume that observed and latent variables have been transformed to be uniform random variables in the interval $(0,1)$. Generalizing the Gaussian factor models, the copula densities of (structured) factor copula models with weak residual dependence are summarized below based on [11].

Definition 4. (1-factor copula model with residual dependence). Let $\boldsymbol{U}_D = (U_1, U_2, \dots, U_d)$ be a vector of dependent $U(0,1)$ random variables and let $V \sim U(0,1)$ be the latent variable. For $j \in \{1, \dots, D\}$, let C_{j0} be the absolutely continuous bivariate copula cdf for (U_j, V) with density c_{j0} and conditional cdf $C_{j|0}(u|v) = \Pr(U_j \le u|V = v) = \partial C_{j0}(u,v)/\partial v$. Assume the conditional distribution of $\boldsymbol{U}_D$ given V has copula density c_{res}. The joint density of $\boldsymbol{U}_D$ is:

$$c_{\boldsymbol{U}_D}(\boldsymbol{u}_D) = \int_0^1 \underbrace{c_{\text{res}}\big(C_{1|0}(u_1|v),\ldots,C_{D|0}(u_D|v)\big)}_{\text{part 1}} \times \underbrace{\Big\{\prod_j c_{j0}(u_j,v)\Big\}}_{\text{part 2}} \,\mathrm{d}v. \quad (5)$$

Part 1 involves the conditional dependence of $[U_1|V],\ldots,[U_D|V]$. If $c_{\text{res}} \equiv 1$ on $(0,1)^D$ (independence copula) to represent conditional independence given V, then the integrand only involves part 2, and $c_{\boldsymbol{U}_D}$ becomes the density for the 1-factor copula model.

Definition 5. (Bi-factor copula model with residual dependence). With G groups, let $D = \sum_{g=1}^{G} d_g$, with d_g variables from group g. Let $\boldsymbol{D} = (d_1,\ldots,d_G)$. Let $\boldsymbol{V} = (V_0,V_1,\ldots,V_G)$ be the vector of mutually independent latent variables. Let $\boldsymbol{U}_g^\top = (U_{1g},\ldots,U_{d_g g})$, $g \in \{1,\ldots,G\}$, and $\boldsymbol{U}_{\boldsymbol{D}}^\top = (\boldsymbol{U}_1^\top,\ldots,\boldsymbol{U}_G^\top)$. Let

$$\widetilde{\boldsymbol{u}}_g = \big(C_{U_{1g}|V_0,V_g}(u_{1g}|v_0,v_g),\ldots,C_{U_{d_g g}|V_0,V_g}(u_{d_g g}|v_0,v_g)\big).$$

Assuming the conditional distribution of $\boldsymbol{U}_D$ given $\boldsymbol{V} = \boldsymbol{v}$ has copula density $c_{\text{res}}(\widetilde{\boldsymbol{u}}_1,\ldots,\widetilde{\boldsymbol{u}}_G)$, the joint density of $\boldsymbol{U}_D$ is:

$$c_{\boldsymbol{U_D}}(\boldsymbol{u_D}) = \int_0^1 \cdots \int_0^1 \underbrace{\Big\{\prod_{g=1}^{G}\prod_{j=1}^{d_g} c_{U_{jg}V_0}(u_{jg},v_0)\cdot c_{U_{jg}V_g;V_0}\big(C_{U_{jg}|V_0}(u_{jg}|v_0),v_g\big)\Big\}}_{\text{part1}}$$
$$\times \underbrace{c_{\text{res}}\big(\widetilde{\boldsymbol{u}}_1,\ldots,\widetilde{\boldsymbol{u}}_G\big)}_{\text{part2}} \,\mathrm{d}v_0\cdots\mathrm{d}v_g. \quad (6)$$

For notation, $c_{U_{jg}V_g;V_0}$ is the copula density to represent conditional dependence of U_{jg},V_g given V_0. Part 1 in (6) has the joint density of observed and latent variables assuming conditional independence, while part 2 accounts for conditional dependence of the U_{jg} given the latent variables.

Definition 6. (Oblique factor copula model with residual dependence). With G groups, let $D = \sum_{g=1}^{G} d_g$, with d_g variables from group g. Let $\boldsymbol{D} = (d_1,\ldots,d_G)$. Let $\boldsymbol{V} = (V_1,\ldots,V_G)$. Let $\boldsymbol{U}_g^\top = (U_{1g},\ldots,U_{d_g})$, $g \in \{1,\ldots,G\}$, $\boldsymbol{U}_{\boldsymbol{D}}^\top = (\boldsymbol{U}_1^\top,\ldots,\boldsymbol{U}_G^\top)$. Let $\widetilde{\boldsymbol{u}}_g = (C_{U_{1g}|\boldsymbol{V}}(u_{1g}|\boldsymbol{v}),\ldots,C_{U_{d_g g}|\boldsymbol{V}}(u_{d_g g}|\boldsymbol{v}))$. Assume the conditional distribution of $\boldsymbol{U}_D$ given $\boldsymbol{V} = \boldsymbol{v}$ has copula density $c_{\text{res}}(\widetilde{\boldsymbol{u}}_1,\ldots,\widetilde{\boldsymbol{u}}_G)$. Let $c_{\boldsymbol{V}}$ be the joint copula density of the latent variables. The joint density of $\boldsymbol{U}_D$ is:

$$c_{\boldsymbol{U_D}}(\boldsymbol{u_D}) = \int_0^1 \cdots \int_0^1 \underbrace{\Big\{\prod_{g=1}^{G}\prod_{j=1}^{d_g} c_{U_{jg},V_g}(u_{jg},v_g)\Big\} c_{\boldsymbol{V}}(\boldsymbol{v})}_{\text{part1}}$$
$$\times \underbrace{c_{\text{res}}\big(\widetilde{\boldsymbol{u}}_1,\ldots,\widetilde{\boldsymbol{u}}_G\big)}_{\text{part2}} \,\mathrm{d}v_1\cdots\mathrm{d}v_G. \quad (7)$$

Part 1 in (7) has the joint density of observed and latent variables assuming conditional independence, while part 2 accounts for conditional dependence of the U_{jg} given the latent variables.

3 Proxies for the latent variables

Proxy estimates are defined in [4] for the factor models in Section 2 assuming $\boldsymbol{\Omega}$ is an identity matrix or c_{res} is an independence copula, that is, assuming conditional independence of observed variables given the latent variables. In high dimensions, some weak residual or conditional dependence given the latent variables is not unreasonable and one would hope that the proxy variables are robust. The next Section 4 has some sufficient conditions to explain and interpret weak residual dependence.

3.1 Proxies in Gaussian factor models

Consider a $D \times 1$ random vector $\boldsymbol{z}_D$ from the p-factor Gaussian model with weak residual dependence as discussed in Section 2.1. The correlation matrix of $\boldsymbol{z}_D$ has a factor structure $\boldsymbol{\Sigma}_D = \boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Gamma}_D$, where $\boldsymbol{A}_D$ is a $D \times p$ matrix of factor loadings, with p being the number of factors. Here, $\boldsymbol{\Gamma}_D$ represents the covariance of $(\psi_j \varepsilon_j : 1 \le j \le D)$. The matrix $\boldsymbol{\Gamma}_D$ is the conditional covariance matrix of the observed variables given the latent variables, denoted as $\boldsymbol{\Gamma}_D = \mathrm{Cor}(\boldsymbol{Z}_D, \boldsymbol{Z}_D | \boldsymbol{W})$; it is not diagonal under the assumption of residual dependence. Define $\boldsymbol{\Psi}_D^2 = \mathrm{diag}(\psi_1^2, \ldots, \psi_D^2)$, where the off-diagonal elements of $\boldsymbol{\Gamma}_D$ have been set to 0. The proxies are defined as the conditional expectation of the latent variables given the observed variables, assuming conditional independence.

In Gaussian factor models with residual dependence in Section 2.1, the conditional expectation proxies are defined with the loading matrix $\boldsymbol{A}_D$ and $\boldsymbol{\Psi}_D$ following the same expressions as presented in [4].

1. For the 1-factor model in Definition 1, let $(w^0, \boldsymbol{z}_D)$ be a realization of $(W, \boldsymbol{Z}_D)$. The proxy for W (or estimate of w^0) given $\boldsymbol{z}_D$, based on $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$, is:

$$\begin{aligned} \widetilde{w}_D &= {\boldsymbol{A}_D}^\top (\boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2)^{-1} \boldsymbol{z}_D \\ &= (1 + {\boldsymbol{A}_D}^\top \boldsymbol{\Psi}_D^{-2} \boldsymbol{A}_D)^{-1} \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2} \boldsymbol{z}_D, \end{aligned} \tag{8}$$

   if $\boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2$ is non-singular and $\boldsymbol{\Psi}_D$ has no zeros on its diagonal.
2. For the bi-factor model in Definition 2, let $(w_0^0, w_1^0, \ldots, w_G^0, \boldsymbol{z}_D)$ be a realization of $(W_0, W_1, \ldots, W_G, \boldsymbol{Z}_D)$. The proxies (estimates of $w_0^0, w_1^0, \ldots, w_G^0$) given $\boldsymbol{z}_D$, based on $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$, are:

$$\widetilde{w}_0 = (\boldsymbol{a}_0)^\top (\boldsymbol{A_D}\boldsymbol{A_D}^\top + \boldsymbol{\Psi}_{\boldsymbol{D}}^2)^{-1}\boldsymbol{z_D}, \tag{9}$$

$$\widetilde{w}_g(\widetilde{w}_0) = (\boldsymbol{b}_g^\top, 0)\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{z}_g^\top, \widetilde{w}_0)^\top, \quad g \in \{1,\ldots,G\}, \tag{10}$$

where $\boldsymbol{a}_0 = \boldsymbol{a}_{0,D}$ is the first column of the loading matrix, $\boldsymbol{z}_g = (z_{1g},\ldots,z_{d_g g})^\top$, $\boldsymbol{b}_{0g} = \boldsymbol{b}_{0,d_g g}$ and $\boldsymbol{b}_g = \boldsymbol{b}_{d_g g}$ are the $d_g \times 1$ global and local loading vectors for group g. Let $\boldsymbol{B}_g = [\boldsymbol{b}_{0g}, \boldsymbol{b}_g]$ (matrix of size $d_g \times 2$). Let $\boldsymbol{\Sigma}_g$ be the correlation matrix of $(\boldsymbol{Z}_g^\top, W_0)$. Then $\boldsymbol{\Sigma}_g = \begin{bmatrix} \boldsymbol{B}_g\boldsymbol{B}_g^\top + \boldsymbol{\Psi}_g^2 & \boldsymbol{b}_{0g} \\ \boldsymbol{b}_{0g}^\top & 1 \end{bmatrix}$ for $g \in \{1,\ldots,G\}$ with $\boldsymbol{\Psi}_D^2 = \mathrm{diag}(\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_G)$.

3. For the oblique factor model in Definition 3, the proxies (estimates of $w_1^0,\ldots,w_G^0$) given $\boldsymbol{z}_D$, based on $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$, are based on separate 1-factor models for each group.

3.2 Proxies in factor copula models

In this section, the proxies, based on the assumption that c_{res} comes from the independence copula, are summarized.

1. For the 1-factor copula model in Definition 4, let $(v^0, \boldsymbol{u}_D)$ be a realization of $(V, \boldsymbol{U}_D)$. The proxy estimate of v^0 is:

$$\widetilde{v}_D = \widetilde{v}_D(\boldsymbol{u}_D) = \frac{\int_0^1 v \prod_{j=1}^D c_{j0}(u_j, v)\,\mathrm{d}v}{\int_0^1 \prod_{j=1}^D c_{j0}(u_j, v)\,\mathrm{d}v}. \tag{11}$$

2. For the bi-factor copula model in Definition 5, let $(v_0^0, v_1^0, \ldots, v_G^0, \boldsymbol{u}_D)$ be a realization of $(V_0, V_1, \ldots, V_G, \boldsymbol{U}_D)$ with $\boldsymbol{u}_D = (\boldsymbol{u}_1,\ldots,\boldsymbol{u}_G)$. Based on conditional expectations, the proxy estimates of v_0^0 and v_g^0 are given below as $\widetilde{v}_{0\boldsymbol{D}}(\boldsymbol{u_D})$ and $\widetilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g, \widetilde{v}_{0\boldsymbol{D}})$. Let

$$f_g(\boldsymbol{u}_g, v_0) = \int_0^1 \prod_{j=1}^{d_g} c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)\,\mathrm{d}v_g, \quad g \in \{1,\ldots,G\}.$$

With $\boldsymbol{D} = (d_1,\ldots,d_g)$,

$$\widetilde{v}_{0\boldsymbol{D}}(\boldsymbol{u_D}) = \frac{\int_0^1 v_0 \prod_{g=1}^G \{\prod_{j=1}^{d_g} c_{U_{jg},V_0}(u_{jg}, v_0) \cdot f_g(\boldsymbol{u}_g, v_0)\}\,\mathrm{d}v_0}{\int_0^1 \prod_{g=1}^G \{\prod_{j=1}^{d_g} c_{U_{jg},V_0}(u_{jg}, v_0) \cdot f_g(\boldsymbol{u}_g, v_0)\}\,\mathrm{d}v_0}. \tag{12}$$

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

$$\widetilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g,\widetilde{v}_{0\boldsymbol{D}}) = \frac{\int_0^1 v_g \prod_{j=1}^{d_g} c_{U_{jg},V_0}(u_{jg},\widetilde{v}_{0\boldsymbol{D}})\cdot c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|\widetilde{v}_{0\boldsymbol{D}}),v_g)\,\mathrm{d}v_g}{\int_0^1 \prod_{j=1}^{d_g} c_{U_{jg},V_0}(u_{jg},\widetilde{v}_{0\boldsymbol{D}})\cdot c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|\widetilde{v}_{0\boldsymbol{D}}),v_g)\,\mathrm{d}v_g}. \tag{13}$$

3. For the oblique factor copula in Definition 6, let $(v_1^0,\dots,v_G^0,\boldsymbol{u}_1,\dots,\boldsymbol{u}_G)$ be a realization of $(V_1,\dots,V_G,\boldsymbol{U}_D)$. The proxies are given below as $\widetilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g)$ for $g\in\{1,\dots,G\}$:

$$\widetilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g) = \frac{\int_0^1 v_g \prod_{j=1}^{d_g} c_{U_{jg},V_g}(u_{jg},v_g)\,\mathrm{d}v_g}{\int_0^1 \prod_{j=1}^{d_g} c_{U_{jg},V_g}(u_{jg},v_g)\,\mathrm{d}v_g}. \tag{14}$$

4 Consistency of proxies: model known

General proofs for factor copula models are not possible or would take too much space because of the generality for c_{res}. For Gaussian factor models with parameters that are loading and correlation matrices, proofs are possible with interpretable conditions for weak residual dependence. These provide insights into conditions for the consistency of the proxies in the factor copula counterparts, and these conditions provide more concise definitions of weak residual dependence.

The theorems are established assuming all parameters in the factor model and residual dependence are known. We have to know that proxy estimates are consistent in this case in order to discuss a method for determining parametric bivariate linking copula families for factor copula models and estimating the parameters.

With proxy estimates of latent variables in high dimensions, existing methods for vine copulas can be used to choose bivariate linking copula families in Definitions 4 to 6.

To consider factor copula models, the variables are monotonically related to each other and to the latent variables. Without this assumption, there could be identifiability issues. For data, assuming the variables are monotonically related and the correlation matrix $\boldsymbol{R}_{data}$ of normal scores summarizes the dependence well, the difference with $\boldsymbol{R}_{model}$ based on a 1-factor or structured factor model can be computed to assess which variables have residual dependence; see more details in Section 6.

In this section, sufficient conditions are obtained on the residual dependence so that the proxies defined in Section 3 for Gaussian and copula factor models with weak residual dependence remain consistent as the dimension $D\to\infty$. The interpretation of the sufficient conditions is that the total dependence of any variable with the other variables, conditioned on the latent variables, is bounded as $D\to\infty$.

Section 4.1 discusses results for the consistency of factor scores in Gaussian factor models, while Section 4.2 presents results for the consistency of proxy variables in factor copula models. The assumptions regarding the strength of the residual dependence are similar for both Gaussian factor models and factor copula models.

4.1 Consistency in Gaussian factor models

For the 1-factor Gaussian model with weak residual dependence in Definition 1, suppose the maximum eigenvalue of $\boldsymbol{\Gamma}_D$ is bounded as $D \to \infty$, the model is an approximate factor model from the definition in [1]. This assumption is sufficient for the proxy in (8) to be asymptotically consistent. An assumption that is more straightforward and can be extended to the copula case is provided below.

Assumption 1 Let $\boldsymbol{\Gamma_D} = (\gamma_{st})_{1\le s,t\le D}$ be the conditional covariance matrix of observed variables given the latent variables in Definition 1. Let $S_D = \sum_{j=1}^{D} \psi_j \varepsilon_j$ and $\bar{\varepsilon} = S_D/D$, then $\mathrm{E}(S_D^2) = \sum_{s=1}^{D}\sum_{t=1}^{D} \gamma_{st}$. Assume

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

where M is a positive constant.

Remark 1. Let $\gamma_{s+}^{(D)} = \sum_{t=1}^{D} \gamma_{st}$ for $s \in \{1,\dots,D\}$. Below are examples where Assumption 1 is satisfied and not satisfied.

- (Satisfied). If $\gamma_{s+}^{(D)}$ is $O(1)$ as $D \to \infty$ for all s, then Assumption 1 is satisfied; for example, ε's are indexed to have ante-dependence of order 1: $0 < r_1 < \gamma_{j,j+1} < r_2 < 1$ for all j, and $\gamma_{jk} = \prod_{i=j}^{k-1} \gamma_{i,i+1}$ for $k - j \ge 2$.
- (Not satisfied). If $\gamma_{s+}^{(D)}$ is $O(D)$ as $D \to \infty$ for at least one s, then Assumption 1 is not satisfied; for example, ε_1 is dominating: $0 < r_1 < \gamma_{1j} < r_2 < 1$ for all j, and $\gamma_{jk} = \gamma_{1j} \cdot \gamma_{1k}$.

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

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

With Assumption 1 on residual dependence, for the proxy defined in equation (8), then $\widetilde{w}_D - w^0 = O_p(D^{-1/2})$ as $D \to \infty$.

Proof. The proof is in Section 9.1.

Remark 2. Assumption 1 implies $\mathrm{Var}(\bar{\varepsilon}) = O(D^{-1})$, the same order as the case of independent and identically distributed. The convergence rate at $\sqrt{D}$ for the proxy variable is the same with the assumption of conditional independence of observed variables given the latent variable.

The consistency of proxies in the oblique factor model is a straightforward corollary of Theorem 1, with an assumption of residual dependence similar to Assumption 1. The observed variables $\boldsymbol{Z}_g$, $g \in \{1,\dots,G\}$ in one group g are linked to the same (realized) latent variable w_g^0 and these variables satisfy a 1-factor model with weak residual dependence. The details of the theorems are omitted.

The conditions of the bi-factor Gaussian model for consistency of the proxy variables are a bit stronger than that required for the 1-factor model.

Assumption 2 *For the bi-factor model with weak residual dependence and a fixed number of G groups in* (3), $D = \sum_{g=1}^{G} d_g$. *Let* $\boldsymbol{\Gamma_D} = (\gamma_{st})_{1\le s,t\le D}$ *be the conditional covariance matrix of disturbances* $\{\varepsilon_{jg}\}$. *Assume*

$$0 < \liminf_{D\to\infty} D^{-1} \sum_{s=1}^{D}\sum_{t=1}^{D} |\gamma_{st}| < \limsup_{D\to\infty} D^{-1} \sum_{s=1}^{D}\sum_{t=1}^{D} |\gamma_{st}| < M,$$

where M is a positive constant.

Theorem 2. *(Asymptotic properties of factor scores in bi-factor Gaussian model with residual dependence model). For the bi-factor Gaussian factor model with fixed G groups defined in Definition 2, with given sequences* $\{\alpha_{0,1g},\alpha_{0,2g},\dots\}$ *and* $\{\alpha_{1g},\alpha_{2g},\dots\}$ *for* $g\in\{1,\dots,G\}$, *suppose there are realized infinite sequences of observed variables* $\boldsymbol{z}_1^\top = (z_{1,1},z_{2,1},\dots)$, $\dots,\boldsymbol{z}_G^\top = (z_{1,G},z_{2,G},\dots)$ *and realized sequences of disturbance terms* $\boldsymbol{e}_1^\top = (e_{1,1},e_{2,1},\dots)$, $\dots,\boldsymbol{e}_G^\top = (e_{1,G},e_{2,G},\dots)$, *with latent variables* $\boldsymbol{w}^0 = (w_0^0,w_1^0,\dots,w_G^0)$. *Truncate the sequences to the first* d_g *variables* $\boldsymbol{z}_{d_g,g}$ *for* $g\in\{1,\dots,G\}$, *with no* d_g *dominating others. Let* $\boldsymbol{z_D} = (\boldsymbol{z}_{d_1,1}^\top,\dots,\boldsymbol{z}_{d_G,G}^\top)^\top$, $\boldsymbol{e_D} = (\boldsymbol{e}_{d_1,1}^\top,\dots,\boldsymbol{e}_{d_G,G}^\top)$ *and assume that the loading matrices*

$$\boldsymbol{A_D} = [\boldsymbol{a}_{0,D}, \mathrm{diag}(\boldsymbol{b}_{d_1 1},\dots,\boldsymbol{b}_{d_G G})]$$

are of full rank, with bounded condition number over $d_g\to\infty$ *for all g. With* $\boldsymbol{b}_{d_g g} = (\alpha_{1g},\dots,\alpha_{d_g,g})^\top$, *assume* $D^{-1}\|\boldsymbol{a}_{0,D}\|_1 \not\to 0$, $d_g^{-1}\|\boldsymbol{b}_{d_g g}\|_1 \not\to 0$ *for* $g\in\{1,\dots,G\}$, *and* $\liminf_{j,g}\psi_{jg}^2 > 0$. *Let* $\widetilde{w}_{\boldsymbol{D}} = (\widetilde{w}_0,\dots,\widetilde{w}_G)$ *be the factor scores defined in* (9) *and* (10). *With Assumption 2 on the residual dependence, then*

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

Proof. See Section 9.2 for the proof.

The assumptions $D^{-1}\|\boldsymbol{a}_{0,D}\|_1 \not\to 0$, $d_g^{-1}\|\boldsymbol{b}_{d_g g}\|_1 \not\to 0$ for $g\in\{1,\dots,G\}$ are similar to an assumption in Theorem 1. The average absolute loading in any column of $\boldsymbol{A_D}$ does not go to 0 as $D\to\infty$ (otherwise, the latent variables do not explain enough dependence in the observed variables).

4.2 Consistency in factor copula models

In this section, we present the conditions on the residual dependence for the consistency of the proxy variables in the 1-factor copula models with residual dependence defined in Section 2.2. We assume that the linking copulas between the observed variables and the latent variables, as well as the copula for the residual dependence, are known. (This step is needed before the case where parametric families of linking copulas are estimated.) For tractability of a proof, we also assume that residual dependence conditioned on the latent variables, is summarized by a multivariate Gaussian copula. The results under this simplified assumption provide insights into (a) the 1-factor copula model when the residual dependence is modeled by other well-behaved parametric copulas instead of Gaussian copulas, and (b) the bi-factor copula model with weak residual dependence.

The regularity conditions on the linking copulas stated in Assumption 1 in [4] are also necessary. These conditions require the linking copulas from the observed variables to the latent variables to be well-behaved, and the dependence between the observed and latent variables to be strong enough and bounded away from comonotonicity and countermonotonicity. (If the dependence can approach comonotonicity or countermonotonicity, then one of the variables can be taken as the proxy.) Additionally, further assumptions on the residual dependence are required, analogous to Section 4.1. It is required that for any variable, the total dependence of the variable with the other variables, conditioned on the latent variables, is bounded as $D \to \infty$. In the following remark, we provide an outline of the proof for the consistency of proxy variables in the 1-factor copula model.

Remark 3. Consider a realization $(v^0, \boldsymbol{u}_D)$ of $(V, \boldsymbol{U}_D)$ where v^0 is to be estimated based on $\boldsymbol{u}_D$. Let $\overline{L}_D(v; \boldsymbol{u}_D)$ be the average negative log-likelihood treating v as a parameter, that is,

$$\overline{L}_D(v;\boldsymbol{u}_D) := \underbrace{-D^{-1}\sum_{j=1}^{D}\log c_{j0}(u_j, v)}_{:=\overline{qL}_D(v,\boldsymbol{u}_D)} \underbrace{-\frac{1}{D}\log c_{\mathrm{res}}(C_{1|0}(u_1|v), \ldots, C_{D|0}(u_D|v))}_{:=\overline{L}_D^{(r)}(v,\boldsymbol{u}_D)}$$

so that $\overline{L}_D(v;\boldsymbol{u}_D) = \overline{qL}_D(v;\boldsymbol{u}_D) + \overline{L}_D^{(r)}(v;\boldsymbol{u}_D)$, where $\overline{L}_D^{(r)}$ comes from the residual dependence given the latent variable.

From the proof of Theorem 5 in [4], the proxy $\widetilde{v}_D$ and maximum likelihood estimate v_D^* with $c_{\mathrm{res}} \equiv 1$ are asymptotically equivalent. With residual dependence, the optimum v_D^* estimated from $\overline{qL}_D$ is now a quasi-MLE. Therefore, to show the consistency of proxy variables, we need to show that the quasi-MLE v_D^* converges to the realized latent variable v^0 under some conditions on the weak residual dependence as $D \to \infty$.

The gradient of the negative log-likelihood for v is

$$\frac{\partial \overline{L}_D}{\partial v} = \frac{\partial \overline{qL}_D}{\partial v} + \frac{\partial \overline{L}_D^{(r)}}{\partial v} \tag{15}$$

The proof mainly consists of two steps.
A. Show that the limiting inference (score) function or derivative of $\overline{L}_D$ with respect to v has a unique solution at v^0.
B. Show that the gradient of the $\overline{L}_D^{(r)}$ with respect to v at v^0 goes to zero as $D \to \infty$.

To prove $v_D^* \to v^0$, for Step B, it is sufficient to justify the Fisher consistency, by showing

$$\left.\frac{\partial \overline{L}_D^{(r)}}{\partial v}\right|_{v=v^0} \to_p 0, \quad D \to \infty. \tag{16}$$

This implies $\lim_{D\to\infty} \partial \overline{qL}_D(v;\boldsymbol{u}_D)/\partial v \,\big|_{v=v^0} = 0$ since, from step A (under usual regularity conditions for the likelihood), $\lim_{D\to\infty} \partial \overline{L}_D(v;\boldsymbol{u}_D)/\partial v|_{v=v^0} = 0$.

The Fisher consistency can be derived by a first-order Taylor expansion. Expanding $\partial \overline{qL}_D(v^0;\boldsymbol{u}_D)/\partial v$ around the quasi-MLE v_D^* leads to

$$\frac{\partial \overline{qL}_D(v_D^*;\boldsymbol{u}_D)}{\partial v} + \frac{\partial^2 \overline{qL}_D(v_D^*;\boldsymbol{u}_D)}{\partial v^2}(v^0 - v_D^*)\ .$$

Since $\partial \overline{qL}_D(v_D^*;\boldsymbol{u}_D)/\partial v = 0$ and assuming $\lim_{D\to\infty} |\partial^2 \overline{qL}_D(v_D^*;\boldsymbol{u}_D)/\partial v^2|$ is bounded, then $v_D^* - v^0 \to 0$ when $D \to \infty$.

The Fisher consistency in (16) is justified in Theorem 3 for the 1-factor copula model with Assumption 3 and Condition 1 below.

Assumption 3 Consider the 1-factor model with residual dependence in Definition 4.
(i) Assume the bivariate linking copulas $\{C_{j0}(u,v)\}$ are such that the derivatives $\{\partial C_{j|0}(u|v)/\partial v\}$ are uniformly bounded for all j and the limits of the derivatives go to 0 when $u \to 1$ or $u \to 0$ for $v_A \le v \le v_B$, where v_A is close to 0 and v_B is close to 1.
(ii) Assume the bivariate linking copulas $\{C_{j0}(u,v)\}$ are such that the ratio of derivatives $[\partial^2 C_{j|0}(u|v)/\partial v \partial u]/[\partial C_{j|0}(u|v)/\partial u]$ are finite and uniformly bounded over j when $u \to 1$ or $u \to 0$ and for $v_A \le v \le v_B$, where v_A is close to 0 and v_B is close to 1.
(iii) Let $\widetilde{U}_j^0 = C_{j|0}(U_j|V = v^0)$, and $E_j^0 = \Phi^{-1}(\widetilde{U}_j^0)$ for all j. Assume that $\sum_{j=1}^D \sum_{k=1}^D |\mathrm{Cov}(\widetilde{U}_j^0, \widetilde{U}_k^0)| = O(D)$ or $\sum_{j=1}^D \sum_{k=1}^D |\mathrm{Cov}(E_j^0, E_k^0)| = O(D)$.

Condition 1 *For a Gaussian copula for residual dependence, with correlation matrix parameter $\boldsymbol{\Gamma}_D = (\gamma_{jk})$, with its (j,k) entry given by* $\mathrm{Cov}(E_j^0, E_k^0)$. *Let $\boldsymbol{\Gamma}_D^{-1} = (\gamma^{jk})_{1\le j,k\le D}$. Assume $\sum_{j=1}^D |\gamma^{jk}| = O(1)$ and $\sum_{j=1}^D |\gamma_{jk}| = O(1)$ for all k.*

Remark 4. One example where Condition 1 is satisfied is for an m-truncated Gaussian partial correlation vine. In [6], a positive $D \times D$ correlation matrix can be re-parametrized into D correlations and $D(D-1)/2$ partial correlations which are algebraically independent in $(-1,1)$. The re-parametrization is not unique and can be represented by vine trees. Consider the vine trees truncated to m levels, from Proposition 2 in [7], there are $(D-m)(D-m-1)/2$ positions of the inverse correlation matrix $\boldsymbol{\Gamma}_D^{-1}$ that are zero and thus Condition 1 is satisfied.

Theorem 3. *(Consistency of proxy in 1-factor copula model with weak residual dependence when linking copulas are known). For the 1-factor model with weak residual dependence in Definition 4, suppose there is a realized infinite sequence $u_1, u_2, \ldots,$ with latent variable v^0 (independent of dimension D). Truncate the sequence to the first D variables and let $\boldsymbol{u}_D = (u_1, \ldots, u_D)^\top$. Define the average negative quasi-log-likelihood in v as*

$$\overline{qL}_D(v) = -D^{-1} \sum_{j=1}^{D} \log c_{j0}(u_j, v). \tag{17}$$

Assume $\lim_{D\to\infty} \overline{qL}_D$ has a global minimum for all $D \geq d$ (for some large $d > 0$) and the second derivative $\overline{qL}''_D$ at the global minimum has a limit infimum that is strictly positive. Also, assume the likelihood function in (17) *satisfies the usual regularity conditions in Assumption 1 in [4]. With Assumption 3, Condition 1, stochastic increasingness ($1 - C_{j|0}(\cdot|v)$ increasing in v for all j), and $v_A < v^0 < v_B$, the proxy $\widetilde{v}_D$ defined in* (11) *satisfies $\widetilde{v}_D - v^0 = o_p(1)$ as $D \to \infty$.*

Proof. The proof is in Section 9.3.

Remark 5. Assumption 3 requires the linking copulas from the observed variables to the latent variable are well-behaved such that the limits of some partial derivatives or ratios of some partial derivatives satisfy certain conditions when the observed variables reach the boundary. If the linking copulas for modeling the residual dependence are all bivariate Gaussian copulas, these conditions are easy to check and are satisfied. This assumption is mild, as it holds for commonly used parametric copula families such as the Gumbel, BB1 and Frank families; this can be verified through analytical and numerical checks. The conditions in Assumption 3(iii) on the residual correlation matrix are parallel to the Gaussian copula case. These assumptions on the linking copulas and on the residual dependence provide insights for the bi-factor and oblique-factor models.

5 Consistency of proxy variables with estimated parameters

In this section, rather than completely specified factor models, we assume there are loading parameters for the Gaussian factor models or parameters for bivariate linking copulas in the factor copula models that have to be estimated. For factor models with residual dependence, the estimation based on the log-likelihood assuming $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$ or c_{res} being an independence copula implies estimation with a mis-specified likelihood, and the resulting estimator from this mis-specified log-likelihood is a quasi-MLE.

In this section, weak sufficient conditions for residual dependence are given so that it can be proved for the 1-factor Gaussian models with residual dependence that the quasi-MLEs of loading parameters are consistent. Note that parameters for the residual dependence are not involved in the estimation of the proxies. The log-likelihoods in this section are functions of the loading parameters and not the latent variables (as in Section 4.1).

Let $\boldsymbol{Y}_{1,D},\ldots,\boldsymbol{Y}_{N,D},\ldots$ be an (infinite) random sample from a density g_D and let the realizations be $\boldsymbol{y}_{1,D},\ldots,\boldsymbol{y}_{N,D},\ldots$. ($Y=Z$ for Gaussian factor models and $Y=U$ for factor copula models.) The ith observation vector is $\boldsymbol{y}_{i,D}=(y_{i1},\ldots,y_{iD})^\top$. Suppose $g_D(\boldsymbol{y}_D)$ is the density with residual dependence and $f_D(\boldsymbol{y}_D,\boldsymbol{\theta})$ is the parametric density of the corresponding factor model assuming conditional independence (given latent variables). The quasi-MLE maximizes the quasi-log-likelihood

$$\hat{\boldsymbol{\theta}}_D=\hat{\boldsymbol{\theta}}_{D,N}=\hat{\boldsymbol{\theta}}_D(\boldsymbol{y}_{1,D},\ldots,\boldsymbol{y}_{N,D})=\operatorname{argmax}\sum_{i=1}^{N}\log f_D(\boldsymbol{y}_{i,D};\boldsymbol{\theta}).$$

Under some regularity conditions, such as in [13], including the MLE asymptotically being in the interior of the parameter space, then $\hat{\boldsymbol{\theta}}_D\to_p\boldsymbol{\theta}_D^*$ as $N\to\infty$ (with fixed D), where $\boldsymbol{\theta}_D^*$ minimizes the Kullback-Leibler (KL) divergence

$$\int g_D(\boldsymbol{y}_D)\log\Big[\frac{g_D(\boldsymbol{y}_D)}{f_D(\boldsymbol{y}_D;\boldsymbol{\theta})}\Big]\,\mathrm{d}\boldsymbol{y}_D. \tag{18}$$

For $j=1,2,\ldots$, let $\boldsymbol{\theta}_j$ be the dependence parameter vector for observation j linking to the latent variables. The parameter vectors $\boldsymbol{\theta}_j$ are specified in the Gaussian factor models with weak residual dependence as below.

- Gaussian 1-factor: θ_j is a loading parameter or correlation parameter with the latent variable, that is $\boldsymbol{\theta}_j=\alpha_j$.
- Gaussian bi-factor: $\boldsymbol{\theta}_j=(\alpha_{0,jg},\alpha_{jg})$ if observation j is in group g.
- Gaussian oblique factor: $\boldsymbol{\theta}_j=\alpha_{jg}$ if observation j is in group g. For Gaussian oblique factor models, there are additional parameters $\{\rho_{ij}:1\le i<j\le G\}$ for the correlations of the latent variables.

For factor copula models, the parameters are in the copulas in Definitions 4 to 6.

For fixed D, let $\boldsymbol{\theta}_D^*=\big\{\theta_{j,D}^*:j\in\{1,\ldots,D\}\big\}$ be the vector minimizing the KL divergence of f_D relative to g_D in (18).

A theorem is given only for the case of the Gaussian 1-factor model (see Theorem 4) and the proof is long. The details are much more complicated than in [4]. The theorem is stated in Section 5.2 with its proof in Section 9.4. The conditions for the theorem can be intuitively extrapolated to 1-factor and group-based structured factor copula models. Section 7 has simulation studies showing the effect of weak residual dependence on parameter and latent variable estimation in factor copula models. For a specific simulation scenario with weak residual dependence, the key properties of the score vector and Hessian matrix in the proof of Theorem 4 and its lemmas can be verified numerically.

5.1 Estimation of parameters in blocks

To avoid the rate of convergence with both N and D increasing to infinity, we use the property of closure under margins for factor models so that parameters can be estimated in blocks as D gets large. This is valid with a super-population assumption for the variables (and corresponding parameters linking to latent variables). This is the methodology of proofs used in [4] for parameter estimation.

With residual dependence, for block-estimation to be valid, there is an additional assumption. The variables are assumed to be organized in blocks such that the conditional correlations between the blocks conditioned on the latent variables is weaker than the conditional correlations for the observed variables within blocks.

With the assumption that the residual dependence is sampled from some super-populations, such that for any variable, the total dependence of the variable with other variables given the latent variables is bounded as $D \to \infty$, the model maintains the "closed-under-margin" property by blocks. Therefore, the block-estimation method discussed in Section 5.3 in [4] can be utilized to estimate the increasing parameters in the loading matrix for Gaussian factor models or on the linking copulas in factor copula models in sequential blocks. The block size d_{block} is similar across blocks, and the number of blocks increases with the growing number of observed variables.

5.2 Differences between quasi-MLEs and true parameters

In this section, we mainly show that with the super-population assumption on the residual dependence, the differences of the quasi-MLEs estimated in sequential blocks and the true parameters are of order $O(N^{-1/2}) + O(d_{\text{block}}^{-1})$. The remark below discusses an outline of this result.

Remark 6. Consider a random sample of size N with D variables. The variables are split into K blocks for parameter estimation. Let the kth block be $\mathscr{B}_k$ and the block size $d_k \approx d_{\text{block}}$, for $k = 1, \dots, K$. The differences between the quasi-MLEs and true parameters in each block come from two aspects.

Firstly, based on analysis of Kullback-Leibler divergence between f_D and g_D, under conditions in Assumption 4 on the weak residual dependence,

$$\theta^*_{j,d_{\text{block}}} - \theta_j = O(d_{\text{block}}^{-1}) \quad \text{for } j \in \mathscr{B}_k. \tag{19}$$

Secondly, consider the sampling variability. The quasi-MLEs for a sample of size N converge to the minimizer of the KL divergence at rate $O(1/\sqrt{N})$. Let the quasi-MLEs be $\hat{\theta}_{j,N,d_{\text{block}}}$ in one block $\mathscr{B}_k$. Under the general regularity conditions for quasi-maximum likelihood estimation [13],

$$\hat{\theta}_{j,N,d_{\text{block}}} - \theta^*_{j,\text{block}} = O_p(N^{-1/2}), \quad j \in \mathscr{B}_k.$$

Therefore, the differences between the quasi-MLEs and the true parameters are of order $O(N^{-1/2}) + O(d_{\text{block}}^{-1})$. If the block size is sufficiently large, the differences between the quasi-MLEs and the true parameters in each block will be dominated by sampling variability.

The assumptions on residual dependence, ensuring (19) is satisfied, are outlined in Assumption 4. The statements for the 1-factor Gaussian model are presented in Theorem 4. The conditions in Assumption 4 can be extended to other factor (copula) models, such as the 1-factor copula model and bi-factor Gaussian (copula) models with residual dependence. The mathematical proof for the 1-factor Gaussian model is tractable but lengthy. The details of the other models are omitted due to the tedious calculations.

Assumption 4 *(Super-population): (1) In the first-level super-population, the observed variables (or their correlations, partial correlations, or linking copulas) are sampled from appropriate super-populations. (2) In the second-level super-population, the conditional dependencies of any pair of variables are assumed to be sampled from some super-populations such that for any observed variable j, the total dependence of this variable with other variables with indices in* $\{1,\dots,j-1,j+1,\dots\}$*, given the latent variables, is bounded* $O(1)$ *as* $D \to \infty$*.*

In the 1-factor Gaussian factor model with weak residual dependence in (2)*, for a fixed dimension, let* $\boldsymbol{\Gamma}_D = (\gamma_{st})_{1\le s,t\le D}$ *be the conditional covariance matrix of observed variables given the latent variables. Assume that*

$$\sum_{\ell=1}^{D} |\gamma_{\ell k}| = \sum_{\ell=1}^{D} |\gamma_{k\ell}| = O(1).$$

uniformly holds for all k.

The super-population assumption in the first level essentially avoids the cases where additional variables, indexed by j as $j \to \infty$, have dependence with the latent variable that converges to independence or perfect dependence. The super-population assumption in the second level means that the pattern of conditional dependence of a variable j with the remaining variables is similar over index j.

The conditions specified in the second paragraph of Assumption 4 are stricter than those in Assumption 1, because the condition in Assumption 4 serves as a sufficient condition to establish the condition in Assumption 1. The examples provided in Remark 1 show cases where the assumption could fail or be satisfied.

Theorem 4. *(Difference between KL minimizer and vector of true loadings in 1-factor Gaussian model with residual dependence). For the model in Definition 1 with loadings* α_j *for* $1 \le j \le D$*, let the correlation matrix of* g_D *be* $\boldsymbol{R}_D^0 = \boldsymbol{\alpha}_D\boldsymbol{\alpha}_D^\top + \boldsymbol{\Gamma}_D$*, where* $\boldsymbol{\Gamma}_D = (\gamma_{ij})_{1\le i,j\le D}$*. The correlation matrix of* f_D *is* $\boldsymbol{\Sigma}_D = \boldsymbol{\beta}_D\boldsymbol{\beta}_D^\top + \boldsymbol{\Psi}_D^2$*, with* $\boldsymbol{\Psi}_D^2 = \mathrm{diag}(1-\beta_j^2; 1 \le j \le D)$.

Let $\boldsymbol{\beta}_D^*$ *be the minimizer of the KL divergence between true* g_D *and* f_D*. With the super-population assumption and conditions in Assumption 4, as well as the loading parameters being bounded away from* ± 1*, then* $\boldsymbol{\beta}_D^* - \boldsymbol{\alpha}_D = O(D^{-1})$.

Proof. See Section 9.4 for the proof.

In [4], when $\boldsymbol{\Omega}_D = \boldsymbol{I}_D$ or c_{res} is the independence copula, under the super-population assumption, the differences between the estimated loading matrix for the Gaussian factor models or the estimated parameters on the linking copulas for the factor copula model and the true parameters are of order $O(N^{-1/2})$ uniformly in each block when the block size is large enough. The proxies are defined the same way as in [4], involving only the links from the observed variables to the latent variables. Therefore, Lipschitz continuity of the proxy variables presented in Section 5.1 of [4] still holds. Following the same logic, the consistency also holds when the parameters are estimated as both N and D increase to infinity.

6 Sequential estimation of parameters and latent variables

This section summarizes how estimation for factor copula models can proceed with weak residual dependence. The estimation method is used in Section 7 for some simulation examples. It is also used in the applications in [3].

We suppose the context of the data suggests that a factor model should be suitable. The steps in using proxies and assessing weak residual dependence are the following, assuming there is a random sample $\{(y_{i1},\dots,y_{id}) : i = 1,\dots,N\}$, from a multivariate cdf $F_{\boldsymbol{Y}}$.

1. Fit d univariate distributions and apply probability integral transforms, or apply rank transforms to get $\{\boldsymbol{u}_i = (u_{i1},\dots,u_{id}) : i = 1,\dots,N\}$.
2. Apply the standard normal quantile function Φ^{-1} to each u_{ij} to get vectors of normal scores $\{(z_{i1},\dots,z_{id}) : i = 1,\dots,n\}$.
3. Let $\boldsymbol{R}_{data}$ be the $d \times d$ correlation matrix of normal scores.
4. If there is no group structure, fit a 1-factor Gaussian model to $\boldsymbol{R}_{data}$. If there is a group structure with $G \geq 2$ groups, fit bi-factor and oblique-factor Gaussian models to $\boldsymbol{R}_{data}$ and keep the better fitting model. Let $\boldsymbol{R}_{model}$ be the model-based correlation matrix computed as a function of the loading parameters.
5. Compare $\boldsymbol{R}_{data}$ and $\boldsymbol{R}_{model}$ to note the (j,k) entries where $|R_{data,jk} - R_{model,jk}|$ exceeds a threshold and might indicate weak residual dependence.
6. Assuming few instances of exceeding the threshold, use the residual dependence pattern to find a parsimonious truncated vine structure (e.g., [7]) for the weak residual dependence.
7. If normal scores plots suggest departures from a Gaussian copula for the multivariate dependence, apply the proxy estimation theory in previous Sections of this article and in [4] for estimation of a parametric factor copula model.

The steps for estimating latent variables with proxies are as follows. Simple proxies (as in [4]) are used in the first stage to help in selecting parametric linking copula families from the observed to their corresponding latent variables and obtaining the initial estimates of the copula parameters. Then, “conditional expectation” proxies in Section 3.2 are constructed and used to estimate the parameters by optimizing the approximate (complete) log-likelihood with the latent variables replaced by the proxy variables.

Some details are repeated from [4] for the 1-factor copula model; they are similar for the bi-factor model except the simple proxy is based on a bi-factor Gaussian approximation.

For the parametric 1-factor copula with weak residual dependence in Definition 4, suppose there is realized random sample of size N, with ith observation vector $\boldsymbol{u}_i = (u_{i1}, \dots, u_{iD})$ as an independent realization of $\boldsymbol{U} = (U_1, \dots, U_D)$. If the latent variables $(v_1, \dots, v_N)$ are assumed observed, the complete log-likelihood, with parametric bivariate linking copula $C_{j0}(\cdot; \boldsymbol{\theta}_j)$ for U_j ($1 \le j \le D$) and copula $C_{\text{res}}(\cdot; \boldsymbol{\theta}_{res})$ for residual dependence, is

$$\sum_{i=1}^{N} \log c_{\boldsymbol{U},V}(u_{i1}, \dots, u_{iD}, v_i; \boldsymbol{\theta}_D) = \sum_{i=1}^{N} \sum_{j=1}^{D} \log c_{j0}(u_{ij}, v_i; \boldsymbol{\theta}_j) + \sum_{i=1}^{N} \log c_{\text{res}}(C_{1|0}(u_{i1}|v_i; \boldsymbol{\theta}_1), \dots, C_{D|0}(u_{iD}|v_i; \boldsymbol{\theta}_D); \boldsymbol{\theta}_{\text{res}}). \quad (20)$$

There are two stages in the proxy estimation.

- Stage 1: Define the simple proxy random variable as

$$U_0 = P_D^U \Big(D^{-1} \sum_{j=1}^{D} U_j \Big), \quad (21)$$

  where P_D^U is the cdf of $\bar{U}_D := D^{-1} \sum_{j=1}^{D} U_j$. (21) has the version of proxy in [11]. For observation i, let $\bar{u}_i = D^{-1} \sum_{j=1}^{D} u_{ij}$ and let the sample proxy be $u_{i,0} = [\text{rank}(\bar{u}_i) - 0.5]/N$; $\text{rank}(\bar{u}_i)$ is defined as the rank of $\bar{u}_i$ based on $\bar{u}_1, \dots, \bar{u}_N$. Combining the bivariate scatterplots of proxy variables versus observed variables, empirical tail-weighted dependence measures and Akaike information criterion (AIC) values of a few candidate parametric copula families, one can decide on copula families C_{j0}. Substitute $v_i = u_{i,0}$ in the log-likelihood (20), and obtain the first-stage estimates of the parameters $\boldsymbol{\theta}$ from this as an approximate log-likelihood. Also, for a simple parametric model for weak residual dependence (such as for a Markov tree structure), $\boldsymbol{\theta}_{res}$ can be estimated.
- Stage 2: Construct the conditional expectation proxies based on (11) with the first-stage estimated parameters of $\boldsymbol{\theta}_D = \{\boldsymbol{\theta}_j, j = 1, \dots, D\}$. One-dimensional Gauss-Legendre quadrature can be used. Denote the improved proxies as $\widetilde{u}_{i,0}$ for $i \in \{1, \dots, N\}$. Substitute $v_i = \widetilde{u}_{i,0}$ in the log-likelihood (20) to obtain the second-stage estimates of the parameters.

7 Simulation studies

The theorems in previous sections provide intuitive conditions for weak residual dependence in factor copula models. Proofs emulating Theorems 3 and 4 would take too much space, so instead this section summarizes some representative simulation results for factor copulas to support and explain ideas in previous sections, and show the effects of weak residual dependence.

For the simulations, to concentrate on the behavior under weak residual dependence, we assume univariate margins are known or have been well estimated so that the probability integral transforms lead to a sample from a factor copula. 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; these are examples of sampling from some super-populations. A few different scenarios were assessed and some summaries are given in two subsections for the considered factor copula models.

7.1 1-factor copula with weak residual dependence

The sample size is N and there are D variables with bivariate linking copulas $C_{jV}(\cdot,\cdot;\boldsymbol{\theta}_j)$, $j \in \{1,\cdots,D\}$, to the latent variable. The parameters $\{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_D\}$ are uniformly sampled from $U(\theta_L,\theta_U)$, 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 parameters change for different generated random samples in different simulation runs. For simplicity, the weak residual dependence in simulated data sets assume the variable indices have been permuted so that a 1-truncated or 2-truncated D-vine structure can be used – for 1-truncated, there is weak conditional dependence for variables with neighboring indices, and for 2-truncated, there is additional weak dependence for variables indexed 2-apart. The parameters for weak residual dependence have Kendall tau values in the interval $(0.1,0.3)$.

The details of the three simulation settings are shown in Table 1. Setting 1 in Table 1 is designed to compare with the results presented in [4] in the 1-factor copula model. To further explore the influence of the weak residual dependence on parameter estimation and proxy estimation, additional settings are designed within Setting 2. The linking copulas in the truncated vine for residual dependence include Gaussian or Frank copulas. In Setting 3, the linking copulas from the observed to the latent variables comes from different families.

The simulation results for Settings 1, 2, and 3 are presented in Tables 2, 3 and 4, respectively. In Settings 1 and 2, the family linking copulas are assumed known and only copula parameters are estimated. In Setting 3, the parametric linking copula families are chosen from among a few candidates based on AIC. Similar to [4], the main summary is the Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) of the estimated parameters for the D linking copulas from the observed to the latent variables for two different methods, denoted as $m \in \{1,2\}$ in the equation

(23). Here, $m=1$ indicates the simple proxy approach (based on (21)) and $m=2$ represents the sequential proxy approach.

Table 1 Three simulation settings for the 1-factor copula model; in setting 3, the number of linking copulas in different families is approximately $D/3$, and the ν parameter of t copulas is fixed at 5. The parameters on the linking copulas from the observed to the latent variables are chosen with Kendall's tau in [0.4,0.8] while the parameters on the linking copulas in the residual dependence are chosen such that Kendall's tau is in $[0.1,0.3]$. For Setting 2.2, the residual dependence is modeled by a 2-truncated D-vine with Kendall's tau in the first level of the D-vine in $[0.2,0.3]$ and that in the second level of the D-vine in $[0.1,0.2]$.

	Linking Families	Residual Dependence Structure
setting 1	Frank	1-truncated D-vine (all Gaussian copulas)
setting 2.1	Gumbel	1-truncated D-vine (all Gaussian copulas)
setting 2.2	Gumbel	2-truncated D-vine (all Gaussian copulas)
setting 2.3	Gumbel	1-truncated D-vine (all Frank copulas)
setting 3	Gumbel, t, BB1	1-truncated D-vine (all Gaussian copulas)

	$[\boldsymbol{\theta}_L,\boldsymbol{\theta}_U]$ (Kendall's τ)	$[\boldsymbol{\theta}_L^{\text{res}},\boldsymbol{\theta}_U^{\text{res}}]$ (Kendall's τ)
setting 1	$[4.2,18.5]\,(0.4,0.8)$	$[0.15,0.45](0.1,0.3)$
setting 2.1	$[1.67,5]\,(0.4,0.8)$	$[0.15,0.45](0.1,0.3)$
setting 2.2	$[1.67,5]\,(0.4,0.8)$	$U(0.30,0.45)(0.2,0.3)$ $U(0.15,0.30)(0.1,0.2)$
setting 2.3	$[1.67,5]\,(0.4,0.8)$	$[0.95,2.95](0.1,0.3)$
setting 3	$[1.67,5],[0.59,0.95],$ $[0.5,1.30]\times[1.35,3.0]\,(0.4,0.8)$	$[0.15,0.45](0.1,0.3)$

Summaries from the simulation settings are the following based on N^* random samples of size N being generated.

a. The RMSEs of the proxies of the two methods are:

$$\hat{v}^m_{\text{RMSE}} = \left\{ \frac{1}{N^*N} \sum_{s=1}^{N^*} \sum_{i=1}^{N} \left(\hat{v}^m_{si} - v_{si}\right)^2 \right\}^{1/2}, \quad m \in \{1,2\}. \tag{22}$$

b. The MAE of the differences between the estimates obtained from the proxy methods and the true parameters is:

$$\hat{\theta}^m_{\text{MAE}} = \frac{1}{N^*D} \sum_{s=1}^{N^*} \sum_{j=1}^{D} |\hat{\theta}^m_{sj} - \theta_{sj}|,$$

where θ_{sj} is the parameter of C_{j0} generated at the sth simulation and $\hat{\theta}^m_{sj}$ is the corresponding estimate using method m.

c. The differences of Kendall's taus (functions of the estimated bivariate linking copula) of the estimates obtained from the proxy methods and the maximum likelihood with exact likelihood:

$$\hat{\theta}^m_{|\mathrm{diff}|} = \frac{1}{N^*D}\sum_{s=1}^{N^*}\sum_{j=1}^{D}|\hat{\theta}^m_{sj} - \hat{\theta}^0_{sj}|,\ \hat{\tau}^m_{|\mathrm{diff}|} = \frac{1}{N^*D}\sum_{s=1}^{N^*}\sum_{j=1}^{D}|\hat{\tau}^m_{sj} - \hat{\tau}^0_{sj}|,\ m \in \{1,2\}. \quad (23)$$

d. In Setting 3, the averaged differences of some dependence measures between the true and fitted models over the D bivariate linking copulas:

$$[\hat{M}^{\mathrm{diff}}]^m_{\mathrm{mean}} = \frac{1}{N^*D}\sum_{s=1}^{N^*}\sum_{j=1}^{D}|[\widehat{M}^{\mathrm{model}}]_j - [M^{\mathrm{true}}]_j|, \quad m \in \{1,2\}, \quad (24)$$

In Setting 3, summaries include the averaged differences of the dependence measures between the true and fitted models over the D bivariate linking copulas in (24). The three dependence measures are Kendall's tau and tail-weighted upper/lower tail ζ_α (with $\alpha = 20$), as defined in [12], to assess central dependence and tail properties.

From the results in Table 2 for Setting 1, the simple and sequential proxy approaches can provide accurate parameter estimates when D is large, and the sequential procedure performs better compared to the simple proxy, as it has more accurate estimates of the parameters and the realized latent variables. In the 1-factor model with weak residual dependence, the differences in the estimated parameters and those obtained from the exact likelihood are in a similar scale as those in the 1-factor copula model (without residual dependence) presented in [4]. However, the RMSEs of the estimated proxies are larger compared to those results in [4] with the assumption the conditional independence given the latent variables.

Table 2 1-factor copula models with all Frank linking copulas (Setting 1); simulation size $N^* = 1000$, sample size $N = 500$, $\boldsymbol{\theta}$ uniform in $(\boldsymbol{\theta}_L, \boldsymbol{\theta}_U)$ as specified in Table 1. Summaries from (23), (24), and (22) for 3 approaches — superscript $m = 0$: exact likelihood; superscript $m = 1$: simple proxy; superscript $m = 2$: stage 2 proxy.

D	$\hat{\theta}^0_{\mathrm{MAE}}$	$\hat{\theta}^1_{\mathrm{MAE}}$	$\hat{\theta}^2_{\mathrm{MAE}}$	$\hat{\theta}^1_{\lvert\mathrm{diff}\rvert}$	$\hat{\theta}^2_{\lvert\mathrm{diff}\rvert}$	$\lvert\hat{\tau}^{\mathrm{diff}}\rvert^1_{\mathrm{mean}}$	$\lvert\hat{\tau}^{\mathrm{diff}}\rvert^2_{\mathrm{mean}}$	$\hat{v}^{m=1}_{\mathrm{RMSE}}$	$\hat{v}^{m=2}_{\mathrm{RMSE}}$
20	0.456	0.598	0.646	0.477	0.469	0.012	0.009	0.051	0.037
40	0.437	0.483	0.488	0.298	0.217	0.007	0.004	0.039	0.027
60	0.429	0.461	0.458	0.234	0.163	0.005	0.003	0.033	0.023
80	0.418	0.443	0.437	0.203	0.140	0.004	0.003	0.030	0.020

From the results in Table 3 for Setting 2, comparing Settings 2.1 and 2.2, when the weak residual dependence becomes more complex (going from 1-truncated to 2-truncated for modeling the residual dependence), the RMSE of the estimated parameters (of linking copulas from the observed variables to the latent variable) of the two approaches is slightly larger, as well as the RMSE of the proxies for the two approaches. However, as the dimension increases, the sequential approach can still provide accurate parameter estimates and proxy estimates, even in cases where the conditional independence is farther from holding in Setting 2.2. Comparing Settings 2.3 and 2.1, the results are similar. This indicates that the sequential approach can be useful when the linking copulas for modeling the residual dependence are not Gaussian but belong to some other well-behaved copula families.

Table 3 Results for Setting 2: 1-factor copula models with all Gumbel linking copulas with weak residual dependence; simulation size $N^* = 1000$, sample size $N = 500$. The three stacked panels show the results for Settings 2.1, 2.2, and 2.3, respectively, from top to bottom. Summaries from (23), (24), and (22) for 2 approaches —superscript $m = 1$: simple proxy; superscript $m = 2$: stage 2 proxy. The differences in (23), (24) are between the estimated parameters and true parameters (instead of estimated parameters from the exact approach). The implementation of the exact maximum likelihood approach is limited to scenarios where the residual dependence is modeled using a 1-truncated vine.

D	$\hat{\theta}^1_{\text{RMSE}}$	$\hat{\theta}^2_{\text{RMSE}}$	$\hat{\theta}^1_{\lvert\text{diff}\rvert}$	$\hat{\theta}^2_{\lvert\text{diff}\rvert}$	$\lvert\hat{\tau}^{\text{diff}}\rvert^1_{\text{mean}}$	$\lvert\hat{\tau}^{\text{diff}}\rvert^2_{\text{mean}}$	$\hat{v}^{m=1}_{\text{RMSE}}$	$\hat{v}^{m=2}_{\text{RMSE}}$
20	0.225	0.218	0.167	0.165	0.015	0.014	0.054	0.042
40	0.203	0.149	0.148	0.112	0.013	0.011	0.041	0.030
60	0.192	0.136	0.139	0.103	0.012	0.010	0.035	0.024
80	0.188	0.135	0.138	0.103	0.012	0.010	0.032	0.022
20	0.237	0.308	0.187	0.246	0.019	0.021	0.067	0.053
40	0.194	0.176	0.143	0.134	0.013	0.012	0.051	0.039
60	0.190	0.146	0.138	0.111	0.012	0.011	0.043	0.032
80	0.182	0.137	0.133	0.104	0.012	0.010	0.039	0.028
20	0.235	0.233	0.175	0.177	0.017	0.015	0.057	0.044
40	0.201	0.152	0.146	0.114	0.013	0.011	0.043	0.032
60	0.194	0.138	0.141	0.105	0.012	0.010	0.036	0.026
80	0.192	0.135	0.139	0.102	0.012	0.010	0.033	0.023

Table 4 Results for Setting3: 1-factor copula model with linking copulas from Gumbel, t, and BB1 families; Simulation size $N^* = 1000$, sample size $N = 500$. In each simulation $\boldsymbol{\theta}$ is uniform in $(\boldsymbol{\theta}_L, \boldsymbol{\theta}_U)$. The summaries are for (24) and (22) and are shown in order $m = 1/m = 2$ respectively in columns 2 to 4. With linking copula families to be decided, $[\tau^{\text{diff}}]_{\text{mean}}$, $[\zeta^{\text{diff}}_{\alpha,U}]_{\text{mean}}$,$[\zeta^{\text{diff}}_{\alpha,L}]_{\text{mean}}$, are the averaged differences in the dependence measures between true and fitted models over D linking copulas. $\alpha = 20$ in $\zeta_{\alpha,U}$ and $\zeta_{\alpha,L}$ defined in [12] to summarize the dependence in the upper and lower joint tails.

D	$[\hat{\tau}^{\text{diff}}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,U}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,L}]^m_{\text{mean}}$	$\hat{v}^{m=1}_{\text{RMSE}}$	$\hat{v}^{m=2}_{\text{RMSE}}$
30	0.016/0.016	0.027/0.022	0.033/0.029	0.051	0.038
45	0.014/0.014	0.024/0.020	0.029/0.025	0.042	0.031
60	0.014/0.013	0.023/0.019	0.027/0.024	0.038	0.027
90	0.013/0.012	0.022/0.019	0.025/0.022	0.033	0.022

Setting 3 indicates that the sequential approach can identify linking copula families with similar tail behaviors to the correct ones in most cases, based on the small differences in upper/lower tail-weighted dependence measures and Kendall's tau when the dimension is large. Comparing the results in Table 3 to those results in [4], when the conditional independence assumption is slightly violated, the RMSE of the proxies is slightly larger, while the differences in the dependence measures are on a similar scale.

7.2 *Bi-factor copula with weak residual dependence*

A simulation setting is summarized in Table 5 to illustrate the sequential approach in bi-factor copula models. 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 $D/3$. 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. As 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. The residual dependence is assumed to be a 1-truncated D-vine structure, and the parameters on the linking copulas are uniformly from $(\boldsymbol{\theta}_L^{\text{res}}, \boldsymbol{\theta}_U^{\text{res}})$ such that the residual dependence is weak. The simulation size is $N^* = 1000$.

Table 5 The simulation setting for the bi-factor copula model with weak residual dependence; the linking families are for global and group; weak residual dependence is modeled by a 1-truncated D-vine. The range of Kendall's tau corresponding to the range of parameters is included after the parameter interval.

	linking families	$[\boldsymbol{\theta}_L, \boldsymbol{\theta}_U](\tau)$	$[\boldsymbol{\theta}_{LL}, \boldsymbol{\theta}_{UU}](\tau)$	$[\boldsymbol{\theta}_L^{\text{res}}, \boldsymbol{\theta}_U^{\text{res}}](\tau)$
setting 1	Frank/Frank/Gaussian	[1.87,8] (0.2,0.6)	[4.2,11.5] (0.4,0.7)	[0.1,0.3]

Summaries are given in Table 6. When the bi-factor models slightly deviate from the conditional independence assumption, the MAEs of the estimated parameters on the global linking copulas are slightly larger, but those and MAEs of the estimated parameters on the local linking copulas are on the same scale as the previous results in bi-factor models (with conditional independence given latent variables) shown in [4]. With proxies, the RMSEs for global latent variables are slightly smaller, and the RMSEs for local latent variables are larger than those in bi-factor models with conditional independence. The estimation improves as D increases.

Table 6 Bi-factor copula model with all linking copulas in the Frank family; Simulation size $N^* = 1000$, sample size $N = 1200$, $d_g \in \{10, 20, 30\}$ for $g \in \{1, 2, 3\}$. The parameters are generated as specified in Table 5. 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 likelihood; superscript $m = 2$: stage 2 proxy are shown; For MAE, the results are shown for m=0/m=2 respectively.

D	Global linking copulas			Local linking copulas			RMSE$_{\text{proxy}}$
Frank	$\hat{\theta}^{m=0/m=2}_{\text{glob:MAE}}$	$\hat{\theta}^{m=2}_{\text{glob:}\lvert\text{diff}\rvert}$	$\hat{\tau}^{m=2}_{\text{glob:}\lvert\text{diff}\rvert}$	$\hat{\theta}^{m=0/m=2}_{\text{loc:MAE}}$	$\hat{\theta}^{m=2}_{\text{loc:}\lvert\text{diff}\rvert}$	$\hat{\tau}^{m=2}_{\text{loc:}\lvert\text{diff}\rvert}$	$\hat{v}_0/\hat{v}_g$
30	0.240/0.458	0.375	0.013	0.333/1.027	0.812	0.038	0.057/0.151
60	0.215/0.351	0.277	0.009	0.236/0.447	0.320	0.017	0.044/0.111
90	0.210/0.325	0.257	0.008	0.212/0.302	0.200	0.012	0.038/0.096

8 Conclusion

The proxy variable estimates of latent variables in factor models in [4] are defined based on the assumption of conditional independence given latent variables. In this article, some interpretable sufficient conditions on weak residual (conditional) dependence are obtained so that the proxies are still consistent when the conditional independence assumption is slightly violated.

For high-dimensional factor copula models with a large sample size, simulation studies show that the sequential estimation approach is robust when there is weak residual dependence. It is possible to efficiently estimate the latent variables, select the families of parametric linking copulas, and estimate the copula parameters. As shown in [4], a latent variable can be well estimated when the cumulative dependence of observed variables to a latent variable is strong enough; more variables would be needed if the individual dependencies with the latent variable are weaker. Small amounts of mis-specification are acceptable either in the copula families linking to the latent variables or weak residual dependence given the latent variables.

Results on consistency of factor scores for 1-factor and bi-factor Gaussian models with weak residual dependence are useful results on their own, but our motivation is to get an idea of weak residual dependence conditions that could apply for high-dimensional factor copula models.

For a large number of variables that can be divided into G groups with stronger within-group dependence, several parsimonious dependence structures given below can be considered that make use of the correlation matrix of normal scores of the observed variables and proxy variables. These are applied in [3] for data sets with daily stock returns, changes in daily currency exchange rates and gene expressions, and will be used in subsequent research.

- An oblique factor model with weak residual dependence, where there is a 1-factor model with weak residual dependence for each of G groups, and the G latent variables are dependent to explain the between-group dependence.

- A bi-factor structure with weak residual dependence, where there is global latent variable and G group-based latent variables that are mutually independent, and there is weak residual dependence of variables in each group conditioned on the global latent variable and the corresponding group-based latent variable.
- A parsimonious dependence structure constructed from applying truncated vine algorithms to the combination of observed and proxy variables, with one latent variable for each group.

Each of the above leads to a truncated vine structure (rooted at latent or proxy variables). With initially estimated proxy variables based on the Gaussian copula then one can check if there can be improvements to bivariate linking Gaussian copulas between each observed variable and the corresponding proxy variables.

Finally, the techniques of proofs in Section 9 are useful for studying asymptotic properties of estimators with slightly mis-specified models.

9 Proofs

9.1 Proof of Theorem 1

The short proof is modified from [4] in order to highlight the steps for the longer proof of Theorem 2.

Proof. Let $\boldsymbol{e}_D = (e_1, \dots, e_D)$ be one realization of $\boldsymbol{\varepsilon}_D$. Let $q_D = \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2} \boldsymbol{A}_D > 0$. Then, from (8),

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

Note that $Y = D^{-1/2} \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1} \boldsymbol{e}_D$ is a realization of $D^{-1/2} \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1} \boldsymbol{\varepsilon}_D$ which can be written as $D^{-1/2} \sum_{j=1}^D \alpha_j \varepsilon_j / \psi_j$. The latter random quantity has a variance $D^{-1} \sum_{j=1}^D \sum_{k=1}^D \alpha_j \alpha_k \gamma_{jk} / [\psi_j \psi_k]$. By Assumption 1 and with loadings that are bounded away from ± 1, this variance is $O(1)$ so that Y can be considered as $O_p(1)$. With the boundedness assumptions for the α_j's, $\bar{q}_D = D^{-1} q_D = O(1)$. Then $\widetilde{w}_D - w^0 = D^{-1/2} (D^{-1} + \bar{q}_D)^{-1} (D^{-1/2} \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1} \boldsymbol{\varepsilon}_D) + O(D^{-1})$ is asymptotically $O_p(D^{-1/2})$. □

9.2 Proof of Theorem 2

Proof. Let $\boldsymbol{Q_D} = \boldsymbol{A_D}^\top \boldsymbol{\Psi_D}^{-2} \boldsymbol{A_D}$. With $p = G + 1$, (8) extends to

$$\widetilde{w}_{\boldsymbol{D}} - \boldsymbol{w}^0 = (\boldsymbol{I}_p + \boldsymbol{Q_D})^{-1} \boldsymbol{A_D}^\top \boldsymbol{\Psi_D}^{-1} \boldsymbol{e_D} + \left[(\boldsymbol{I}_p + \boldsymbol{Q_D}^{-1})^{-1} \boldsymbol{w}^0 - \boldsymbol{w}^0 \right]. \tag{25}$$

The proof is outlined in two steps. For **step1**, the second term in the right side of (25) is shown to be $O(D^{-1})$. For **step2**, the first term in the right side of (25) is shown to be $O_p(D^{-1/2})$.

step1: Since $\boldsymbol{A_D}$ is of full rank and the entries of $\boldsymbol{\Psi_D}$ are uniformly bounded away from 0. Then $\overline{\boldsymbol{Q}}_{\boldsymbol{D}} = D^{-1}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{A_D}$ is positive definite for any fixed D and $\overline{\boldsymbol{Q}}_D = O(1)$. For the second term in the right side of (25), by first-order Taylor expansion, $(\boldsymbol{I}_p+\boldsymbol{Q}_{\boldsymbol{D}}^{-1})^{-1}-\boldsymbol{I}_p = (\boldsymbol{I}_p+D^{-1}\overline{\boldsymbol{Q}}_D^{-1})^{-1}-\boldsymbol{I}_p = \boldsymbol{I}_p - D^{-1}\overline{\boldsymbol{Q}}_D + O(D^{-2}) - \boldsymbol{I}_p = O(D^{-1})$.

step2: The first term in the right side of (25) can be rewritten as

$$D^{-1/2}(D^{-1}\boldsymbol{I}_p+\overline{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}D^{-1/2}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{e_D}. \tag{26}$$

$D^{-1/2}\boldsymbol{A}_D^{\top}\boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D$ is a realization of $D^{-1/2}\boldsymbol{A}_D^{\top}\boldsymbol{\Psi}_D^{-1}\boldsymbol{\varepsilon}_D$. The latter quantity will be shown to be $O_p(1)$. It is sufficient to justify that the covariance matrix of this quantity is bounded. Since $\boldsymbol{\varepsilon_D} \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{\Omega_D})$, the covariance matrix of $D^{-1/2}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{\varepsilon_D}$ is

$$D^{-1}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{\Omega_D}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{A_D} := D^{-1}\boldsymbol{B},$$

where $\boldsymbol{B}$ is $p\times p$. Let $\boldsymbol{1}$ be the $p\times 1$ vector with all elements equal to 1. Then, the sum of the elements in the $p\times p$ covariance matrix can be written as

$$\begin{aligned}\text{trace}[D^{-1}\boldsymbol{1}^{\top}\boldsymbol{B}\boldsymbol{1}] &= \text{trace}[D^{-1}\boldsymbol{1}^{\top}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{\Omega_D}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{A_D}\boldsymbol{1}] \\ &= \text{trace}[D^{-1}\boldsymbol{\Omega_D}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{A_D}\boldsymbol{1}\boldsymbol{1}^{\top}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}] = \text{trace}[D^{-1}\boldsymbol{\Omega_D}\boldsymbol{K_D}],\end{aligned}$$

where $\boldsymbol{K}_D = \boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{A_D}\boldsymbol{1}\boldsymbol{1}^{\top}\boldsymbol{A}_{\boldsymbol{D}}^{\top}\boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}$. Let $\boldsymbol{\alpha}_i^{\top}$ to represent the ith row of $\boldsymbol{A}_D$ so that $\boldsymbol{A}_D = [\boldsymbol{\alpha}_1^{\top},\ldots,\boldsymbol{\alpha}_D^{\top}]^{\top}$. Then $[\boldsymbol{K}_D]_{ij} = \psi_i^{-1}\psi_j^{-1}\boldsymbol{\alpha}_i^{\top}\boldsymbol{1}\boldsymbol{1}^{\top}\boldsymbol{\alpha}_j$.

With the assumptions on the loading parameters, there are constants $0 < M_1 < M_2 < 1$ such that $M_1 < \psi_j < M_2$ and $\|\boldsymbol{\alpha}_j\|_2 < 1$ for $j = 1,\ldots,D$. For the absolute value of the (i,j) entry (i can equal j) in $\boldsymbol{K_D}$, $|k_{ij}| = |\psi_i^{-1}\psi_j^{-1}\boldsymbol{\alpha}_i^{\top}\boldsymbol{1}\boldsymbol{1}^{\top}\boldsymbol{\alpha}_j| \le \psi_i^{-1}\psi_j^{-1}\|\boldsymbol{\alpha}_i\|_1\|\boldsymbol{\alpha}_j\|_1 \le \psi_i^{-1}\psi_j^{-1}\times p\|\boldsymbol{\alpha}_i\|_2\|\boldsymbol{\alpha}_j\|_2 < pM_1^{-2}$, by the Cauchy-Schwarz inequality. That is, the elements in $\boldsymbol{K_D}$ are uniformly bounded.

The summation of absolute values of all elements in the covariance matrix $D^{-1}\boldsymbol{B}$ is:

$$\left|\text{trace}[D^{-1}\boldsymbol{\Omega_D}\boldsymbol{K_D}]\right| = D^{-1}\sum_{s=1}^{D}\sum_{t=1}^{D}|\omega_{st}k_{st}| \le pM_1^{-2}D^{-1}\sum_{s=1}^{D}\sum_{t=1}^{D}|\omega_{st}|.$$

Based on the Assumption 2 on the order of residual dependence, the summation of absolute values of entries in the covariance matrix $D^{-1}\boldsymbol{B}_{p\times p}$ is of $O(1)$. Therefore, $D^{-1/2}\boldsymbol{A}_D^{\top}\boldsymbol{\Psi}_D^{-1}\boldsymbol{\varepsilon}_D$ is $O_p(1)$. Since $D^{-1/2}\boldsymbol{A}_D^{\top}\boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D$ is a realization of $D^{-1/2}\boldsymbol{A}_D^{\top}\boldsymbol{\Psi}_D^{-1}\boldsymbol{\varepsilon}_D$, so it is also $O_p(1)$. Since $(D^{-1}\boldsymbol{I}_p+\overline{\boldsymbol{Q}}_D)$ in (26) is $O(1)$ as $D\to\infty$, then $\widetilde{w}_D - \boldsymbol{w}^0 = O_p(D^{-1/2}) + O(D^{-1}) = O_p(D^{-1/2})$. □

9.3 Proof of Theorem 3

Proof. For the model defined in equation (5), write the averaged negative log-likelihood as the summation of two parts:

$$\overline{L}_D(v,\boldsymbol{u}_D) := \underbrace{-\frac{1}{D}\sum_{j=1}^{D}\log c_{j0}(u_j,v)}_{:=\overline{qL}_D(v,u_D)} \underbrace{-\frac{1}{D}\log c_{\text{res}}(C_{1|0}(u_1|v),\ldots,C_{D|0}(\boldsymbol{u}_D|v))}_{:=\overline{L}_D^{(r)}(v,\boldsymbol{u}_D)},$$

where $\overline{L}_D^{(r)}$ comes from the residual dependence given the latent variables and $\overline{qL}_D$ is the averaged negative quasi-log-likelihood based on conditional independence given the latent variables.

Step A: As mentioned in Remark 3, consider v as the parameter in the log-likelihood function $\overline{L}_D$. According to the general theory of MLE under standard regularity conditions in [13], and mixing conditions such as [2], the negative log-likelihood $\overline{L}_D$ has a minimizer located at the true parameters v^0 as $D \to \infty$.

Step B in Remark 3 is verified under the Assumptions 3 and Condition 1. It will be shown that with the assumptions that

$$\left.\frac{\partial \overline{L}_D^{(r)}}{\partial v}\right|_{v=v^0} \to_p 0, \quad D \to \infty.$$

Define random variables $\widetilde{U}_j(v) = C_{j|0}(U_j|v)$, and $\widetilde{U}_j^0 = C_{j|0}(U_j|v^0)$ with realized value v^0. Then $\widetilde{U}_j^0 \sim U(0,1)$. Let $\widetilde{u}_j(v) = C_{j|0}(u_j|v)$ and $\widetilde{u}_j^0 = \widetilde{u}_j(v^0)$ be realized values of $\widetilde{U}_j(v)$ and $\widetilde{U}_j^0$ respectively. Let $E_j^0 := \Phi^{-1}(\widetilde{U}_j^0) \sim \mathcal{N}(0,1)$ and let one realization be $e_j^0 = \Phi^{-1}(\widetilde{u}_j^0)$.

Denote the derivative of $\Phi^{-1}(\widetilde{u}_j(v))$ with respect to v evaluated at v^0 as m_j^0, so that

$$\begin{aligned} m_j^0 &= \left.\frac{\partial \Phi^{-1}(C_{j|0}(u_j|v))}{\partial v}\right|_{v=v^0} = \frac{1}{\phi(\Phi^{-1}(\widetilde{u}_j^0))}\left.\frac{\partial C_{j|0}(u_j|v)}{\partial v}\right|_{v=v^0} \\ &= \frac{1}{\phi(e_j^0)}\left.\frac{\partial \widetilde{u}_j(v)}{\partial v}\right|_{v=v^0}. \end{aligned}$$

Let $m_j(u,v) = \partial \Phi^{-1}(C_{j|0}(u_j|v))/\partial v$. With Assumption 3 and also Assumption 1 in [4], $\lim_{u\to 1} m_j(u,v)$ and $\lim_{u\to 0} m_j(u,v)$ are finite for $v_A < v < v_B$, and $|m_j(u,v)|$ is uniformly bounded over j, $v_A \le v \le v_B$ and $0 \le u \le 1$. Furthermore, from the stochastic increasing condition with $1 - C_{j|0}(\cdot|v)$ increasing in v for all j, then $\partial C_{j|0}(u_j|v)/\partial v \le 0$ and $m_j^0 \le 0$ for all j. Let B_m be an upper bound on $-m_j(u,v)$ over all j, all $0 < u < 1$, and $v_A < v < v_B$.

Write the derivative of $\overline{L}^{(r)}(v,\boldsymbol{u}_D)$ with respect to v evaluated at v^0 as

$$\frac{\partial \overline{L}_D^{(r)}(v,\boldsymbol{u}_D)}{\partial v}\Big|_{v=v^0} = D^{-1}\frac{\partial \log c_{\text{res}}(\widetilde{u}_1(v),\ldots,\widetilde{u}_D(v))}{\partial v}\Big|_{v=v^0}. \tag{27}$$

From Condition 1, then

$$c_{\text{res}}(\widetilde{u}_1(v),\ldots,\widetilde{u}_D(v)) = \frac{\phi_D\big(\boldsymbol{\Phi}^{-1}(\widetilde{u}_1(v)),\ldots,\boldsymbol{\Phi}^{-1}(\widetilde{u}_D(v));\boldsymbol{\Gamma}_D\big)}{\prod_{j=1}^D \phi(\boldsymbol{\Phi}^{-1}(\widetilde{u}_j(v))},$$

where ϕ_D is the D-variate Gaussian density, and ϕ, Φ are respectively the univariate standard normal densiy and cdf. Therefore,

$$\begin{aligned}
&D^{-1}\partial \log c_{\text{res}}(\widetilde{u}_1(v),\ldots,\widetilde{u}_D(v);\boldsymbol{\theta}^{(r)}) \;/\; \partial v \,\big|_{v=v^0} \\
&= D^{-1}\partial \log \phi_d\big(\boldsymbol{\Phi}^{-1}(\widetilde{u}_1(v)),\ldots,\boldsymbol{\Phi}^{-1}(\widetilde{u}_D(v));\boldsymbol{\Gamma}_D\big) \;/\; \partial v \,\big|_{v=v^0} - \\
&D^{-1}\sum_{j=1}^D \partial \log \phi\big(\boldsymbol{\Phi}^{-1}(\widetilde{u}_j(v))\big) \;/\; \partial v \,\big|_{v=v^0} \\
&= D^{-1}(\boldsymbol{m}_D^0)^\top \boldsymbol{\Gamma}_D^{-1}\boldsymbol{e}_D^0 - D^{-1}\sum_{j=1}^D \frac{\phi'(e_j^0)}{\phi(e_j^0)} m_j^0 &(28)\\
&= D^{-1}(\boldsymbol{m}_D^0)^\top (\boldsymbol{\Gamma}_D^{-1} - \boldsymbol{I}_D)\boldsymbol{e}_D^0, &(29)
\end{aligned}$$

where $\boldsymbol{m}_D^0 = (m_1^0,\ldots,m_D^0)^\top$ and $\boldsymbol{e}_D^0 = (e_1^0,\ldots,e_D^0)^\top$. In the last term of (28), $\phi'(x) = -x\phi(x)$, so that it simplifies to $-D^{-1}\sum_{j=1}^D e_j^0 m_j^0 = -D^{-1}(\boldsymbol{m}_D^0)^\top \boldsymbol{I}_D \boldsymbol{e}^0$.

It will proved in Lemma 1 below that $\{D^{-1}(\boldsymbol{m}_D^0)^\top(\boldsymbol{\Gamma}_D^{-1} - \boldsymbol{I}_D)\boldsymbol{e}_D^0\}$ is a realization of a random sequence whose variance converges to 0 as $D \to \infty$. Then, combining (27) and (29), $\partial \overline{L}_D^{(r)}(v,\boldsymbol{u}_D)/\partial v \,\big|_{v=v^0} \to 0$ as $D \to \infty$. Hence Step B is justified.

As noted in Remark 3 after step B, the quasi-MLE v_D^* which is the root $\partial \overline{qL}_D(v,u_D)/\partial v = 0$ is such that $v_D^* - v^0 = o_p(1)$ as $D \to \infty$ with a Taylor expansion. Also, due to the asymptotic equivalence of the proxy $\widetilde{v}_D$ and quasi-MLE v_D^* based on the Laplace approximation in [4], $\widetilde{v}_D - v_D^* = O(D^{-1})$. The consistency of the proxy variable is still satisfied since $\widetilde{v}_D - v^0 = (\widetilde{v}_D - v_D^*) + (v_D^* - v^0) = O(D^{-1}) + o_p(1) = o_p(1)$ as $D \to \infty$. □

Lemma 1. $\{D^{-1}(\boldsymbol{m}_D^0)^\top(\boldsymbol{\Gamma}_D^{-1} - \boldsymbol{I}_D)\boldsymbol{e}_D^0\}$ *is a realization of a random sequence whose variance converges to 0 as* $D \to \infty$*.*

Proof. The subscript D is suppressed for $\boldsymbol{\Gamma}, \boldsymbol{m}^0, \boldsymbol{e}^0$ etc. Let $\boldsymbol{M} = (M_j)$ and $\boldsymbol{E}^0 = (E_j^0)$ be the random counterparts of $\boldsymbol{m}^0$ and $\boldsymbol{e}^0$ in (29). The component e_j^0 is the realization of $E_j^0 \sim \mathcal{N}(0,1)$ and that $\boldsymbol{E}^0 \sim \mathcal{N}_D(\boldsymbol{0},\boldsymbol{\Gamma})$. M_j is a function of a random variable with distribution $C_{j|0}(\cdot|v^0)$, and is correlated with E_j^0. With $v_A < v_0 < v_B$, and the assumptions of Theorem 3, there is a positive constant B_m such that $0 \le -M_j \le B_m$.

Let $\boldsymbol{Y} := D^{-1}\boldsymbol{M}^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0$ be the random version of $D^{-1}(\boldsymbol{m}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{e}^0$. Let $\boldsymbol{X}^\top = \boldsymbol{M}^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})$ be a random variable so that $\boldsymbol{X} \in \mathbb{R}^D$, then $\boldsymbol{Y} = D^{-1}\boldsymbol{X}^\top\boldsymbol{E}^0$. It will be shown that $\mathrm{Var}(\boldsymbol{Y}) = D^{-2}\,\mathrm{Var}(\boldsymbol{X}^\top\boldsymbol{E}^0) \to 0$ when $D \to \infty$. Write the variance of $\boldsymbol{Y}$ as

$$D^{-2}\,\mathrm{Var}(\boldsymbol{X}^\top \boldsymbol{E}^0) = D^{-2}\mathrm{E}([\boldsymbol{X}^\top \boldsymbol{E}^0]^2) - [D^{-1}\mathrm{E}(\boldsymbol{X}^\top \boldsymbol{E}^0)]^2. \tag{30}$$

In (30), the expectation in the second term can be rewritten as

$$\begin{aligned} D^{-1}\mathrm{E}(\boldsymbol{X}^\top \boldsymbol{E}^0) &= D^{-1}\mathrm{E}[\boldsymbol{M}^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0] \\ &= D^{-1}\mathrm{E}\big[\mathrm{E}(\boldsymbol{M}^\top|\boldsymbol{E}^0)(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0\big]. \end{aligned} \tag{31}$$

Let $\boldsymbol{K} = \mathrm{E}[\boldsymbol{M}^\top|\boldsymbol{E}^0](\boldsymbol{\Gamma}^{-1} - \boldsymbol{I}) \in \mathbb{R}^D$, so that the ith entry of the vector is $K_i = \sum_{\ell=1}^{D}\mathrm{E}[M_\ell|\boldsymbol{E}^0]\gamma^{k\ell} - \mathrm{E}[M_i|\boldsymbol{E}^0] = O_p(1)$ with Condition 1, Assumption 3 on weak residual dependence and uniformly bounded $\{M_\ell\}$. Thus, $\boldsymbol{K}$ is a realization of an $O_p(1)$ random vector with uniformly bounded entries. Then, based on the law of large numbers,

$$D^{-1}\sum_{j=1}^{D} K_j E_j^0 \to_p 0. \tag{32}$$

Based on the dominating convergence theorem, the limit in (31) is 0.

Without D^{-2}, the first term in the right side of (30) can be rewritten as

$$\begin{aligned} \mathrm{E}[(\boldsymbol{X}^\top \boldsymbol{E}^0)^2] &= \mathrm{E}\big[\boldsymbol{M}^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{M}\big] \\ &= \mathrm{E}\big[\mathrm{trace}\{(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{M}\boldsymbol{M}^\top\}\big] \\ &= \mathrm{E}\big[\mathrm{trace}\{(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\,\mathrm{E}(\boldsymbol{M}\boldsymbol{M}^\top|\boldsymbol{E}^0)\}\big]. \end{aligned} \tag{33}$$

The entries of $\mathrm{E}(\boldsymbol{M}\boldsymbol{M}^\top|\boldsymbol{E}^0)$ are positive and uniformly bounded by B_m^2. Let $\boldsymbol{A} = (\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I}) = (A_{ks})$. Then from (33),

$$D^{-2}\,\mathrm{trace}\{(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\,\mathrm{E}(\boldsymbol{M}\boldsymbol{M}^\top|\boldsymbol{E}^0) \le D^{-2}B_m^2\sum_{k=1}^{D}\sum_{s=1}^{D}|A_{ks}|.$$

Let $\boldsymbol{\Xi} = (\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I}) = \boldsymbol{\Gamma}^{-1} + \boldsymbol{\Gamma} - 2\boldsymbol{I} = (\xi_{ks})$. Then

$$\begin{aligned} D^{-2}\mathbf{1}^\top\boldsymbol{\Xi}\mathbf{1} &= D^{-2}\sum_{k=1}^{D}\sum_{s=1}^{D}\xi_{ks} \\ &\le D^{-2}\sum_{k=1}^{D}\sum_{s=1}^{D}\big\{|\gamma^{ks} + \gamma_{ks} - 2I(k=s)|\big\} = O(D^{-1}), \end{aligned}$$

with Condition (1). Note that $D^{-2}\boldsymbol{A}$ is a random matrix with expectation $D^{-2}\boldsymbol{\Xi}$. Since $\boldsymbol{E}^0$ is a multivariate normal random vector, as $D \to \infty$,

$$D^{-2}\{(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\boldsymbol{E}^0(\boldsymbol{E}^0)^\top(\boldsymbol{\Gamma}^{-1} - \boldsymbol{I})\} - D^{-2}\boldsymbol{\Xi} \to_p 0. \tag{34}$$

Hence,

$$\lim_{D\to\infty} D^{-2}\mathrm{E}([\boldsymbol{X}^\top \boldsymbol{E}^0]^2) \to 0. \tag{35}$$

Combining (32) and (35),

$$D^{-2}\,\mathrm{Var}(\boldsymbol{X}^\top \boldsymbol{E}^0) = D^{-2}\mathrm{E}([\boldsymbol{X}^\top \boldsymbol{E}^0]^2) - [D^{-1}\mathrm{E}(\boldsymbol{X}^\top \boldsymbol{E}^0)]^2 \to 0.$$

□

9.4 Proof of Theorem 4

Proof. The objective KL divergence function is defined as

$$\begin{aligned}
&D_{\mathrm{KL}}(f_D, g_D) \overset{\boldsymbol{Z}_D \sim g_D}{=} \int \log[g_D(z)/f_D(z)] \cdot g_D(z)\,\mathrm{d}z \\
&= \mathrm{E}\Big\{ -\tfrac{1}{2}\log|\boldsymbol{R}_D^0| - \tfrac{1}{2}\boldsymbol{Z}_D^\top \boldsymbol{R}_D^{-1}\boldsymbol{Z}_D + \tfrac{1}{2}\log|\boldsymbol{\Sigma}_D(\boldsymbol{\beta}_D)| + \tfrac{1}{2}\boldsymbol{Z}_D^\top \boldsymbol{\Sigma}_D^{-1}(\boldsymbol{\beta}_D)\boldsymbol{Z}_D \Big\} \\
&= -\tfrac{1}{2}\log|\boldsymbol{R}_D^0| + \tfrac{1}{2}\log|\boldsymbol{\Sigma}_D(\boldsymbol{\beta}_D)| - \tfrac{1}{2}D + \tfrac{1}{2}\,\mathrm{trace}[\boldsymbol{\Sigma}_D^{-1}(\boldsymbol{\beta}_D)\boldsymbol{R}_D^0] =: D_{\mathrm{KL}}(\boldsymbol{\beta}_D),
\end{aligned}$$

and the minimizer is defined as

$$\boldsymbol{\beta}_D^* = \arg\min_{\boldsymbol{\beta}} D_{\mathrm{KL}}(\boldsymbol{\beta}_D).$$

Hence $\boldsymbol{\beta}_D^*$ minimizes the objective function

$$\log|\boldsymbol{\Sigma}_D(\boldsymbol{\beta}_D)| + \mathrm{trace}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_D)\boldsymbol{R}_D^0]. \tag{36}$$

Let the score vector (gradient of objective function (36) with respect to $\boldsymbol{\beta}_D$ be $\boldsymbol{s}_D(\boldsymbol{\beta}_D)$ and let the Hessian matrix be $\boldsymbol{H}_D(\boldsymbol{\beta}_D)$. Let the jth element of $\boldsymbol{s}_D(\boldsymbol{\beta}_D)$ be $s_j(\boldsymbol{\beta}_D)$. Based on the mean-value theorem,

$$s_D(\boldsymbol{\beta}_D^*) = s_D(\boldsymbol{\alpha}_D) + \boldsymbol{H}_D(\bar{\boldsymbol{\alpha}}_D)(\boldsymbol{\beta}_D^* - \boldsymbol{\alpha}_D), \tag{37}$$

where $\bar{\boldsymbol{\alpha}}_D$ is a vector between $\boldsymbol{\alpha}_D$ and $\boldsymbol{\beta}_D^*$. Since the estimator $\boldsymbol{\beta}_D^*$ is the solution of a system of inference (or score) functions, $\boldsymbol{s}_D(\boldsymbol{\beta}_D^*) = \boldsymbol{0}$. Thus,

$$\boldsymbol{\beta}_D^* - \boldsymbol{\alpha}_D = -\boldsymbol{H}_D^{-1}(\bar{\boldsymbol{\alpha}}_D)\boldsymbol{s}_D(\boldsymbol{\alpha}_D). \tag{38}$$

All rates below are as $D \to \infty$ and are under the assumption that the α_j's are uniformly bounded away from -1 and 1. It will be proved in Lemma 3 that (I) the elements of the score vector $\boldsymbol{s}(\boldsymbol{\beta}_D)$ evaluated at the true parameters $\boldsymbol{\alpha}_D$ are uniformly $O(D^{-1})$, and in Lemma 4 that (II) the diagonal entries in the Hessian matrix $\boldsymbol{H}_D(\boldsymbol{\beta}_D)$ are uniformly $O(1)$ and the off-diagonal entries are uniformly $O(D^{-1})$.

Next, from Lemma 2, $\boldsymbol{H}_D^{-1}(\bar{\boldsymbol{\alpha}}_D)$ has uniformly bounded diagonal elements and off-diagonal elements that are uniformly $O(D^{-1})$. Let $\beta_{jD}^* - \alpha_{jD}$ be an element of the left side of (38). The properties of the score vector and inverse Hessian imply that $\boldsymbol{\beta}^*_{jD} - \boldsymbol{\alpha}_{jD} = O(D^{-1}) + D \cdot O(D^{-2}) = O(D^{-1})$ uniformly over j. □

Lemma 2. *Let $\boldsymbol{H}_D$ be a non-singular Hessian matrix with diagonal entries that are uniformly bounded and off-diagonal entries that are $O(D^{-1})$. Then the inverse Hessian matrix has the same properties and is diagonally dominant.*

Proof. Write $\boldsymbol{H}_D = \boldsymbol{H}_{0D} + D^{-1}\boldsymbol{\delta}_D$ where $\boldsymbol{H}_{0D} = \mathrm{diag}(\boldsymbol{H}_D)$ and $\boldsymbol{\delta}_D$ is a $D \times D$ matrix that is uniformly $O(1)$. Then

$$\begin{aligned}\boldsymbol{H}_D^{-1} &= (\boldsymbol{H}_{0D} + D^{-1}\boldsymbol{\delta}_D)^{-1} = [(\boldsymbol{I}_D + D^{-1}\boldsymbol{\delta}_D\boldsymbol{H}_{0D}^{-1})\boldsymbol{H}_{0D}]^{-1} \\ &= \boldsymbol{H}_{0D}^{-1}\big(\boldsymbol{I}_D + D^{-1}\boldsymbol{\delta}_D\boldsymbol{H}_{0D}^{-1}\big)^{-1} \\ &= \boldsymbol{H}_{0D}^{-1}\big\{\boldsymbol{I}_D - D^{-1}\boldsymbol{\delta}_D\boldsymbol{H}_{0D}^{-1} + D^{-2}[\boldsymbol{\delta}_D\boldsymbol{H}_{0D}^{-1}]^2 + \dots\big\}.\end{aligned}$$

Since $\boldsymbol{H}_{0D}^{-1}$ is diagonal and has all entries are $O(1)$, then $\boldsymbol{\delta}_D^{(-j)}\boldsymbol{H}_{0D}^{-1}$ is uniformly bounded. Hence, in the Taylor expansion, the second-order term is $O(D^{-1})$ and the third order is $O(D^{-2})$, etc. This proves that the inverse Hessian matrix is diagonally dominant and the off-diagonal elements are uniformly $O(D^{-1})$. □

Below, we sequentially show the results (I) and (II) in the proof of Theorem 4. The ideas in the proofs are straightforward, but there are many calculations of partial derivatives and determining orders of different terms as $D \to \infty$. All the presented orders in D are under the assumption that α_j's are uniformly bounded away from ± 1.

Lemma 3. *, The elements of the score vector $\boldsymbol{s}_D(\boldsymbol{\beta}_D)$ in* (37)*, evaluated at the true parameters $\boldsymbol{\alpha}_D$, are uniformly $O(D^{-1})$.*

Proof. For simpler notation, the subscript D will be suppressed for most variables.

An outline is summarized below in a few steps and substeps:

Step A: compute the expression of score vector below shown in (42);

Step B: evaluate the order of score vector at the true parameters $\boldsymbol{\alpha}_D$.

Step A:

Consider the jth element of the gradient of (36).

From Section 15.8 of [5], for a covariance matrix $\boldsymbol{\Sigma}$ parametrized by $\boldsymbol{\beta}$,

$$\frac{\log|\boldsymbol{\Sigma}|}{\partial\beta_j} = \mathrm{trace}\Big(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\beta_j}\Big), \quad \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\beta} = -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\beta_j}\boldsymbol{\Sigma}^{-1}.$$

Thus, the gradient of (36) with respect to $\boldsymbol{\beta}_j$ is

$$\mathrm{trace}\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\beta_j}\Big] - \mathrm{trace}\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\beta_j}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_D^0\Big]. \tag{39}$$

Next, we simplify each term in (39). For the first term, the derivative of $\partial\boldsymbol{\Sigma}(\boldsymbol{\beta})/\partial\beta_j$has only jth row and jth column being non-zero; specifically:

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \beta_j} = \begin{bmatrix} 0 & 0 & \dots & 0 & \beta_1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \beta_2 & 0 & \dots & 0 \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ \beta_1 & \beta_2 & \dots & \beta_{j-1} & 0 & \beta_{j+1} & \dots & \beta_D \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & 0 & \beta_D & 0 & \dots & 0 \end{bmatrix}.$$

Let $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}) = (\sigma^{ij})_{1\le i,j\le D}$. After some algebraic computations, the first term in (39) becomes

$$\text{trace}\Big[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\beta}_j}\Big] = 2\sum_{1\le \ell\le D, \ell\ne j}^{D} \beta_\ell \sigma^{j\ell} =: 2b_j^{(-j)}(\boldsymbol{\beta}_D). \tag{40}$$

For $k \ne j$, $k \in \{1,\dots,j-1,j+1,D\}$, let

$$b_k^{(-j)}(\boldsymbol{\beta}_D) = \sum_{1\le \ell\le D, \ell\ne j}^{D} \beta_\ell \sigma^{kl}. \tag{41}$$

For a fixed j, the second term of (39) becomes

$$\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \beta_j} = \begin{bmatrix} \beta_1\sigma^{1j} & \beta_2\sigma^{1j} & \dots & \beta_{j-1}\sigma^{1j} & b_1^{(-j)} & \beta_{j+1}\sigma^{1j} & \dots & \beta_D\sigma^{1j} \\ \beta_1\sigma^{2j} & \beta_2\sigma^{2j} & \dots & \beta_{j-1}\sigma^{2j} & b_2^{(-j)} & \beta_{j+1}\sigma^{2j} & \dots & \beta_D\sigma^{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \\ \beta_1\sigma^{jj} & \beta_2\sigma^{jj} & \dots & \beta_{j-1}\sigma^{jj} & b_j^{(-j)} & \beta_{j+1}\sigma^{jj} & \dots & \beta_D\sigma^{jj} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \\ \beta_1\sigma^{Dj} & \beta_2\sigma^{Dj} & \dots & \beta_{j-1}\sigma^{Dj} & b_D^{(-j)} & \beta_{j+1}\sigma^{Dj} & \dots & \beta_D\sigma^{Dj} \end{bmatrix}.$$

By symmetry of $\boldsymbol{\Sigma}^{-1}$,

$$\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \beta_j}\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 2b_1^{(-j)}\sigma^{1j} & 2b_1^{(-j)}\sigma^{2j} & \dots & 2b_1^{(-j)}\sigma^{Dj} \\ 2b_2^{(-j)}\sigma^{1j} & 2b_2^{(-j)}\sigma^{2j} & \dots & 2b_2^{(-j)}\sigma^{Dj} \\ \vdots & \vdots & \vdots & \vdots \\ 2b_D^{(-j)}\sigma^{1j} & 2b_D^{(-j)}\sigma^{2j} & \dots & 2b_D^{(-j)}\sigma^{Dj} \end{bmatrix}.$$

Letting $\boldsymbol{R}_D^0 = (r_{ij})_{1\le i,j\le D}$, the second term of the (39) is

$$2\sum_{k=1}^{D}\sum_{m=1}^{D} b_k^{(-j)}\sigma^{mj} r_{mk}.$$

Overall, the gradient in (39) at the fixed index j is

$$s_j(\boldsymbol{\beta}_D) = 2\Big(\underbrace{b_j^{(-j)}(\boldsymbol{\beta}_D)}_{\text{term1}} - \sum_{k=1}^{D}\sum_{m=1}^{D} b_k^{(-j)}\sigma^{mj} r_{mk}\Big). \tag{42}$$

This concludes Step A.

Step B: Consider $s_j(\boldsymbol{\alpha}_D)$ on the right side of (37). We will show that $s_j(\boldsymbol{\alpha}_D) \to 0$ as $D \to \infty$. More specifically, $s_j(\boldsymbol{\alpha}_D) = O(D^{-1})$ with Assumption 1. We need to evaluate the order of two terms in (42). The second term in (42) can be written as

$$\sum_{k=1}^{D}\sum_{m=1}^{D} b_k^{(-j)}\sigma^{mj} r_{mk} = \underbrace{\sum_{m=1}^{D} b_j^{(-j)}\sigma^{mj} r_{mj}}_{\text{term2}} + \underbrace{\sum_{k=1,k\neq j}^{D}\sum_{m=1}^{D} b_k^{(-j)}\sigma^{mj} r_{mk}}_{\text{term3}}. \tag{43}$$

Step B can be carried out in four sub-steps:

- **Step B.0**: Some simplified expression are provided.
- **Step B.1**: The term1 in (42) with $\beta_D = \alpha_D$ is simplified and evaluated.
- **Step B.2**: The term2 in (43) with $\beta_D = \alpha_D$ is simplified and evaluated.
- **Step B.3**: The term3 in (43) with $\beta_D = \alpha_D$ is simplified and evaluated.

Step B.0 The simplified expressions of (1) σ^{mi}, (2) $b_j^{(-j)}(\boldsymbol{\alpha}_D)$, (3) $b_k^{(-j)}(\boldsymbol{\alpha}_D)$, (4) $\sum_{m=1}^{D}\sigma^{mj}\alpha_m\alpha_k$, (5) $\sum_{m=1}^{D}\sigma^{mj}\gamma_{mk}$, and (6) $\sum_{m=1}^{D}\sigma^{mj} r_{mk}$ are provided for later use.

(1) For the inverse of $\boldsymbol{\Sigma}$ with 1-factor structure, from special case of $\boldsymbol{\Sigma}^{-1}$, page 136 in [6],

$$\sigma^{mi} = -(1+q_D)^{-1}\psi_m^{-2}\alpha_m\alpha_i\psi_i^{-2} \qquad \text{if } m \neq i, \tag{44a}$$

$$\sigma^{mi} = \psi_m^{-2} - (1+q_D)^{-1}\psi_m^{-2}\alpha_m\alpha_m\psi_m^{-2} \qquad \text{if } m = i, \tag{44b}$$

where

$$q_D = \sum_{\ell=1}^{D} \alpha_\ell^2/\psi_\ell^2 \tag{45}$$

is $O(D)$. Therefore, σ^{mi} for $m \neq i$ is $O(D^{-1})$, and $\sigma^{mm} = \psi_m^{-2} + O(D^{-1})$. The inverse matrix of $\boldsymbol{\Sigma}$ will approach a diagonal matrix as $D \to \infty$.

(2) Define $q_D^{(-j)} = q_D - \alpha_j^2\psi_j^{-2}$. Let $\beta_\ell = \alpha_\ell$ in (40). Then

$$\begin{aligned} b_j^{(-j)}(\boldsymbol{\alpha}_D) &= \sum_{\ell=1,\ell\neq j}^{D} \alpha_\ell\sigma^{j\ell} = -\sum_{1\le\ell\le D,\ell\neq j} \alpha_\ell(1+q_D)^{-1}\psi_j^{-2}\alpha_j\alpha_\ell\psi_\ell^{-2} \\ &= -\psi_j^{-2}\alpha_j(1+q_D)^{-1}\sum_{1\le\ell\le D,\ell\neq j}\alpha_\ell[\alpha_\ell\psi_\ell^{-2}] = -\psi_j^{-2}\alpha_j(1+q_D)^{-1}q_D^{(-j)}. \end{aligned}$$

(3) Similarly, let $\beta_\ell = \alpha_\ell$ in (41). With $k \neq j$,

$$b_k^{(-j)}(\boldsymbol{\alpha}_D) = \sum_{\ell=1,\ell\neq j}^{D} \alpha_\ell \sigma^{k\ell} = \alpha_k \sigma^{kk} + \sum_{\ell=1,\ell\neq j,\ell\neq k}^{D} \alpha_\ell \sigma^{\ell k}$$

$$= \alpha_k \big[\psi_k^{-2} - (1+q_D)^{-1}\psi_k^{-2}\alpha_k\alpha_k\psi_k^{-2}\big] - \sum_{\ell=1,\ell\neq j,\ell\neq k}^{D} \alpha_\ell (1+q_D)^{-1}\psi_\ell^{-2}\alpha_\ell\alpha_k\psi_k^{-2}$$

$$= \alpha_k\psi_k^{-2} - \sum_{\ell=1,\ell\neq j}^{D} \alpha_\ell (1+q_D)^{-1}\psi_l^{-2}\alpha_\ell\alpha_k\psi_k^{-2}$$

$$= \alpha_k\psi_k^{-2} - (1+q_D)^{-1}\alpha_k\psi_k^{-2}\sum_{\ell=1,\ell\neq j}^{D} \alpha_\ell^2\psi_\ell^{-2}$$

$$= \alpha_k\psi_k^{-2}\Big[1-(1+q_D)^{-1}\sum_{\ell=1,\ell\neq j}^{D}\alpha_\ell^2\psi_\ell^{-2}\Big] = \alpha_k\psi_k^{-2}\big[1-(1+q_D)^{-1}q_D^{(-j)}\big].$$

Let $q_D^{(-j)}/(1+q_D) = \eta_D^{(-j)}$. In summary,

$$b_j^{(-j)}(\boldsymbol{\alpha}_D) = -\psi_j^{-2}\alpha_j\eta_D^{(-j)} =: \delta_D^{(-j)}, \tag{46}$$

$$b_k^{(-j)}(\boldsymbol{\alpha}_D) = \alpha_k\psi_k^{-2} - \alpha_k\psi_k^{-2}\eta_D^{(-j)} = (1-\eta_D^{(-j)})\alpha_k\psi_k^{-2}, \quad k\neq j. \tag{47}$$

Also, from the definition of $\eta_D^{(-j)}$, algebraic calculations lead to:

$$\eta_D^{(-j)} + (1+q_D)^{-1} = 1-(1+q_D)^{-1}\alpha_j^2\psi_j^{-2} = O(1), \tag{48}$$

$$\eta_D^{(-j)} = 1-\alpha_j^2\psi_j^{-2}(1+q_D)^{-1} - (1+q_D)^{-1} = O(1). \tag{49}$$

(4) By substituting σ^{mj} in (44b), $\sum_{m=1}^{D}\sigma^{mj}\alpha_m\alpha_k$ can be simplified.
For $m\neq j$,

$$\sum_{m=1,m\neq j}^{D}\sigma^{mj}\alpha_m\alpha_k = -\alpha_k\sum_{m=1,m\neq j}^{D}(1+q_D)^{-1}\psi_m^{-2}\alpha_m^2\alpha_j\psi_j^{-2}$$

$$= -(1+q_D)^{-1}\alpha_k\alpha_j\psi_j^{-2}\sum_{m=1,m\neq j}^{D}\psi_m^{-2}\alpha_m^2 = -\eta_D^{(-j)}\alpha_k\alpha_j\psi_j^{-2}.$$

For $m=j$, with (48),

$$\sigma^{jj}\alpha_j\alpha_k = \alpha_k\alpha_j\psi_j^{-2}\big[1-(1+q_D)^{-1}\alpha_j^2\psi_j^{-2}\big] = \alpha_k\alpha_j\psi_j^{-2}\big[\eta_D^{(-j)}+(1+q_D)^{-1}\big].$$

Therefore, from summing the two parts,

$$\sum_{m=1}^{D}\sigma^{mj}\alpha_m\alpha_k = \sum_{m=1,m\neq j}^{D}\sigma^{mj}\alpha_m\alpha_k + \sigma^{jj}\alpha_j\alpha_k = (1+q_D)^{-1}\alpha_k\alpha_j\psi_j^{-2}. \tag{50}$$

(5) Substituting (44a), (44b) and (48) leads to

$$\sum_{m=1}^{D}\sigma^{mj}\gamma_{mk}=\sum_{m=1,m\neq j}^{D}\sigma^{mj}\gamma_{mk}+\sigma^{jj}\gamma_{jk}$$
$$=-\sum_{m=1,m\neq j}^{D}(1+q_D)^{-1}\psi_m^{-2}\alpha_m\alpha_j\psi_j^{-2}\gamma_{mk}+\psi_j^{-2}\left[1-(1+q_D)^{-1}\alpha_j^2\psi_j^{-2}\right]\gamma_{jk}$$
$$=-(1+q_D)^{-1}\alpha_j\psi_j^{-2}\sum_{m=1,m\neq j}^{D}\psi_m^{-2}\alpha_m\gamma_{mk}+\psi_j^{-2}\eta_D^{(-j)}\gamma_{jk}+\psi_j^{-2}(1+q_D)^{-1}\gamma_{jk}. \tag{51}$$

(6) With $r_{mk}=\alpha_m\alpha_k+\gamma_{mk}$, combining the results in (4) and (5) leads to

$$\sum_{m=1}^{D}\sigma^{mj}r_{mk}=\sum_{m=1}^{D}\sigma^{mj}\alpha_m\alpha_k+\sum_{m=1}^{D}\sigma^{mj}\gamma_{mk}$$
$$=(1+q_D)^{-1}\alpha_k\alpha_j\psi_j^{-2}-(1+q_D)^{-1}\alpha_j\psi_j^{-2}\sum_{m=1,m\neq j}^{D}\psi_m^{-2}\alpha_m\gamma_{mk}$$
$$+\psi_j^{-2}\eta_D^{(-j)}\gamma_{jk}+\psi_j^{-2}(1+q_D)^{-1}\gamma_{jk}. \tag{52}$$

Step B.1: By (46), the term 1 in (42) is $\delta_D^{(-j)}$.

Step B.2: For term2 in (43), substituting (46) and (52) with $k=j$ leads to:

$$\sum_{m=1}^{D}b_j^{(-j)}\sigma^{mj}r_{mj}$$
$$=\delta_D^{(-j)}\Big[(1+q_D)^{-1}\alpha_j^2\psi_j^{-2}-(1+q_D)^{-1}\alpha_j\psi_j^{-2}\sum_{m=1,m\neq j}^{D}\psi_m^{-2}\alpha_m\gamma_{mj}$$
$$+\psi_j^{-2}\eta_D^{(-j)}\gamma_{jj}+\psi_j^{-2}(1+q_D)^{-1}\gamma_{jj}\Big]. \tag{53}$$

Using (49), $\delta_D^{(-j)}=O(1)$ in (53) assuming loading parameters bounded away from ±1. As for the four terms inside the brackets in (53), the order of the first and the fourth term is $O(D^{-1})$ because $(1+q_D)^{-1}=O(D^{-1})$. With Assumption 4 for the $\{\gamma_{mj}\}$, the second term is also $O(D^{-1})$ with $\sum_{m=1,m\neq j}\psi_m^{-2}\alpha_m\gamma_{mk}=O(1)$. Furthermore, the third term is $\eta_D^{(-j)}=O(1)$ by noticing $\gamma_{jj}=\psi_j^2$. Overall, the order of term2 in (43) is $\eta_D^{(-j)}\delta_D^{(-j)}+O(D^{-1})$.

Step B.3: For term3 in (43), the substitution (52) and (47) leads to,

$$
\begin{aligned}
&\sum_{k=1,k\neq j}^{D}\sum_{m=1}^{D} b_k^{(-j)}\sigma^{mj} r_{mk} \\
&= (1-\eta_D^{(-j)})\Big[(1+q_D)^{-1}\alpha_j\psi_j^{-2}\sum_{k=1,k\neq j}^{D}\alpha_k^2\psi_k^{-2} \qquad (54)\\
&-(1+q_D)^{-1}\alpha_j\psi_j^{-2}\sum_{k=1,k\neq j}^{D}\alpha_k\psi_k^{-2}\sum_{m=1,m\neq j}^{D}\psi_m^{-2}\alpha_m\gamma_{mk} \\
&+\psi_j^{-2}\eta_D^{(-j)}\sum_{k=1,k\neq j}^{D}\alpha_k\psi_k^{-2}\gamma_{jk}+\psi_j^{-2}(1+q_D)^{-1}\sum_{k=1,k\neq j}^{D}\alpha_k\psi_k^{-2}\gamma_{jk}\Big].
\end{aligned}
$$

In (54), the scalar $(1-\eta_D^{(-j)})$ is $O(D^{-1})$; the first term, second term, and third term inside the bracket are all $O(1)$ while the fourth term is $O(D^{-1})$. Thus, term3 in (43) is $O(D^{-1})$.

Combining the results in **B.1,B.2,B.3**, term1 in (42) is $\delta_D^{(-j)}$, term2 in (43) is $\delta_D^{(-j)}\eta_D^{(-j)}+O(D^{-1})$, and term 3 in (43) is $O(D^{-1})$. The score vector in (42) equals to 2[term1-(term2+term3)] $=2[\delta_D^{(-j)}(1-\eta_D^{(-j)})+O(D^{-1})]$. Since $\delta_D^{(-j)}(1-\eta_D^{(-j)})=O(D^{-1})$ by (49), $s_j(\boldsymbol{\beta}_D)$ in (42) is of order $O(D^{-1})$. With the super-population assumption, the indices are "exchangeable" and the $\{s_j(\boldsymbol{\beta}_D)\}$ are uniformly $O(D^{-1})$. □

Lemma 4. *The diagonal entries in the Hessian matrix $\boldsymbol{H}_D(\boldsymbol{\beta}_D)$ in (37) are uniformly $O(1)$ and the off-diagonal entries are uniformly $O(D^{-1})$.*

Proof. For simpler notation, the subscript D will be suppressed for most variables.

Following similar techniques to the proof of the preceding lemma, the order of off-diagonal and diagonal elements in the Hessian matrix will be shown respectively in **Step C** and **Step D**. Based on the super-population assumption, considering the non-diagonal $(1,2)$ element and diagonal (1,1) element in the Hessian matrix are sufficient, and this will simplify the notation.

Step C. The order of off-diagonal entries in the Hessian matrix.

Take the derivatives of the score vector in (42), the off-diagonal (1,2) entry in the Hessian matrix can be expressed as

$$
\begin{aligned}
\frac{1}{2}\frac{\partial s_1(\boldsymbol{\beta}_D)}{\partial\boldsymbol{\beta}_2} &= \underbrace{\partial\Big[b_1^{(-1)}(\boldsymbol{\beta}_D)\Big(1-\sum_{m=1}^{D}\sigma^{m1}r_{m1}\Big)\Big]\Big/\partial\beta_2}_{\text{term1}} \\
&\quad \underbrace{-\sum_{k=2}^{D}\sum_{m=1}^{D}\frac{\partial(b_k^{(-1)}\sigma^{m1}r_{mk})}{\partial\beta_2}}_{\text{term2}}. \qquad (55)
\end{aligned}
$$

The term2 in (55) can be rewritten as

$$\underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D}\frac{\partial b_k^{(-1)}}{\partial \beta_2}\sigma^{m1} r_{mk}}_{\text{term21}} + \underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D} b_k^{(-1)}\frac{\partial \sigma^{m1}}{\partial \beta_2} r_{mk}}_{\text{term22}} + \underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D} b_k^{(-1)}\sigma^{m1}\frac{\partial r_{mk}}{\partial \beta_2}}_{\text{term23}}. \tag{56}$$

For notational simplicity, let

$$\begin{aligned}
\Delta_1 &= \frac{\partial(1-\alpha_2^2)^{-1}}{\partial \alpha_2} = 2\alpha_2/(1-\alpha_2^2)^2,\\
\Delta_2 &= \frac{\partial \alpha_2(1-\alpha_2^2)^{-1}}{\partial \alpha_2} = (1+\alpha_2^2)/(1-\alpha_2^2)^2, \text{ and}\\
\Delta_3 &= \frac{\partial q_D}{\partial \alpha_2} = \frac{\partial[\alpha_2^2/(1-\alpha_2^2)]}{\partial \alpha_2} = 2\alpha_2/(1-\alpha_2^2)^2 \text{ from (45).}
\end{aligned}$$

The evaluation of the order of off-diagonal entries is done in four sub-steps.

- **Step C.0**: Some simplified expressions are provided for later use.
- **Step C.1**: The term1 in (55) is simplified and evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$.
- **Step C.2**: The term21 in the expression (56) is simplified and evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$.
- **Step C.3**: The term22 in the expression (56) is simplified and evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$.
- **Step C.4**: The term23 in the expression (56) is simplified and evaluated.

Step C.0: With some algebraic calculations, the derivatives of $\sigma^{k\ell}$ in (44a) and (44b) with respect to β_2 when $k \neq 2$ is

$$\frac{\partial \sigma^{k\ell}(\boldsymbol{\beta}_D)}{\partial \beta_2}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \begin{cases} (1+q_D)^{-2}\Delta_3\psi_k^{-2}\alpha_k\alpha_\ell\psi_\ell^{-2} = O(D^{-2}) & \ell \neq 2,\\ -(1+q_D)^{-1}\alpha_k\psi_k^{-2}\Delta_2 + O(D^{-2}) = O(D^{-1}) & \ell = 2. \end{cases} \tag{57}$$

When $k = 2$, the derivative of $\sigma^{2\ell}$ with respect to β_2 evaluated is

$$\frac{\partial \sigma^{2\ell}(\boldsymbol{\beta}_D)}{\partial \beta_2}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \begin{cases} -\frac{[(1+q_D)\alpha_\ell\psi_\ell^{-2}\Delta_2 - \psi_2^{-2}\alpha_2\alpha_\ell\psi_\ell^{-2}\Delta_3]}{(1+q_D)^2} \\ \quad = O(D^{-1}) & \ell \neq 2,\\ \Delta_1 + O(D^{-1}) & \ell = 2. \end{cases} \tag{58}$$

For $k \neq 2$, the derivative of $b_k^{(-1)}(\boldsymbol{\beta}_D) = \sum_{\ell=2}^{D}\beta_\ell\sigma^{k\ell}$ in (41) with respect to β_2 evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ is simplified below by substituting the order of expressions in (44a) and (57):

$$\frac{\partial b_k^{(-1)}(\boldsymbol{\beta}_D)}{\partial \beta_2}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \sigma^{k2}(\boldsymbol{\alpha}_D) + \beta_2\left(\frac{\partial \sigma^{k2}}{\partial \beta_2}\Big|_{\boldsymbol{\alpha}_D}\right) + \sum_{\ell=3}^{D}\beta_\ell\left(\frac{\partial \sigma^{k\ell}}{\partial \beta_2}\Big|_{\boldsymbol{\alpha}_D}\right)$$
$$= O(D^{-1}) + O(D^{-1}) + O(D^{-1}) = O(D^{-1}). \quad (59)$$

In the case of $k=2$, the derivative of $b_2^{(-1)}$ with respect with to β_2 evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$, using (58) and q_D in (45), is

$$\frac{\partial b_2^{(-1)}(\boldsymbol{\beta}_D)}{\partial \beta_2}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \left[\sigma^{22} + \beta_2\frac{\partial \sigma^{22}}{\partial \beta_2} + \sum_{\ell=3}^{D}\beta_\ell\frac{\partial \sigma^{2\ell}}{\partial \beta_2}\right]\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D}$$
$$= (1-\alpha_2^2)^{-1} + 2\alpha_2^2(1-\alpha_2^2)^{-2} - \sum_{\ell=3}^{D}(1+q_D)^{-1}\alpha_\ell^2\psi_\ell^{-2}\Delta_2 + O(D^{-1})$$
$$= \Delta_2 - (1+q_D)^{-1}\sum_{\ell=3}^{D}\Delta_2\alpha_\ell^2\psi_\ell^{-2} + O(D^{-1}) = O(D^{-1}).$$

Step C.1. From (52) with Assumption 4, $\sum_{m=1}^{D}\sigma^{m1}r_{m1} = \psi_1^{-2}\eta_D^{(-1)}\gamma_{11} + O(D^{-1}) = \eta_D^{(-1)} + O(D^{-1})$ as $\gamma_{11} = \psi_1^2$, and thus from (49),

$$1 - \sum_{m=1}^{D}\sigma^{m1}r_{m1} = O(D^{-1}). \quad (60)$$

Next,

$$\sum_{m=1}^{D}\frac{\partial(\sigma^{m1}r_{m1})}{\partial \beta_2} = \sum_{m=1}^{D}\frac{\partial \sigma^{m1}}{\partial \beta_2}r_{m1} + \sum_{m=1}^{D}\frac{\partial r_{m1}}{\partial \beta_2}\sigma^{m1}$$
$$= \frac{\partial \sigma^{21}}{\partial \beta_2}r_{21} + \sum_{m=1,m\neq 2}^{D}\frac{\partial \sigma^{m1}}{\partial \beta_2}r_{m1} + \sigma^{21}\beta_1 \quad (61)$$
$$= O(D^{-1}) + O(D^{-1}) + O(D^{-1}) = O(D^{-1}), \quad (62)$$

using $r_{m1} = \beta_m\beta_1 + \gamma_{m1}$, and $r_{11} = 1$. The orders in (62) are based on the orders computed in (57), (58), and (44a). The derivative of $\partial\sigma^{k\ell}/\partial\beta_2 = O(D^{-1})$ when $k=2, \ell\neq 2$ or $k\neq 2, \ell = 2$ and is $O(D^{-2})$ when $k\neq 2, \ell\neq 2$.

With the above calculation and the term evaluated from step C.0 in (59), the order of term1 in (55) evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ is

$$\frac{\partial b_1^{(-1)}(\boldsymbol{\alpha}_D)}{\partial \beta_2} \times \left(1 - \sum_{m=1}^{D}\sigma^{m1}r_{m1}\right) - b_1^{(-1)}(\boldsymbol{\alpha}_D)\sum_{m=1}^{D}\frac{\partial(\sigma^{m1}r_{m1})}{\partial \beta_2} \quad (63)$$
$$= O(D^{-2}) + O(D^{-1}) = O(D^{-1}).$$

The first term in (63) is based on (59) and (60) while the second term in (63) is based on (62) and $b_1^{(-1)}(\boldsymbol{\alpha}_D) = O(1)$ in (46). Therefore, the term1 in (55) is $O(D^{-1})$.

Step C.2. From equation (52), $\sum_{m=1}^{D}\sigma^{m1}r_{mk} = \psi_1^{-2}\eta_D^{(-1)}\gamma_{1k} + O(D^{-1})$. Then, the term21 in (56) evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ equals to

$$\sum_{k=2}^{D}\frac{\partial b_k^{(-1)}}{\partial\beta_2}\big(\psi_1^{-2}\eta_D^{(-1)}\gamma_{1k} + O(D^{-1})\big). \tag{64}$$

From (59) in Step C.0, $\partial b_k^{(-1)}/\partial\beta_2$ is uniformly $O(D^{-1})$ for $k = 2,\ldots,D$. With the assumption on the weak residual dependence $\sum_{k=1}^{D}|\gamma_{1k}| = O(1)$ and loadings are bounded away from ± 1, and hence the order of term in (64) is $O(D^{-1})$. Thus, the term21 in (56) is $O(D^{-1})$.

Step C.3. Substituting the expression (47) in the term22 in (56), and evaluating at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ leads to

$$(1-\eta_D^{(-1)})\sum_{k=2}^{D}\alpha_k\psi_k^{-2}\sum_{m=1}^{D}\frac{\partial\sigma^{m1}}{\partial\beta_2}\times(\alpha_m\alpha_k + \gamma_{mk}). \tag{65}$$

For the double summation term in (65), substitute (57) leads to

$$\begin{aligned}
&\sum_{k=2}^{D}\alpha_k\psi_k^{-2}\sum_{m=1}^{D}\frac{\partial\sigma^{m1}}{\partial\beta_2}(\alpha_m\alpha_k + \gamma_{mk})\\
&= \sum_{k=2}^{D}\alpha_k\psi_k^{-2}\sum_{m\neq 2}\frac{\partial\sigma^{m1}}{\partial\beta_2}(\alpha_m\alpha_k + \gamma_{mk}) + \sum_{k=2}^{D}\alpha_k\psi_k^{-2}\frac{\partial\sigma^{21}}{\partial\beta_2}(\alpha_2\alpha_k + \gamma_{2k})\\
&= \sum_{k=2}^{D}\alpha_k\psi_k^{-2}\Big\{\sum_{m\neq 2}(1+q_D)^{-2}\Delta_3\psi_m^{-2}\alpha_m\alpha_1\psi_1^{-2})(\alpha_m\alpha_k + \gamma_{mk})\\
&+ \big[-(1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2 + O(D^{-2})\big](\alpha_2\alpha_k + \gamma_{2k})\Big\}\\
&= \sum_{k=2}^{D}\alpha_k\psi_k^{-2}\Big\{(1+q_D)^{-2}\Delta_3\alpha_1\psi_1^{-2}\alpha_k\sum_{m\neq 2}\alpha_m^2\psi_m^{-2}\\
&+ (1+q_D)^{-2}\Delta_3\alpha_1\psi_1^{-2}\sum_{m\neq 2}\alpha_m\psi_m^{-2}\gamma_{mk} - (1+q_D)^{-1}\alpha_1\alpha_2\psi_1^{-2}\Delta_2\alpha_k\\
&- (1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2\gamma_{2k} + O(D^{-2})\Big\}\\
&= \sum_{k=2}^{D}\alpha_k\psi_k^{-2}\Big\{\Delta_3(1+q_D)^{-2}q_D^{(-2)}\alpha_1\alpha_k\psi_1^{-2}\\
&+ \Delta_3(1+q_D)^{-2}\alpha_1\psi_1^{-2}\sum_{m\neq 2}^{D}\alpha_m\psi_m^{-2}\gamma_{mk}\\
&+ O(D^{-2}) - (1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2\alpha_2\alpha_k - (1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2\gamma_{2k}\Big\}.
\end{aligned}$$

Further algebraic simplications lead to:

$$
\begin{aligned}
&= \Delta_3(1+q_D)^{-1}\eta_D^{(-2)}\alpha_1\psi_1^{-2}\sum_{k=2}^{D}\alpha_k^2\psi_k^{-2} \\
&+ \Delta_3(1+q_D)^{-2}\alpha_1\psi_1^{-2}\sum_{k=2}^{D}\alpha_k^2\psi_k^{-2}\sum_{m\neq 2}^{D}\alpha_m\psi_m^{-2}\gamma_{mk} + O(D^{-1}) \\
&- (1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2\alpha_2\sum_{k=2}^{D}\alpha_k^2\psi_k^{-2} - (1+q_D)^{-1}\alpha_1\psi_1^{-2}\Delta_2\sum_{k=2}^{D}\alpha_k\psi_k^{-2}\gamma_{2k} \qquad (66) \\
&= O(1) + O(D^{-1}) + O(D^{-1}) + O(1) + O(D^{-1}) = O(1).
\end{aligned}
$$

$\eta^{(-1)}$ and $\eta^{(-2)}$ are in (48) with $j=1$ and $j=2$ respectively. By noticing $\sum_{k=2}^{D}\alpha_k^2\psi_k^{-2} = O(D)$ and $\sum_{m\neq 2}\alpha_m\psi_m^{-2}\gamma_{mk} = O(1)$, the summation in (66) is $O(1)$. Since $1-\eta_D^{(-1)} = O(D^{-1})$, the order of term in (65) is $O(D^{-1})$. Thus, the term22 in (56) is $O(D^{-1})$.

Step C.4. Since $\partial r_{mk}/\partial\beta_2$ is non-zero only when $m=2$ or $k=2$. term23 in (56) can be simplified, using (44a) and (47), to

$$
\begin{aligned}
&b_2^{(-1)}\sum_{m\neq 2}^{D}\sigma^{m1}\alpha_m + \sigma^{12}\sum_{k=3}^{D}\alpha_k b_k^{(-1)} \\
&= -b_2^{(-1)}\alpha_1\psi_1^{-2}(1+q_D)^{-1}\sum_{m\neq 2}^{D}\alpha_m^2\psi_m^{-2} + \sigma^{12}(1-\eta_D^{(-1)})\sum_{k=3}^{D}\alpha_k^2\psi_k^{-2} \\
&= O(D^{-1}) + O(D^{-1}) = O(D^{-1}).
\end{aligned}
$$

Recall $b_2^{(-1)}$ in (47), σ^{12} in (44a) and $1-\eta_D^{(-1)}$ are all $O(D^{-1})$. The term23 in (56) is also $O(D^{-1})$.

Combining Steps C.0, C.1, C.2, and C.3 and using the super-population assumption, the off-diagonal entries are uniformly of order $O(D^{-1})$.

Step D. The order of diagonal entries in the Hessian matrix.

For the diagonal entries, consider

$$
\frac{1}{2}\frac{\partial s_1(\boldsymbol{\beta}_D)}{\partial\beta_1} = \underbrace{\partial\left[b_1^{(-1)}(\boldsymbol{\beta}_D)\Big(1-\sum_{m=1}^{D}\sigma^{m1}r_{m1}\Big)\right]\Big/\,\partial\beta_1}_{\text{term1}} \underbrace{-\sum_{k=2}^{D}\sum_{m=1}^{D}\frac{\partial(b_k^{(-1)}\sigma^{m1}r_{mk})}{\partial\beta_1}}_{\text{term2}}. \qquad (67)
$$

Similar to the analysis of the off-diagonal entries, we evaluate the order of term 1 and term2 in (67) at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ respectively.

Step D.1: Evaluate term1 in (67)

(1) Using the expressions in (40) and (44a),

$$\frac{\partial b_1^{(-1)}(\boldsymbol{\beta}_D)}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \sum_{\ell=2}^{D} \beta_\ell \frac{\partial \sigma^{\ell 1}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = O(1). \tag{68}$$

(2) Using the expressions in (44a) and (44b),

$$\begin{aligned}
&\sum_{m=1}^{D}\left(\frac{\partial \sigma^{1m}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D}\right) r_{m1} = \frac{\partial \sigma^{11}}{\partial \beta_1} + \sum_{m=2}^{D} \frac{\partial \sigma^{1m}}{\partial \beta_1}(\alpha_m\alpha_1 + \gamma_{1m}) \\
&= 2\alpha_1\psi_1^{-4} + O(D^{-1}) - (1+q_D)^{-1}(1+\alpha_1^2)\psi_1^{-4}\alpha_1 \sum_{m\neq 2}^{D} \alpha_m^2\psi_m^{-2} \\
&\quad - (1+q_D)^{-1}(1+\alpha_1^2)\psi_1^{-4}\sum_{m=2}^{D}\alpha_m\psi_m^{-2}\gamma_{1m} + O(D^{-1}) \qquad (69)\\
&= O(1) + O(D^{-1}) + O(1) + O(D^{-1}) = O(1).
\end{aligned}$$

(3) Using $r_{m1} = \beta_m\beta_1 + \gamma_{1m}$, when $m \neq 1$, $\partial r_{m1}/\partial \beta_1 = \beta_m$ and when $m = 1$, $\partial r_{m1}/\partial \beta_1 = 0$. Therefore using (44b),

$$\sum_{m=1}^{D} \frac{\partial r_{m1}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} \sigma^{m1} = \sum_{m=2}^{D} \alpha_m \sigma^{m1} = O(1). \tag{70}$$

Based on the order in (68), (69), (70), and (60), for the derivative of the term1 in (67) evaluated at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$,

$$\begin{aligned}
&\frac{\partial b_1^{(-1)}(\boldsymbol{\alpha}_D)}{\partial \beta_1} \times \Big(1 - \sum_{m=1}^{D} \sigma^{m1} r_{m1}\Big) - \sum_{m=1}^{D} \frac{\partial(\sigma^{m1} r_{m1})}{\partial \beta_1} \times b_1^{(-1)}(\boldsymbol{\alpha}_D) \\
&= O(D^{-1}) - b_1^{(-1)}(\boldsymbol{\alpha}_D)\left[\sum_{m=1}^{D} \frac{\partial \sigma^{1m}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} r_{m1} + \sum_{m=1}^{D} \frac{\partial r_{m1}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} \sigma^{m1}\right] \\
&= O(D^{-1}) + O(1) = O(1). \qquad (71)
\end{aligned}$$

Step D.2: Evaluate term2 in (67).

For $k = 2, \ldots, D$, from (41),

$$\frac{\partial b_k^{(-1)}(\boldsymbol{\beta}_D)}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = \sum_{\ell=2}^{D} \beta_\ell \frac{\partial \sigma^{\ell k}}{\partial \beta_1}\Big|_{\boldsymbol{\beta}_D=\boldsymbol{\alpha}_D} = O(D^{-1}). \tag{72}$$

The term2 in (67) can be rewritten as

$$\underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D} \frac{\partial b_k^{(-1)}}{\partial \beta_1}\sigma^{m1} r_{mk}}_{\text{term21}} + \underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D} b_k^{(-1)} \frac{\partial \sigma^{m1}}{\partial \beta_1} r_{mk}}_{\text{term22}} + \underbrace{\sum_{k=2}^{D}\sum_{m=1}^{D} b_k^{(-1)} \sigma^{m1} \frac{\partial r_{mk}}{\partial \beta_1}}_{\text{term23}}. \tag{73}$$

Since the technique applied to evaluate the order of each term in (73) is quite similar to Step C, we just briefly discuss the order of terms in (73).

The term21 in (73) is $O(1)$ because $\partial b_k^{(-1)}/\partial\beta_1 = O(D^{-1})$ uniformly for $k = 2,\dots,D$, $\sum_{m=1}^D \sigma^{m1} r_{mk} = \psi_1^{-2}\eta_D^{(-1)}\gamma_{1k} + O(D^{-1})$ uniformly, and $\sum_{k=1}^D \gamma_{1k} = O(1)$ with Assumption 4.

The term22 in (73) is $O(1)$. Substitute the expression (47) in the term22 in (73), and evaluate at $\boldsymbol{\beta}_D = \boldsymbol{\alpha}_D$ leads to

$$(1-\eta_D^{(-1)})\sum_{k=2}^D \alpha_k\psi_k^{-2}\sum_{m=1}^D \frac{\partial\sigma^{m1}}{\partial\beta_1}(\alpha_m\alpha_k+\gamma_{mk}). \tag{74}$$

Similar to the derivation of (65), excluding the scalar $1-\eta_D^{(-1)}$, for the summation term in (74), let $\Delta_{11} = 2\alpha_1\psi_1^{-4}$ and $\Delta_{22} = (1+\alpha_1^2)\psi_1^{-4}$. Substitute $\partial\sigma^{m1}/\partial\beta_1$ computed using the expression in (44b) leads to

$$\begin{aligned}
&\sum_{k=2}^D \alpha_k\psi_k^{-2}\Big(\Delta_{11}(\alpha_1\alpha_k+\gamma_{1k})\\
&-\sum_{m=2}^D\Big[(1+q_D)^{-1}\alpha_m\psi_m^{-2}\Delta_{22}+O(D^{-2})\Big](\alpha_m\alpha_k+\gamma_{mk})\Big)\\
&=\Delta_{11}\alpha_1\sum_{k=2}^D\alpha_k^2\psi_k^{-2}+\Delta_{11}\sum_{k=2}^D\alpha_k\psi_k^{-2}\gamma_{1k}\\
&-(1+q_D)^{-1}\Delta_{22}\sum_{k=2}^D\alpha_k^2\psi_k^{-2}\sum_{m=2}^D\alpha_m^2\psi_m^{-2}\\
&-(1+q_D)^{-1}\Delta_{22}\sum_{k=2}^D\alpha_k\psi_k^{-2}\sum_{m=2}^D\alpha_m\psi_m^{-2}\gamma_{mk}\\
&=O(D)+O(1)+O(D)+O(1)=O(D).
\end{aligned}\tag{75}$$

As $1-\eta_D^{(-1)} = O(D^{-1})$, the term in (74) is $O(1)$.

As for the term23 in (73), $\partial r_{mk}/\partial\beta_1$ is not zero only when $m=1$ or $k=1$. Similar with step C.4, the term23 in (73) can be simplified into

$$b_1^{(-1)}\sum_{m=2}^D\sigma^{m1}\alpha_m+\sigma^{11}\sum_{k=2}^D\alpha_k b_k^{(-1)} = O(1)+O(1)=O(1) \tag{76}$$

by noticing that $\{\alpha_m\}$ are uniformly bounded, $b_1^{(-1)} = O(1)$ in (46), $\sum_{m=2}^D|\sigma^{m1}| = O(1)$ from (44a), $\sigma^{11} = O(1)$ in (44b), and $b_k^{(-1)} = O(D^{-1})$ in (47) for $k \neq 1$.

Combining all these results and under the super-population assumption, the diagonal terms in the Hessian matrix are uniformly $O(1)$. □

Acknowledgements This research has been supported by NSERC Discovery Grant GR010293. Thanks to the referees for their feedback and careful reading.

References

1. Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica **51**(5), 1281–1304 (1983)
2. Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications **84**(2), 313–342 (1999)
3. Fan, X.: Dependence modeling in high dimensions with latent variables. Ph.D. thesis, University of British Columbia (2024)
4. Fan, X., Joe, H.: High-dimensional factor copula models with estimation of latent variables. Journal of Multivariate Analysis **201**(105263) (2024)
5. Harville, D.A.: Matrix Algebra From a Statistician's Perspective. Springer, New York (1997)
6. Joe, H.: Dependence Modeling with Copulas. Chapman & Hall/CRC, Boca Raton, FL (2014)
7. Joe, H.: Parsimonious graphical dependence models constructed from vines. Canadian Journal of Statistics **46**(4), 532–555 (2018)
8. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, 5th edn. Prentice Hall, Englewood Cliffs, NJ (2002)
9. Krupskii, P., Joe, H.: Factor copula models for multivariate data. Journal of Multivariate Analysis **120**, 85–101 (2013)
10. Krupskii, P., Joe, H.: Structured factor copula models: Theory, inference and computation. Journal of Multivariate Analysis **138**, 53–73 (2015)
11. Krupskii, P., Joe, H.: Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models. Statistical Papers **63**(2), 543–569 (2022)
12. Lee, D., Joe, H., Krupskii, P.: Tail-weighted dependence measures with limit being the tail dependence coefficient. Journal of Nonparametric Statistics **30**(2), 262–290 (2018)
13. White, H.: Maximum likelihood estimation of misspecified models. Econometrica **50**(1), 1–25 (1982)