

Properties of stepwise parameter estimation in high-dimensional vine copulas

Jana Gauss¹ and Thomas Nagler^{1,2}

¹*Department of Statistics, LMU, Munich*

²*Munich Center for Machine Learning (MCML)*

November 24, 2025

Abstract

The increasing use of vine copulas in high-dimensional settings, where the number of parameters is often of the same order as the sample size, calls for asymptotic theory beyond the traditional fixed- p , large- n framework. We establish consistency and asymptotic normality of the stepwise maximum likelihood estimator for vine copulas when the number of parameters diverges as $n \rightarrow \infty$. Our theoretical results cover both parametric and nonparametric estimation of the marginal distributions, as well as truncated vines, and are also applicable to general estimation problems, particularly other sequential procedures. Numerical experiments suggest that the derived assumptions are satisfied if the pair copulas in higher trees converge to independence copulas sufficiently fast. A simulation study substantiates these findings and identifies settings in which estimation becomes challenging. In particular, the vine structure strongly affects estimation accuracy, with D-vines being more difficult to estimate than C-vines, and estimates in Gumbel vines exhibit substantially larger biases than those in Gaussian vines.

1 Introduction

The increasing availability of complex, high-dimensional data necessitates flexible methods to describe and estimate dependence between random variables. Copulas enable the modeling of the dependence structure independently of the marginal distributions, thereby providing a framework for constructing multivariate distributions beyond the normal and Student's t -distribution (see [Nelsen \(2007\)](#) and [Genest and Favre \(2007\)](#) for introductions). This decomposition allows the combination of arbitrary, possibly heavy-tailed or skewed, univariate distributions into a multivariate distribution with an arbitrary dependence structure. Although numerous bivariate parametric copula families can capture asymmetric dependence and heavy tails, such flexibility is not readily available in higher dimensions.

Vine copulas provide a versatile approach for modeling multivariate distributions by decomposing a copula into a hierarchical structure of bivariate copulas ([Joe, 1996](#), [Bedford and Cooke, 2001, 2002](#)). Applications include finance and insurance ([Erhardt and Czado, 2012](#), [Dißmann et al., 2013](#), [Aas, 2016](#)), medical sciences ([Stöber et al., 2015](#)), weather forecasting ([Möller et al., 2018](#)), and hydrology ([Hobæk Haff et al., 2015](#)). Most of them use the sequential maximum-likelihood-based approach proposed by [Aas et al. \(2009\)](#) for the parametric estimation of vine copulas. While the joint maximum-likelihood estimator becomes computationally infeasible in high dimensions, the stepwise estimator exploits the hierarchical structure of the model, thereby

substantially reducing the computational complexity. The asymptotic properties of this estimator in finite dimensions have been studied in [Hobæk Haff \(2013\)](#) and [Stöber and Schepsmeier \(2013\)](#).

In classical settings, including many of the aforementioned references, the copula dimension typically does not exceed $d = 15$, while the sample size is larger than $n = 1000$. As the dimension increases, estimation rapidly becomes challenging, since the number of parameters to estimate is approximately $p \approx 0.5d^2$ if each pair copula has one parameter. Comparatively high dimensions are not uncommon, see, for instance, [Hobæk Haff et al. \(2015, \$d = 64, p \approx 2000, n \approx 7500\$ \)](#) and the extreme value copulas in [Kiriliouk et al. \(2024, \$d = 29, p \approx 380, n < 500\$ \)](#). In such applications, the validity of theoretical guarantees for finite-dimensional inference remains unclear. Existing work on high-dimensional vine copulas (see [Czado and Nagler, 2022](#), Section 6.2 for an overview) primarily focuses on sparse models such as truncated vines, especially the choice of an appropriate truncation level ([Kurowicka, 2010](#), [Brechmann et al., 2012](#), [Brechmann and Joe, 2015](#)), and thresholded vine copulas ([Nagler et al., 2019](#)), as well as connections between vines and graphical models, e.g., directed acyclic graphs ([Müller and Czado, 2018, 2019a,b](#)). [Xue and Zou \(2014\)](#) analyze the estimation of covariance matrices of Gaussian copulas when the dimension exceeds the sample size, however, their results provide limited insight into the estimation of vine copulas. The results on model selection in [Nagler et al. \(2019\)](#) rely on the assumption of consistent parameter estimation when the dimension d of the vine copula diverges, which has yet not been proven.

Our main contribution is to provide theoretical results for the sequential estimator by [Aas et al. \(2009\)](#) when $p \rightarrow \infty$. The necessary background on (vine) copulas is presented in [Section 2](#). In [Section 3](#), we establish consistency and asymptotic normality of the sequential estimator, including both parametric and nonparametric estimation of the margins and a discussion of the required assumptions. Although our primary motivation lies in the estimation of parametric, simplified vines, the results are also applicable to non-simplified vines and nonparametric approaches involving splines for example. Moreover, the results are not specific to vine copulas and can also be applied to other sequential procedures as well as general estimation problems. Some of the assumptions are examined numerically in [Section 4](#). The theoretical results are complemented by a simulation study in [Section 5](#). The article concludes with a discussion and some final remarks in [Section 6](#). All proofs can be found in the Appendix.

2 Background

2.1 Copulas

Consider a continuous random vector $\mathbf{X} = (X_1, \dots, X_d) \in \mathcal{X} \subseteq \mathbb{R}^d$ with joint distribution F and marginal distributions F_1, \dots, F_d . Sklar's theorem ([Sklar, 1959](#)) states that any multivariate distribution can be expressed as

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

where $C: [0, 1]^d \mapsto [0, 1]$ is a *copula* that captures the dependence structure among the components of \mathbf{X} . A copula is a multivariate distribution function on $[0, 1]^d$ with uniform marginals ([Czado, 2019](#)). The joint density (or probability mass function) f can be written as

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d),$$

where $c(u_1, \dots, u_d) := \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \dots, u_d)$ denotes the *copula density* and f_1, \dots, f_d are the marginal densities ([Czado, 2019](#)).

2.2 Regular Vines and Vine Copulas

Vine copulas provide a flexible framework for modeling multivariate dependence structures by decomposing a copula into multiple bivariate copulas (also referred to as *pair copulas*) using conditioning. For a comprehensive introduction, we refer to [Czado \(2019\)](#), while [Czado and Nagler \(2022\)](#) provide a concise overview. The underlying idea goes back to [Joe \(1996\)](#) and was formalized by [Bedford and Cooke \(2001, 2002\)](#): any valid factorization of a multivariate density into marginal densities and bivariate conditional densities can be represented as a *regular vine (R-vine) tree structure*. Informally, an R-vine on d elements is a nested sequence of trees $\mathcal{V} = (T_1, \dots, T_{d-1})$, where the edges of each tree become the nodes of the subsequent tree and T_1 has node set $N_1 = \{1, \dots, d\}$. In addition, a certain *proximity condition* must be satisfied. This condition ensures the validity of the following pair copula construction:

Definition 1 (R-vine specification ([Bedford and Cooke, 2002](#))). A *regular vine (R-vine)* specification is a triplet $(\mathcal{F}, \mathcal{V}, \mathcal{B})$, where

1. $\mathcal{F} = (F_1, \dots, F_d)$ is a vector of continuous invertible distribution functions.
2. \mathcal{V} is an R-vine tree sequence on d elements.
3. $\mathcal{B} = \{C_e : e \in E_t, t = 1, \dots, d-1\}$ is a set of bivariate copulas. E_t is the edge set of tree T_t in \mathcal{V} .

In the R-vine tree sequence, each edge e can be labeled as $e = (a_e, b_e; D_e)$, where a_e and b_e are the *conditioned variables* and D_e is the *conditioning set*. Denote by $c_{a_e, b_e; D_e}$ the density of the copula C_e , also written as $C_{a_e, b_e; D_e}$.

[Bedford and Cooke \(2002\)](#) showed that for a triplet $(\mathcal{F}, \mathcal{V}, \mathcal{B})$ satisfying the above conditions, there exists a valid d -dimensional distribution F with density

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d) \prod_{t=1}^{d-1} \prod_{e \in E_t} c_{a_e, b_e; D_e} (F_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), F_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}),$$

such that for each edge $e \in E_t, t = 1, \dots, d-1$, the copula of the bivariate conditional distribution $(X_{a_e}, X_{b_e}) | \mathbf{X}_{D_e} = \mathbf{x}_{D_e}$ is given by $C_{a_e, b_e; D_e}$, i.e.,

$$F_{a_e, b_e|D_e}(x_{a_e}, x_{b_e} | \mathbf{x}_{D_e}) = C_{a_e, b_e; D_e} (F_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), F_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}).$$

Furthermore, the one-dimensional margins of F are given by F_1, \dots, F_d .

One often employs the so-called *simplifying assumption*: the copula $C_{a_e, b_e; D_e}$ does not depend on the specific value of \mathbf{X}_{D_e} . Since our theory is not limited to simplified vines, we consider the general, non-simplified case in the theoretical part of the paper.

Frequently, interest lies primarily in the copula density, i.e., the density of the random vector $\mathbf{U} = F(\mathbf{X}) = (F_1(X_1), \dots, F_d(X_d))$, which is given by

$$c(\mathbf{u}) = \prod_{t=1}^{d-1} \prod_{e \in E_t} c_{a_e, b_e; D_e} (u_{a_e|D_e}, u_{b_e|D_e}; \mathbf{u}_D),$$

where $u_{a_e|D_e} = C_{a_e|D_e}(u_{a_e} | \mathbf{u}_{D_e})$ and $C_{a_e|D_e}$ denotes the conditional distribution of U_{a_e} given \mathbf{U}_{D_e} . These distributions can be computed recursively. In particular, for any variable $c \in D$, it holds

$$C_{a|D}(u_a | \mathbf{u}_D) = \frac{\partial C_{a,c; D \setminus \{c\}}(u_a | D \setminus \{c\}, u_c | D \setminus \{c\}; \mathbf{u}_{D \setminus \{c\}})}{\partial u_c | D \setminus \{c\}}.$$

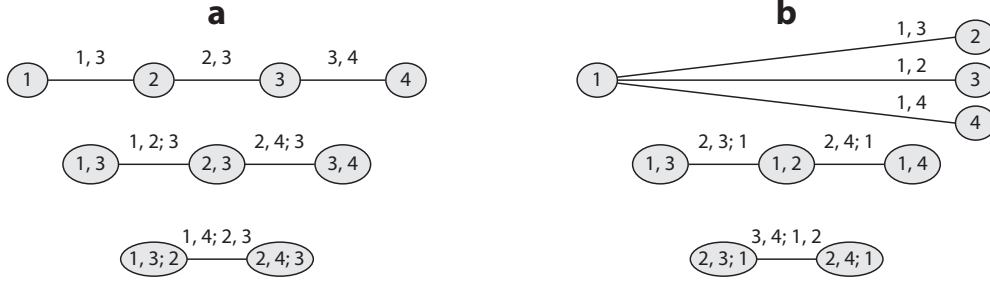


Figure 1: Graphical representations of two R-vine structures in four dimensions: D-vine (a) and C-vine (b). Each edge in the three trees denotes a pair copula.

The aforementioned proximity condition for vine tree structures guarantees that there always exists some variable c such that the copula $C_{a,c;D \setminus \{c\}}$ belongs to the R-vine.

Two common special cases of R-vine structures are illustrated in Fig. 1. In a *drawable* or *D-vine* (a), each tree is a path. Due to the proximity condition, the entire tree structure is determined by the first tree T_1 , which, for example, may consist of the edges $(1, 2), (2, 3), \dots, (d-1, d)$. The second special case is the *canonical* or *C-vine* (b), in which each tree has a root node that is connected to all other nodes.

3 Theoretical Results on Stepwise Maximum Likelihood Estimation in Vine Copulas

3.1 Known Margins

We first assume that the margins F_1, \dots, F_d are known. This makes it easier to study the assumption on the vine copula whose parameters are estimated.

Given parametric models $c_e(\cdot, \cdot; \theta_e)$ for all edges e in the vine, let $\theta = (\theta_e)_{e \in E_t, t=1, \dots, d-1}$ be the stacked parameter vector. Furthermore, denote by $S_a(e)$ the set of edges involved in the recursive computation of $C_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e})$ and define $\theta_{S_a(e)} = (\theta_e)_{e \in S_a(e)}$. The log-likelihood is then given by

$$\ell(\theta; \mathbf{u}) = \sum_{t=1}^{d-1} \sum_{e \in E_t} \ln c_{a_e, b_e; D_e} \left(C_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e}; \theta_{S_a(e)}), C_{b_e|D_e}(u_{b_e}|\mathbf{u}_{D_e}; \theta_{S_b(e)}); \mathbf{u}_{D_e}, \theta_e \right).$$

Given an *iid* sample $\mathbf{U}_1, \dots, \mathbf{U}_n$ from this model, θ can in principle be estimated by maximizing the log-likelihood. However, due to the recursive structure of the conditional copula distributions $C_{a_e|D_e}$, the MLE $\hat{\theta}^{\text{ML}}$ is computationally infeasible in high-dimensional vine copulas (Hobæk Haff, 2013). The solution is a sequential approach, the stepwise MLE (Aas et al., 2009, Hobæk Haff, 2013): The parameters of each bivariate copula are estimated separately, starting with the first tree T_1 and proceeding to tree T_{d-1} :

$$\hat{\theta}_e = \arg \max_{\theta_e} \sum_{i=1}^n \ln c_{a_e, b_e; D_e} \left(C_{a_e|D_e}(U_{i, a_e}|\mathbf{U}_{i, D_e}; \hat{\theta}_{S_a(e)}), C_{b_e|D_e}(U_{i, b_e}|\mathbf{U}_{i, D_e}; \hat{\theta}_{S_b(e)}); \mathbf{U}_{D_e}, \theta_e \right)$$

The vine tree structure ensures that all parameters $\theta_{S_a(e)}, \theta_{S_b(e)}$ involved in the recursive computation of the conditional distribution functions have already been estimated in previous iterations.

Denoting $\boldsymbol{\theta}_{S_{a,b}(e)} = (\boldsymbol{\theta}_{S_a(e)}, \boldsymbol{\theta}_{S_b(e)})$ and

$$\ln c_{a_e, b_e; D_e}(\mathbf{u}, \boldsymbol{\theta}_{S_{a,b}(e)}, \boldsymbol{\theta}_e) = \ln c_{a_e, b_e; D_e}(C_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e}; \boldsymbol{\theta}_{S_a(e)}), C_{b_e|D_e}(u_{b_e}|\mathbf{u}_{D_e}; \boldsymbol{\theta}_{S_b(e)}); \mathbf{u}_{D_e}, \boldsymbol{\theta}_e),$$

the stepwise MLE with known margins, $\hat{\boldsymbol{\theta}}_U$, can be expressed as the solution to the system of equations

$$\sum_{i=1}^n \phi(\mathbf{U}_i; \hat{\boldsymbol{\theta}}_U) = \mathbf{0}, \quad \text{where} \quad \phi(\mathbf{u}; \boldsymbol{\theta}) := \begin{pmatrix} \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e}(\mathbf{u}, \boldsymbol{\theta}_e) \right)_{e \in E_1} \\ \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e; D_e}(\mathbf{u}, \boldsymbol{\theta}_{S_{a,b}(e)}, \boldsymbol{\theta}_e) \right)_{e \in E_2} \\ \vdots \\ \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e; D_e}(\mathbf{u}, \boldsymbol{\theta}_{S_{a,b}(e)}, \boldsymbol{\theta}_e) \right)_{e \in E_{d-1}} \end{pmatrix}. \quad (1)$$

Now define the pseudo-true value $\boldsymbol{\theta}^*$ as the solution to

$$\mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta}^*)] = \mathbf{0}.$$

If the model is correctly specified, these pseudo-true values coincide with the true parameters.

3.1.1 Consistency

We now allow the number of parameters p_n to diverge as the sample size n tends to ∞ . Denote $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{p_n}) \in \mathbb{R}^{p_n}$. If each pair copula has one parameter, p_n coincides with the number of pair copulas, i.e., $p_n = d_n(d_n - 1)/2$, however, our results are formulated in terms of p_n and make no explicit assumption on the relation between d_n and p_n . We assume that the entries of $\boldsymbol{\theta}$ are ordered as $\boldsymbol{\theta} = ((\boldsymbol{\theta}_e)_{e \in E_1}, (\boldsymbol{\theta}_e)_{e \in E_2}, \dots, (\boldsymbol{\theta}_e)_{e \in E_{d_n-1}})$ and that the k -th entry in ϕ , denoted by $\phi(\mathbf{u}; \boldsymbol{\theta})_k$, is the derivative of the respective copula density with respect to θ_k .

Define the $p_n \times p_n$ matrices

$$I(\boldsymbol{\theta}) := \text{Cov}[\phi(\mathbf{U}; \boldsymbol{\theta})], \quad J(\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} \mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta})] = \left(\frac{\partial}{\partial \theta_j} \mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta})_k] \right)_{k,j=1, \dots, p_n}. \quad (2)$$

We assume that derivative and expectation can be interchanged, so $J(\boldsymbol{\theta}) = \mathbb{E}[\nabla_{\boldsymbol{\theta}} \phi(\mathbf{U}; \boldsymbol{\theta})]$. Due to the hierarchical structure of the model, $J(\boldsymbol{\theta})$ is a lower block triangular matrix and, more precisely, a lower triangular matrix if each pair copula has a single parameter.

The parameter $\boldsymbol{\theta}$, functions $\phi(\boldsymbol{\theta})$, $I(\boldsymbol{\theta})$, $J(\boldsymbol{\theta})$ and the support and distribution of the \mathbf{U}_i all depend on n , but we omit the index n to improve readability. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm for vectors, and the spectral norm $\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$ for matrices. $\|\cdot\|_{\infty}$ denotes the maximum norm $\|\mathbf{x}\|_{\infty} = \max\{|x_1|, \dots, |x_p|\}$ and $\|\mathbf{x}\|_0$ denotes the number of non-zero entries in \mathbf{x} .

Denote $r_n = \sqrt{\ln d_n/n}$ and let $\alpha_{j,n}, j = 1, \dots, p_n$ be positive sequences that are bounded away from zero and denote

$$\alpha_n := \max_{1 \leq j \leq p_n} \alpha_{j,n}.$$

Let $\Theta_n \subset \mathbb{R}^{p_n}$ be a sequence of sets with $\Theta_n \supset \{\boldsymbol{\theta} : |\theta_j - \theta_j^*| \leq r_n C \alpha_{j,n} \forall j = 1, \dots, p_n\}$ for all $C < \infty$ and n large. Denote $\Theta_n^{\Delta} = \{\boldsymbol{\Delta} : |\Delta_j| \leq \alpha_{j,n} \forall j = 1, \dots, p_n\}$.

(A1) It holds that $\max_{k=1, \dots, p_n} \mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta}^*)_k^2] = O(1)$.

(A2) It holds that $\mathbb{P}\left(\|\phi(\mathbf{U}; \boldsymbol{\theta}^*)\|_{\infty} > \sqrt{\frac{\sigma_n^2 n}{4 \ln p_n}}\right) = o(1/n)$, where $\sigma_n^2 = \max_{k=1, \dots, p_n} \mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta}^*)_k^2]$.

(A3) There exists a $c > 0$ such that for all $C < \infty$, it holds that

$$\limsup_{n \rightarrow \infty} \max_{1 \leq j \leq p_n} \sup_{\Delta \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{E} [\phi(\mathbf{U}; \boldsymbol{\theta}^* + r_n C \Delta)_j] \leq -c$$

(A4) There are sequences $M_n = o(\sqrt{n/(k_n + \ln p_n)})$ and $D_n = o(n/(k_n + \ln p_n))$ such that

$$\begin{aligned} \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} \mathbb{E} \left[\left(\sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{U}; \boldsymbol{\theta})_j \right| \right)^2 \right] &\leq M_n^2, \\ \mathbb{P} \left(\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} \sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{U}; \boldsymbol{\theta})_j \right| > D_n \right) &= o(1/n), \end{aligned}$$

where k_n is a sequence such that

$$\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n, \mathbf{u} \in \mathbb{R}^{d_n}} \|\nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j\|_0 = k_n.$$

Assumption (A2) is a tail condition which, together with (A1), guarantees that $\|\frac{1}{n} \sum_{i=1}^n \phi(\mathbf{U}_i; \boldsymbol{\theta}^*)\|_\infty = O_p(\sqrt{\ln p_n/n})$ and imposes mild restrictions on the rate of growth of p_n . Assumption (A3) ensures identifiability and is discussed in more detail below. Since this is a population-level condition, (A4) guarantees sufficiently fast convergence of the empirical counterpart, thereby implicitly restricting the rate at which p_n can grow.

Theorem 1 (Consistency with known margins). *Under assumptions (A1)–(A4), with probability tending to 1, the sets Θ_n contain at least one solution $\hat{\boldsymbol{\theta}}_U$ of the estimating equation (1) that satisfies*

$$\max_{1 \leq j \leq p_n} \frac{|\hat{\theta}_{U,j} - \theta_j^*|}{\alpha_{j,n}} = O_p \left(\sqrt{\frac{\ln p_n}{n}} \right).$$

The estimator achieves the optimal global rate of convergence $\|\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*\|_\infty = O_p(\sqrt{\ln p_n/n})$ if $\alpha_n = \max_{1 \leq j \leq p_n} \alpha_{j,n}$ is bounded. The sequences $\alpha_{j,n}$ allow for different rates of convergence for individual parameters: While a single parameter in the first tree can be estimated at the standard \sqrt{n} rate, parameters in higher trees may not achieve this rate, as their estimation depends on the noisy estimates from previous trees. The strength of this dependence is controlled by (A3): For any j , it holds that

$$\begin{aligned} \text{sign}(\Delta_j) \mathbb{E} [\phi(\mathbf{U}; \boldsymbol{\theta}^* + r_n C \Delta)_j] &= \text{sign}(\Delta_j) \mathbb{E} [\phi(\mathbf{U}; \boldsymbol{\theta}^* + r_n C \Delta)_j - \phi(\mathbf{U}; \boldsymbol{\theta}^*)_j] \\ &= \text{sign}(\Delta_j) r_n C \Delta^T \mathbb{E} [\nabla_{\boldsymbol{\theta}} \phi(\mathbf{U}; \tilde{\boldsymbol{\theta}})_j] \end{aligned}$$

with some $\tilde{\boldsymbol{\theta}}$ on the segment between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + r_n C \Delta$. The sensitivity of the estimation of θ_j with respect to other parameters is encoded in $\nabla_{\boldsymbol{\theta}} \phi(\mathbf{U}; \tilde{\boldsymbol{\theta}})_j$. If the estimating equations $\phi(\mathbf{u}; \boldsymbol{\theta})$ consisted of independent estimation problems, only the j -th entry of $\nabla_{\boldsymbol{\theta}} \phi(\mathbf{U}; \tilde{\boldsymbol{\theta}})_j$ would be non-zero and (A3) would simplify to $\mathbb{E} [\frac{\partial}{\partial \theta_j} \phi(\mathbf{U}; \boldsymbol{\theta})_j] \leq -c < 0$ for all $j = 1, \dots, p_n$ and $\boldsymbol{\theta}$ in a neighborhood around $\boldsymbol{\theta}^*$. In vine copulas however, estimation errors in earlier trees influence the estimation in subsequent trees. Assumption (A3) guarantees that this sensitivity sufficiently weak so that each θ_j remains identifiable. Pair copulas that are (close to) independence copulas facilitate estimation in later trees, since the influence of the estimation of θ_j^* on estimates in

subsequent trees diminishes as the corresponding pair copula converges to the independence copula. For simplicity, we assume from now on that $\theta^* = 0$ corresponds to the independence copula, as it is the case for Gaussian copulas.

While $\alpha_{j,n} = 1$ is a suitable choice if $\theta_j^* \rightarrow 0$ sufficiently fast as $d_n \rightarrow \infty$, (A3) together with $\alpha_{j,n} \rightarrow \infty$ allows for consistency results at the cost of a slower rate of convergence when the dependence in the vine decays more slowly. Another special case is that of finite-dimensional vines: As long as $\mathbb{E}[\frac{\partial}{\partial \theta_j} \phi(\mathbf{U}; \boldsymbol{\theta})_j] \leq -c < 0$ for all j (i.e., $J(\boldsymbol{\theta})$ is negative definite), one can always choose sequences $\alpha_{j,n}$ such that (A3) is satisfied with $\alpha_n = O(1)$, since each $\nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j$ has only finitely many non-zero entries. The same holds for truncated vines with a fixed truncation level, provided that $\mathbb{E}[\frac{\partial}{\partial \theta_k} \phi(\mathbf{u}; \boldsymbol{\theta})_j]$ is bounded for all $j, k = 1, \dots, p_n$.

The main restrictions on the growth of p_n are imposed by (A4). Similar to (A3), the sequences M_n and D_n quantify the strength of dependence between the estimation of parameters from different trees. Computations shown in Section 4 suggest that $M_n^2 = O(p_n)$ and $D_n = O(p_n)$ for θ^* small enough, which implies that $p_n^2/n \rightarrow 0$ is sufficient for consistency.

The sequence k_n in (A4) can be interpreted as the maximum number of parameters affecting estimates in subsequent steps. For an untruncated Vine, we typically have $k_n = O(p_n)$. For a vine truncated after tree t_n , there are $t_n d_n - t_n(t_n - 1)/2$ parameters to estimate, however, the estimation of any parameter θ_j depends on at most $O(t_n^2)$ parameters, thus $k_n = O(t_n^2)$. In particular, if the truncation level is fixed, we obtain $k_n = O(1)$. The corresponding estimation problem can be interpreted as estimating a diverging number of finite-dimensional vines and, under standard conditions, Theorem 1 only requires $\ln p_n/n \rightarrow 0$.

If $\alpha_n = O(1)$, Theorem 1 yields $\|\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*\|_2 = O_p(\sqrt{p_n \ln p_n/n})$, which is slightly worse than the optimal rate $\sqrt{p_n/n}$. Convergence at the optimal rate would follow from Gauss and Nagler (2025, Theorem 1), however, this requires that the largest eigenvalue of $J(\boldsymbol{\theta}^*) + J(\boldsymbol{\theta}^*)^\top$ is negative and bounded away from zero. This condition is not satisfied for many truncated C-vines (see Proposition 1 in the appendix). We therefore establish convergence in $\|\cdot\|_\infty$ norm using the Poincaré-Miranda theorem. Notably, while the condition on $J(\boldsymbol{\theta}^*) + J(\boldsymbol{\theta}^*)^\top$ ignores the sequential estimation procedure, i.e., that $J(\boldsymbol{\theta})$ is (block) triangular, this structure is exploited in (A3). Theorem 1 is not specific to vines but can also be applied to other stepwise estimators as well as general estimating problems that can be expressed in the form $\sum_{i=1}^n \phi(\mathbf{X}_i; \hat{\boldsymbol{\theta}}) = \mathbf{0}$ for some function ϕ , e.g., a gradient.

3.1.2 Asymptotic Normality

Since the dimension of $\boldsymbol{\theta}$ increases with the sample size, asymptotic normality of $\hat{\boldsymbol{\theta}}_U$ is formulated in terms of finite-dimensional projections. Let $A_n \in \mathbb{R}^{q \times p_n}$ be some matrix and define $\tilde{r}_n = \alpha_n \sqrt{p_n \ln p_n/n}$.

(A5) For every $C \in (0, \infty)$ and some sequence $D_n = o(n/\tilde{r}_n p_n)$, it holds that

$$\begin{aligned} \sup_{\|\boldsymbol{\Delta}\|, \|\boldsymbol{\Delta}'\| \leq \tilde{r}_n C} \frac{\mathbb{E}[\|A_n[\phi(\mathbf{U}; \boldsymbol{\theta}^* + \boldsymbol{\Delta}) - \phi(\mathbf{U}; \boldsymbol{\theta}^* + \boldsymbol{\Delta}')] \|^2]}{\|\boldsymbol{\Delta} - \boldsymbol{\Delta}'\|^2} &= o\left(\frac{1}{\tilde{r}_n^2 p_n}\right), \\ \mathbb{P}\left(\sup_{\|\boldsymbol{\Delta}\|, \|\boldsymbol{\Delta}'\| \leq \tilde{r}_n C} \frac{\|A_n[\phi(\mathbf{U}; \boldsymbol{\theta}^* + \boldsymbol{\Delta}) - \phi(\mathbf{U}; \boldsymbol{\theta}^* + \boldsymbol{\Delta}')] \|^2}{\|\boldsymbol{\Delta} - \boldsymbol{\Delta}'\|^2} > D_n\right) &= o\left(\frac{1}{n}\right), \\ \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta} \|A_n[J(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta}) - J(\boldsymbol{\theta}^*)]\| &= o\left(\frac{1}{\sqrt{n} \tilde{r}_n}\right). \end{aligned}$$

(A6) It holds that $\mathbb{E}[\|A_n \phi(\mathbf{U}; \boldsymbol{\theta}^*)\|^4] = o(n)$.

Assumption (A5) is a stochastic smoothness condition required to control fluctuations of the estimating equation. The moment condition (A6) typically requires that $p_n^2/n \rightarrow 0$, e.g., if $\|A_n\| = O(1)$ and $\max_k \mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta}^*)_k^4] = O(1)$.

Theorem 2 (Asymptotic normality with known margins). *If conditions (A1)–(A6) hold for some matrix $A_n \in \mathbb{R}^{q \times p_n}$ for which $\Sigma = \lim_{n \rightarrow \infty} A_n I(\boldsymbol{\theta}^*) A_n^\top$ exists, $\hat{\boldsymbol{\theta}}_U$ in Theorem 1 satisfies*

$$\sqrt{n} A_n J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) \rightarrow_d \mathcal{N}(\mathbf{0}, \Sigma).$$

If $\boldsymbol{\theta}$ is finite-dimensional, choosing $A_n = J(\boldsymbol{\theta}^*)^{-1}$ yields $\sqrt{n}(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) \rightarrow_p \mathcal{N}(\mathbf{0}, J(\boldsymbol{\theta}^*)^{-1} I(\boldsymbol{\theta}^*) J(\boldsymbol{\theta}^*)^{-\top})$. In Section 5, we present simulations of $\sqrt{n/p_n} \sum_{j=1}^{p_n} \hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^*$ for different vine copula models. This corresponds to the choice

$$A_n = \frac{1}{\sqrt{p_n}} \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} J(\boldsymbol{\theta}^*)^{-1} \in \mathbb{R}^{1 \times p_n}, \text{ so } \Sigma = \lim_{n \rightarrow \infty} \frac{1}{p_n} \sum_{j=1}^{p_n} \sum_{k=1}^{p_n} (J(\boldsymbol{\theta}^*)^{-1} I(\boldsymbol{\theta}^*) J(\boldsymbol{\theta}^*)^{-\top})_{jk}.$$

3.2 Parametric Estimation of Margins

We now assume parametric models $F_l(x_l; \boldsymbol{\eta}_l)$, $l = 1, \dots, d$ for the marginals and denote $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_d)$. The log-likelihood is then given by

$$\begin{aligned} \ell(\boldsymbol{\eta}, \boldsymbol{\theta}; \mathbf{x}) &= \sum_{l=1}^d \ln f_l(x_l; \boldsymbol{\eta}_l) \\ &\quad + \sum_{t=1}^{d-1} \sum_{e \in E_t} \ln c_{a_e, b_e; D_e} (F_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_a(e)}), F_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_b(e)}); \mathbf{x}_{D_e}, \boldsymbol{\theta}_e) \\ &= \ell_M(\boldsymbol{\eta}; \mathbf{x}) + \ell_C(\boldsymbol{\eta}, \boldsymbol{\theta}; \mathbf{x}). \end{aligned}$$

The maximum likelihood estimator of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ solves the estimating equations

$$\sum_{i=1}^n \begin{pmatrix} \nabla_{\boldsymbol{\eta}} \ell(\hat{\boldsymbol{\eta}}^{\text{ML}}, \hat{\boldsymbol{\theta}}^{\text{ML}}; \mathbf{X}_i) \\ \nabla_{\boldsymbol{\theta}} \ell(\hat{\boldsymbol{\eta}}^{\text{ML}}, \hat{\boldsymbol{\theta}}^{\text{ML}}; \mathbf{X}_i) \end{pmatrix} = \mathbf{0}.$$

Both the joint MLE and the two-step *inference for margins* approach by Joe (2005), which maximizes ℓ_M and ℓ_C sequentially, are computationally too demanding for practical purposes. One therefore again resorts to a sequential approach, slightly modifying the stepwise MLE presented in Section 3.1: first, $\boldsymbol{\eta}$ is estimated by maximizing $\sum_{i=1}^n \ell_M(\hat{\boldsymbol{\eta}}; \mathbf{X}_i)$. Then, the parameters of each bivariate copula are estimated separately, plugging in the estimate $\hat{\boldsymbol{\eta}}$:

$$\hat{\boldsymbol{\theta}}_e = \arg \max_{\boldsymbol{\theta}_e} \sum_{i=1}^n \ln c_{a_e, b_e; D_e} \left(F_{a_e|D_e}(X_{i, a_e} | \mathbf{X}_{i, D_e}; \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_{S_a(e)}), F_{b_e|D_e}(X_{i, b_e} | \mathbf{X}_{i, D_e}; \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_{S_b(e)}); \mathbf{X}_{D_e}, \boldsymbol{\theta}_e \right).$$

Denoting

$$\ln c_{a_e, b_e; D_e}(\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S_a, b(e)}, \boldsymbol{\theta}_e) = \ln c_{a_e, b_e; D_e}(F_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_a(e)}), F_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_b(e)}); \mathbf{x}_{D_e}, \boldsymbol{\theta}_e),$$

the stepwise MLE $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta)$ can be expressed as the solution to the system of equations

$$\sum_{i=1}^n \phi(\mathbf{X}_i; \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta) = \mathbf{0}, \quad \text{where } \phi(\mathbf{x}; \boldsymbol{\eta}, \boldsymbol{\theta}) := \begin{pmatrix} \nabla_{\boldsymbol{\eta}} \sum_{j=1}^d \ln f_j(\mathbf{x}_j; \boldsymbol{\eta}_j) \\ \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e}(\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}_e) \right)_{e \in E_1} \\ \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e; D_e}(\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S_{a,b}(e)}, \boldsymbol{\theta}_e) \right)_{e \in E_2} \\ \vdots \\ \left(\nabla_{\boldsymbol{\theta}_e} \ln c_{a_e, b_e; D_e}(\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S_{a,b}(e)}, \boldsymbol{\theta}_e) \right)_{e \in E_{d-1}} \end{pmatrix}. \quad (3)$$

The pseudo-true values $\boldsymbol{\eta}^*, \boldsymbol{\theta}_\eta^*$ are defined as the solution to

$$\mathbb{E}[\phi(\mathbf{X}; \boldsymbol{\eta}^*, \boldsymbol{\theta}_\eta^*)] = \mathbf{0}.$$

If the model is correctly specified, they coincide with the true parameters.

We again allow the number of parameters to diverge and denote the dimension of the stacked parameter vector $(\boldsymbol{\eta}, \boldsymbol{\theta})$ by p_n . The matrices $I_\eta(\boldsymbol{\eta}, \boldsymbol{\theta})$ and $J_\eta(\boldsymbol{\eta}, \boldsymbol{\theta})$ are defined as

$$I_\eta(\boldsymbol{\eta}, \boldsymbol{\theta}) := \text{Cov}[\phi(\mathbf{X}; \boldsymbol{\eta}, \boldsymbol{\theta})] \in \mathbb{R}^{p_n \times p_n}, \quad J_\eta(\boldsymbol{\eta}, \boldsymbol{\theta}) := \nabla_{(\boldsymbol{\eta}, \boldsymbol{\theta})} \mathbb{E}[\phi(\mathbf{X}; \boldsymbol{\eta}, \boldsymbol{\theta})] \in \mathbb{R}^{p_n \times p_n}.$$

Let $\tilde{\Theta}_n \subset \mathbb{R}^{p_n}$ be a sequence of sets satisfying $\tilde{\Theta}_n \supset \{(\boldsymbol{\eta}, \boldsymbol{\theta}) : |(\boldsymbol{\eta}, \boldsymbol{\theta})_j - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)_j| \leq r_n C \alpha_{j,n} \forall j = 1, \dots, p_n\}$ for all $C < \infty$ and n sufficiently large. Consistency and asymptotic normality of $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta)$ follow from [Theorem 1](#) and [Theorem 2](#) by appropriately adapting the assumptions:

Theorem 3 (Consistency with parametric estimation of margins). *Suppose assumptions (A1)–(A4) hold with Θ_n replaced by $\tilde{\Theta}_n$, $\boldsymbol{\theta}$ replaced by $(\boldsymbol{\theta}, \boldsymbol{\eta})$ and $\phi(\mathbf{u}; \boldsymbol{\theta})$ replaced by $\phi(\mathbf{x}; \boldsymbol{\eta}, \boldsymbol{\theta})$. Then, with probability tending to 1, the sets $\tilde{\Theta}_n$ contain at least one solution $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta)$ of the estimating equation (3) that satisfies*

$$\max_{1 \leq j \leq p_n} \frac{|(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta)_j - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)_j|}{\alpha_{j,n}} = O_p \left(\sqrt{\frac{\ln p_n}{n}} \right).$$

Theorem 4 (Asymptotic normality with parametric estimation of margins). *Suppose assumptions (A1)–(A6) hold with Θ_n replaced by $\tilde{\Theta}_n$, $\boldsymbol{\theta}$ replaced by $(\boldsymbol{\theta}, \boldsymbol{\eta})$ and $\phi(\mathbf{u}; \boldsymbol{\theta})$ replaced by $\phi(\mathbf{x}; \boldsymbol{\eta}, \boldsymbol{\theta})$ for some matrix $A_n \in \mathbb{R}^{q \times p_n}$ for which $\Sigma_\eta = \lim_{n \rightarrow \infty} A_n I_\eta(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) A_n^\top$ exists. Then, the solution $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta)$ from [Theorem 3](#) satisfies*

$$\sqrt{n} A_n J_\eta(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) ((\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\eta) - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)) \rightarrow_d \mathcal{N}(\mathbf{0}, \Sigma_\eta).$$

3.3 Nonparametric Estimation of Margins

Parametric estimation of the marginals is often not the primary focus, or one may prefer an estimator of the copula parameters that is robust to misspecification of the margins. In the semiparametric approach, first proposed by [Genest et al. \(1995\)](#), the marginals $F_l(x_l; \boldsymbol{\eta}_l)$ are replaced by the empirical distribution functions $\hat{u}_l = F_{nl}(x_l) = (n+1)^{-1} \sum_{i=1}^n \mathbb{1}(X_{i,l} \leq x_l)$. Define $\hat{\mathbf{u}} = F_n(\mathbf{x}) = (F_{n1}(x_1), \dots, F_{nd}(x_d))$. In the semiparametric stepwise MLE introduced by [Aas et al. \(2009\)](#), \mathbf{U}_i in (1) is replaced by $\hat{\mathbf{U}}_i$, i.e., we define the estimator $\hat{\boldsymbol{\theta}}_X$ as the solution to

$$\sum_{i=1}^n \phi(\hat{\mathbf{U}}_i; \hat{\boldsymbol{\theta}}_X) = \mathbf{0} \quad (4)$$

with $\phi(\mathbf{u}, \boldsymbol{\theta})$ as defined in (1). The pseudo-true value $\boldsymbol{\theta}^*$ remains the solution to $\mathbb{E}[\phi(\mathbf{U}; \boldsymbol{\theta}^*)] = \mathbf{0}$. Hobæk Haff (2013) showed that, in finite dimensions, this estimator is consistent and asymptotically normal.

We again consider a diverging number of parameters p_n . The notation is the same as in Section 3.1.1, i.e., the expectations in the definitions of $I(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ in (2) and in (A1)–(A4) are still taken with respect to $\mathbf{U} = F(\mathbf{X})$.

Denote

$$\mathbf{Y}_{ii'} := \underbrace{\nabla_{\mathbf{u}} \phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*)}_{\in \mathbb{R}^{p_n \times d_n}} \begin{pmatrix} \mathbb{1}(X_{i'1} \leq X_{i1}) - F_1(X_{i1}) \\ \vdots \\ \mathbb{1}(X_{i'd_n} \leq X_{id_n}) - F_{d_n}(X_{id_n}) \end{pmatrix} \in \mathbb{R}^{p_n} \quad (5)$$

and

$$h(\mathbf{X}_i, \mathbf{X}_{i'}) := \frac{1}{2} \mathbf{Y}_{ii'} + \frac{1}{2} \mathbf{Y}_{i'i} \quad \text{and} \quad h_1(\mathbf{x}) := \mathbb{E}[h(\mathbf{x}, \mathbf{X})]. \quad (6)$$

We require some additional assumptions to account for the estimation of $F(\mathbf{X})$:

(A7) Denote $\sigma_n^2 = \max_{1 \leq k \leq p_n} \mathbb{E}[h_1(\mathbf{X})_k^2]$. It holds that $\sigma_n^2 = O(1)$ and

$$\mathbb{P} \left(\|h_1(\mathbf{X})\|_{\infty} > \sqrt{\frac{\sigma_n^2 n}{4 \ln p_n}} \right) = o(1/n).$$

(A8) It holds that $\mathbb{E}[h(\mathbf{X}_i, \mathbf{X}_{i'})_k^2] = O(1)$ for all $k = 1, \dots, p_n$ and $i, i' = 1, \dots, n$.

(A9) Define

$$\mathcal{G}_n := \left\{ G: \mathbb{R}^{d_n} \rightarrow [0, 1]^{d_n}, G(\mathbf{x}) = (G_1(x_1), \dots, G_{d_n}(x_{d_n})) \mid \right. \\ \left. G_m, m = 1, \dots, d_n \text{ is a continuous distribution function and} \right. \\ \left. \sup_{x \in \mathbb{R}, 1 \leq m \leq d_n} \frac{|G_m(x) - F_m(x)|}{w(F_m(x))} \leq C d_n^b \sqrt{\ln d_n / n} \text{ for all } C < \infty \text{ and } n \text{ large enough} \right\}$$

with $w(s) = s^\gamma(1-s)^\gamma$ for some $\gamma \in (0, 1)$ and $b > \gamma$. Denote $\partial_{ml} \phi(\mathbf{u}; \boldsymbol{\theta})_k = \partial^2 / (\partial u_m \partial u_l) \phi(\mathbf{u}; \boldsymbol{\theta})_k$. It holds that

$$\mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} |\partial_{ml} \phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k|^2 \right] = O(d_n \ln p_n).$$

(A10) For all $l = 1, \dots, d_n$, there exists a function $\psi_l: (0, 1)^{d_n} \rightarrow \mathbb{R}_0^+$ such that for all $\mathbf{u} \in (0, 1)^{d_n}$, $k = 1, \dots, p_n$ and $\boldsymbol{\Delta} \in \mathbb{R}^{p_n}$, it holds that

$$\left| \frac{\partial}{\partial u_l} [\phi(\mathbf{u}; \boldsymbol{\theta}^* + \boldsymbol{\Delta})_k - \phi(\mathbf{u}; \boldsymbol{\theta}^*)_k] \right| \leq \|\boldsymbol{\Delta}\|_{\infty} \psi_l(\mathbf{u}).$$

Denote $\boldsymbol{\psi}(\mathbf{u}) = (\psi_l(\mathbf{u}))_{l=1, \dots, d_n}$. With \mathcal{G}_n as defined in (A9), it holds that

$$\mathbb{E} \left[\sup_{G \in \mathcal{G}_n} \|\boldsymbol{\psi}(G(\mathbf{X}))\|_2 \right] = O(\sqrt{d_n}).$$

(A11) With $\alpha_{j,n}, j = 1, \dots, p_n$, it holds that, for all $j = 1, \dots, p_n$,

$$\frac{\max_{1 \leq k \leq j} \alpha_{k,n}}{\alpha_{j,n}} = O(1).$$

Theorem 5 (Consistency with nonparametric estimation of margins). *Under assumptions (A1)–(A11) and $d_n^3/n = O(1)$, with probability tending to 1, the sets Θ_n contain at least one solution of the estimating equation (4) that satisfies*

$$\max_{1 \leq j \leq p_n} \frac{|\hat{\theta}_{X,j} - \theta_j^*|}{\alpha_{j,n}} = O_p \left(\sqrt{\frac{\ln p_n}{n}} \right).$$

(A12) With \mathcal{G}_n and $\partial_{ml}\phi(\mathbf{u}; \boldsymbol{\theta})_k$ as defined in (A9), holds that

$$\mathbb{E} \left[\sup_{G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \sum_{k=1}^{p_n} |\partial_{ml}\phi(G(\mathbf{X}); \boldsymbol{\theta}^*)_k|^2 \right] = O(p_n)$$

(A13) With $h_1(\mathbf{x})$ as defined in (6), it holds that $\mathbb{E}[\|A_n h_1(\mathbf{X})\|^4] = o(n)$.

Let C be the true copula distribution of \mathbf{X} , i.e., the distribution of $\mathbf{U} = F(\mathbf{X})$.

Theorem 6 (Asymptotic normality with nonparametric estimation of margins). *Suppose assumptions (A1)–(A13) hold with some matrix $A_n \in \mathbb{R}^{q \times p_n}$ with $\|A_n\| = O(1)$ and for which $\Sigma_X = \lim_{n \rightarrow \infty} A_n \tilde{\Sigma}_n A_n^\top$ exists, where*

$$\tilde{\Sigma}_n = \text{Cov} \left(\phi(\boldsymbol{\xi}; \boldsymbol{\theta}^*) + \int \sum_{l=1}^{d_n} \frac{\partial}{\partial u_l} \phi(\mathbf{u}; \boldsymbol{\theta}^*) (\mathbb{1}(\xi_l \leq u_l) - u_l) dC(\mathbf{u}) \right),$$

and $\boldsymbol{\xi}$ is a random variable with distribution C , and $\alpha_n^2 p_n d_n^2 \ln d_n/n \rightarrow 0$. Then, $\hat{\boldsymbol{\theta}}_X$ in Theorem 5 satisfies

$$\sqrt{n} A_n J(\boldsymbol{\theta}^*) (\hat{\boldsymbol{\theta}}_X - \boldsymbol{\theta}^*) \rightarrow_d \mathcal{N}(\mathbf{0}, \Sigma_X).$$

For finite dimensions, choosing $A_n = J^{-1}(\boldsymbol{\theta}^*)$, Theorem 6 recovers the results from Tsukahara (2005) and Hobæk Haff (2013). Compared to the estimation with known margins in Theorem 2, an additional term appears in the covariance matrix stemming from the increased uncertainty due to the estimation of margins.

Since they only cover the finite-dimensional case, Tsukahara (2005) and Hobæk Haff (2013) require fewer assumptions for consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_X$. In particular, standard conditions like continuity of $\phi(\mathbf{u}; \boldsymbol{\theta})$ in its arguments and finite second moments suffice when $d_n = O(1)$. In contrast, when the copula dimension diverges, additional assumptions on the sensitivity of $\phi(\mathbf{u}; \boldsymbol{\theta})$ with respect to \mathbf{u} are required to ensure convergence of expressions involving $\phi(\mathbf{X}_i; \boldsymbol{\theta})$ to their population counterparts $\phi(\mathbf{U}_i; \boldsymbol{\theta})$.

4 Numerical Validation of Assumptions

In this section, we numerically investigate some of the assumptions underlying the theorems presented above for Gaussian vine copulas. While Gaussian vines are of limited use in practice, their simplicity allows computations in high dimensions, as all required derivatives can easily be computed analytically. Despite this limitation, the findings provide valuable insights into the conditions under which the theoretical results from the previous section are valid.

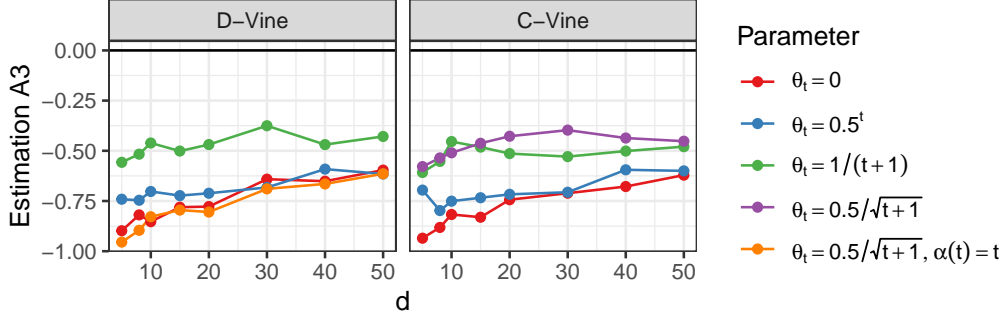


Figure 2: Estimated values for the validation of (A3) for Gaussian vines. The assumption is satisfied if the estimates are negative and bounded away from 0. The supremum over θ is approximated by taking the maximum over $K = 50$ values of θ . $\alpha_{j,n} = 1$ unless otherwise stated.

4.1 Validation of (A3)

The key assumption to ensure identifiability is (A3). A central question is how to choose the sequences $\alpha_{j,n}$ such that the expression in (A3) remains negative and bounded away from 0. If $\max_{1 \leq j \leq p_n} \alpha_{j,n} = O(1)$, the optimal rate of convergence in $\|\cdot\|_\infty$ norm, $\sqrt{\ln p_n/n}$, is attained. As a starting point, we set all $\alpha_{j,n} = 1$. If the numerical approximation produces positive values, we try increasing functions for $\alpha_{j,n}$.

We computed

$$\max_{1 \leq j \leq p} \max_{k=1, \dots, K} \frac{1}{|\Delta_j|} \frac{1}{N} \sum_{i=1}^N \phi(U_i; \theta^* + \Delta_k)_j - \phi(U_i; \theta^*)_j \quad (7)$$

for Gaussian D- and C-vines with various models for the true parameter. Since $\mathbb{E}[\phi(U; \theta^*)] = \mathbf{0}$, we can compute the difference $\phi(u; \theta^* + \Delta) - \phi(u; \theta^*)$ rather than $\phi(u; \theta^* + \Delta)$, which improves numerical stability when approximating the expectation empirically.

For simplicity, all true parameters (the partial correlations in a Gaussian vine) in a given tree are set to the same value. Let θ_t^* denote the true parameter in the t -th tree. As discussed in the previous section, estimation is facilitated, or perhaps only possible, if the pair copulas in higher trees converge to independence copulas. Accordingly, we select functions for θ_t^* that converge to 0 at different rates.

The dimension d ranges from 5 to 50 and the number of samples used to estimate the expectation is $N = 2000 \log(d)$. To approximate the supremum in (A3), we draw $K = 50$ vectors Δ_k with entries $\Delta_{k,j} = \pm \varepsilon \alpha_j$, where $\alpha_j = \alpha(t(j))$, and $t(j)$ denotes the tree of the j -th parameter. The sign of $\Delta_{k,j}$ is drawn uniformly from $\{-1, 1\}$. The constant ε is 0.005 and $\alpha(t) = 1$ for all models except for the D-vine with $\theta_t^* = 0.5/\sqrt{t+1}$.

The results are shown in Fig. 2. For $\theta_t^* = 0$, $\theta_t^* = 0.5^t$ and $\theta_t^* = 1/(t+1)$, and the C-vine with $\theta_t^* = 0.5/(t+1)$, we obtain negative estimates when all $\alpha_{j,n}$ are set to 1. In contrast, for the D-vine with $\theta_t^* = 0.5/\sqrt{t+1}$, the same setup produces positive values that are too large to show here. While the increasing but bounded sequence $\alpha(t) = \sum_{s=1}^t s^{-1.1}$ also yields positive values (not shown here), negative estimates are obtained using $\alpha(t) = t$ with $\varepsilon = 10^{-7}$. Using a diverging sequence for $\alpha_{j,n}$ results in a slower rate of convergence in Theorem 1, here $\|\hat{\theta} - \theta^*\|_\infty = O_p(\sqrt{d_n^2 \ln d_n/n})$.

Although these results are obtained in finite dimensions and for a finite number of sampled θ_k values and should therefore be interpreted with caution, they provide a useful impression of the relationship between (A3) and the rate of convergence in Theorem 1.

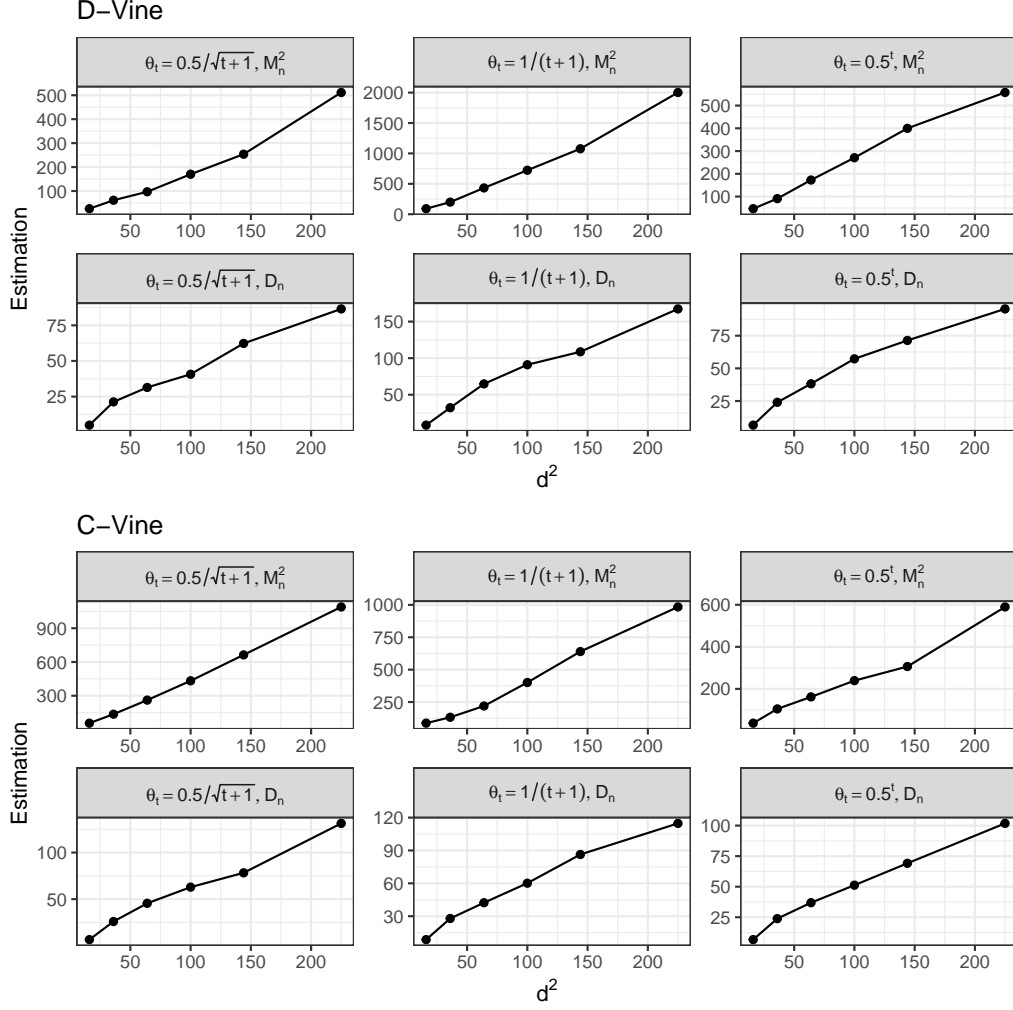


Figure 3: Estimated values for M_n^2 and D_n in (A4). The supremum over θ is approximated by taking the maximum over $K = 30$ values of θ .

4.2 Estimation of M_n and D_n from (A4)

By estimating the quantities M_n and D_n , we can obtain a rough assessment of the restrictions on p_n imposed by (A4). For Gaussian vines, it is relatively easy to analytically compute derivatives of $\phi(\mathbf{u}; \theta)$ w.r.t. θ .

The supremum is again approximated by drawing $K = 30$ vectors Δ_k as described in the previous section. The constant ε is again set to 0.005 and $\alpha(t) = 1$, with the exceptions for the D-vine with $\theta_t^* = 0.5/\sqrt{t+1}$, where $\varepsilon = 10^{-7}$ and $\alpha(t) = t$. Due to the increased computational complexity compared to the previous section, the maximal dimension is limited to $d = 15$. To approximate the expectation in the definition of M_n^2 , we draw $200 \log(d)$ samples. For the estimation of D_n , the $1/n$ term in (A4) must be replaced by some value depending on the number of parameters p . Based on the results for M_n , we compute the $(1 - 15/p^2)$ -quantile since $p_n^2 \sim n$, using the same $200 \log(d)$ samples. The results, shown in Fig. 3, suggest that $M_n = O(d_n)$ and $D_n = O(d_n^2)$ for all considered models, indicating that the condition $p_n^2/n \rightarrow 0$ should be sufficient for (A4) to hold.

5 Simulation Study

To evaluate the performance of the stepwise maximum likelihood estimator, we compare the estimates $\hat{\theta}_U$ and $\hat{\theta}_X$ to the true values θ^* for several (simplified) vine copula models. The structure is either a D- or a C-vine. All bivariate copulas within a given vine are chosen from the same family, either Gaussian, Gumbel, or Student's t. Gaussian vines provide a well-behaved benchmark with light tails, while Gumbel vines exhibit heavy tails and asymmetric dependence. With Student's t, we include a copula family with a second parameter, in which $\rho = 0$ does not correspond to independence.

For Gumbel, we chose the following parameterization: For $\theta \geq 0$, we use the Gumbel copula with parameter $\theta + 1$. For $\theta < 0$, we use the 90-degree rotated Gumbel copula with parameter $-\theta + 1$. This ensures that $\theta = 0$ (independence copula) does not lie on the boundary of the parameter space. For the Student's t vines, the degrees of freedom parameter ν is set to $\nu^* = 4$ for all pair copulas.

The dimension d ranges from 10 to 200 (100 for Student's t for computational reasons), and the sample size n ranges from 100 to 5000. The models for the true parameter are those from the previous section. For each configuration, $N = 100$ data sets ($N = 50$ for Student's t) of size n are generated by sampling copula data U_i or the untransformed data $X_i, i = 1, \dots, n$. We estimate $\hat{\theta}_U$ and $\hat{\theta}_X$ via stepwise ML and compute the normalized maximum norm of the error $\sqrt{n/\log(d_n)}\|\hat{\theta} - \theta^*\|_\infty$, which should be bounded in probability if [Theorem 1](#) and [Theorem 5](#) hold, as well as the normalized sum of errors $\sqrt{n}/d_n \sum_{k=1}^p \hat{\theta}_k - \theta_k^*$. If the assumptions of [Theorem 2](#) and [Theorem 6](#) are satisfied and n is sufficiently large, this sum should be approximately normally distributed with mean zero and variance independent of both d and n .

When assessing consistency, the following question arises: Since the theoretical results concern the regime in which both d and n diverge, which pairs (n, d) should be compared to each other with respect to $\sqrt{n/\log(d_n)}\|\hat{\theta} - \theta^*\|_\infty$? Our approach is as follows: We fix a set of values for d and three different growth regimes for d relative to n , namely $d \sim n$, $d \sim n^2$ and $d \sim n^3$.¹ For each d and each regime, we compute the corresponding sample size n^{new} . Since we usually do not have d -dimensional estimates $\hat{\theta}_{n^{\text{new}}}^{(d)}$ based on samples of size n^{new} , we estimate the mean of the error $\|\hat{\theta}_{n^{\text{new}}}^{(d)} - \theta^{*,(d)}\|_\infty$ using linear interpolation or, when extrapolation is required, a linear model fitted to the estimates using the two n values that are closest to n^{new} . Both interpolation and extrapolation are performed on the log scale, as for fixed d , we expect $\|\hat{\theta}_n - \theta^*\|_\infty \approx c n^a$, i.e., $\log \|\hat{\theta}_n - \theta^*\|_\infty = \log c + a \log n$. In [Fig. 4](#), [5](#) and [7](#), we show the estimated normalized errors for the three growth regimes. A stabilization of the values indicates that, for the given relation between n and d , the parameter estimates are consistent in $\|\cdot\|_\infty$ norm.

Note that the results for $n \sim d^3$ are only computed for $d \geq 75$ (for Gaussian and Gumbel vines): To avoid extreme extrapolation for $d = 200$, the constant in $n \sim d^3$ is so small that for small d , we obtain n^{new} that are too small to yield reasonable estimates.

5.1 Parameter Estimation without Estimation of Margins

In this section, the estimation of marginals is omitted, i.e., the true copula data $U_i = F(X_i)$ are used.

Gaussian and Gumbel Vines [Fig. 4](#) shows the error norm for Gaussian and Gumbel vines.

¹The explicit functions are $n = 25d$, $n = 0.125d^2$ and $n = 0.003(d - 50)^3$ (for $d > 50$) for Gaussian and Gumbel vines, and $n = 0.005d^3$ for Student's t.

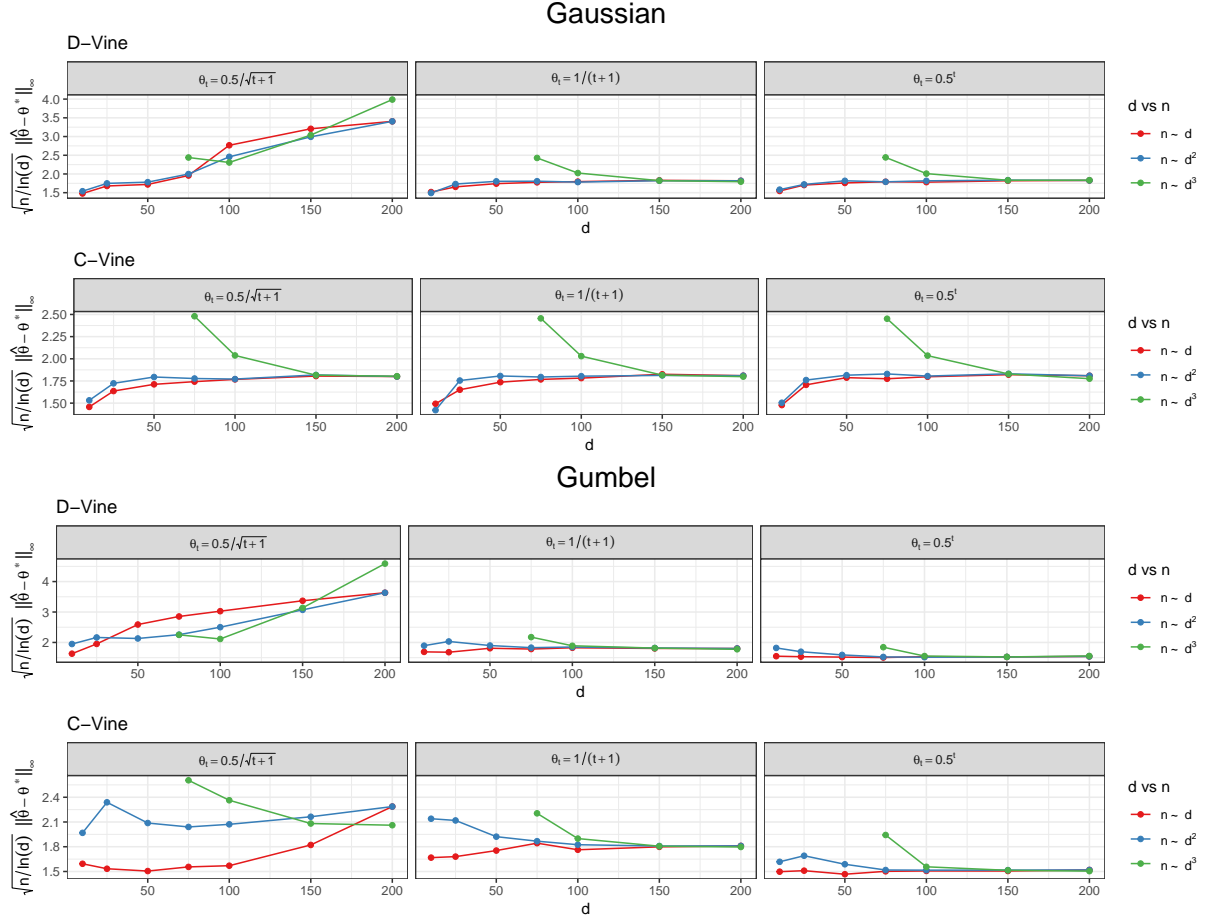


Figure 4: Parameter estimation for Gaussian and Gumbel vines, mean maximum norm of estimation error for different proportions of d and n . Parameterization: $\theta = \rho$ for Gaussian and $\text{Gumbel}(\theta + 1)$ for $\theta \geq 0$, $\text{Gumbel}_{90}(\theta - 1)$ for $\theta < 0$.

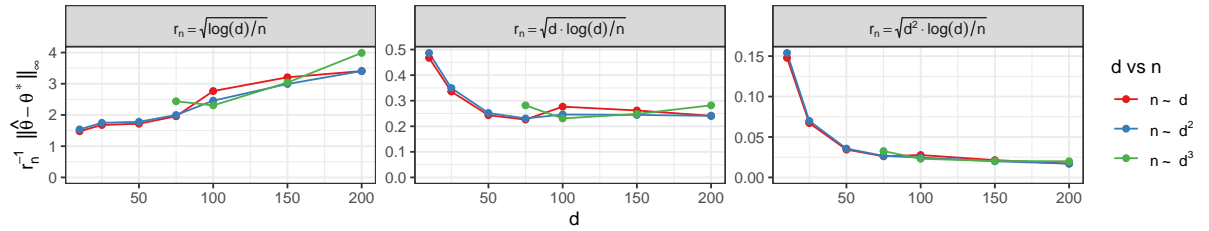


Figure 5: Parameter estimation for Gaussian D-vine with $\theta_t^* = 1/\sqrt{t+1}$, mean maximum norm of estimation error with different normalizations.

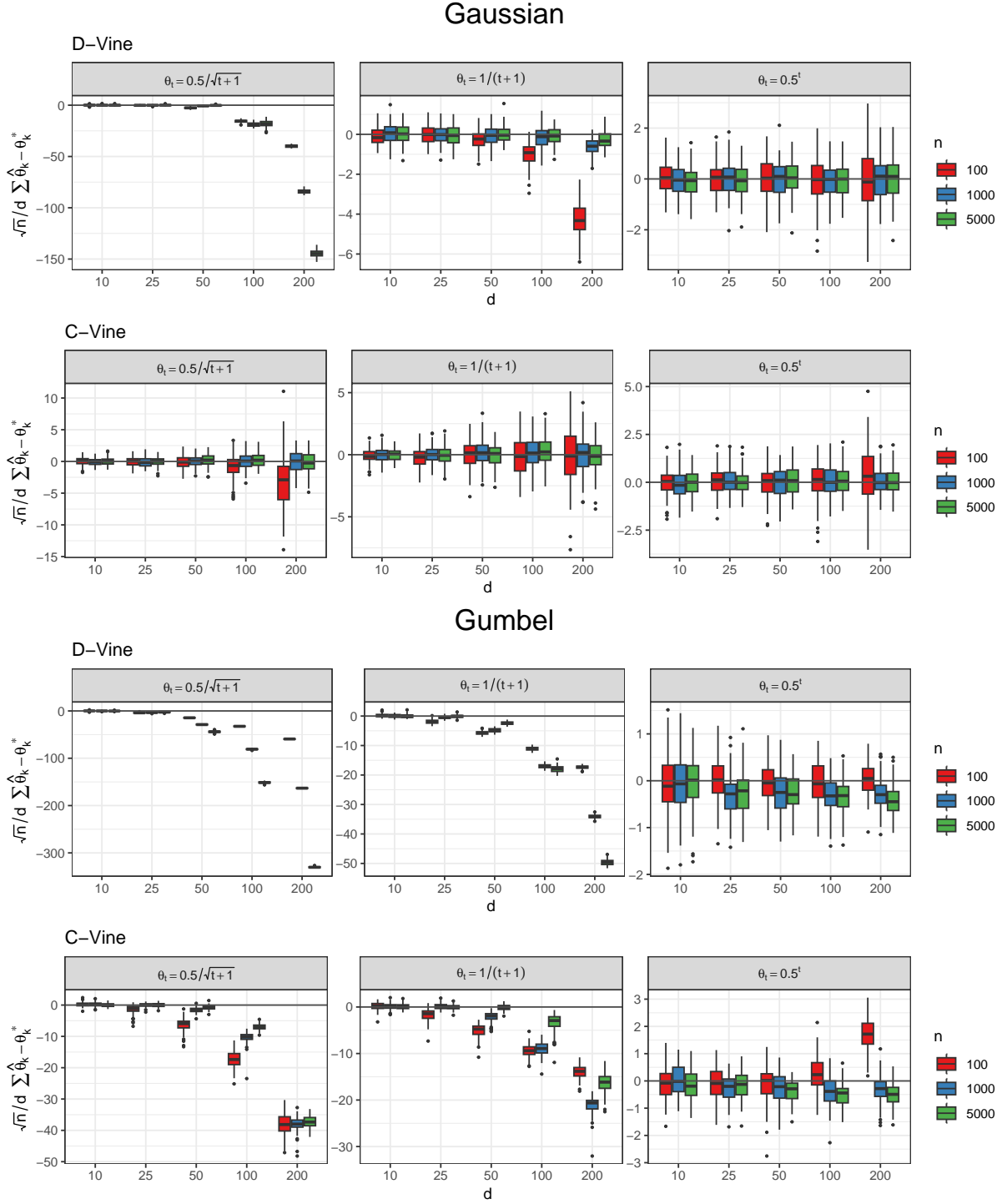


Figure 6: Parameter estimation for Gaussian and Gumbel vines, sum of estimation errors. Parameterization: $\theta = \rho$ for Gaussian and $\text{Gumbel}(\theta + 1)$ for $\theta \geq 0$, $\text{Gumbel}_{90}(\theta - 1)$ for $\theta < 0$. Each boxplot represents $N = 100$ replications.

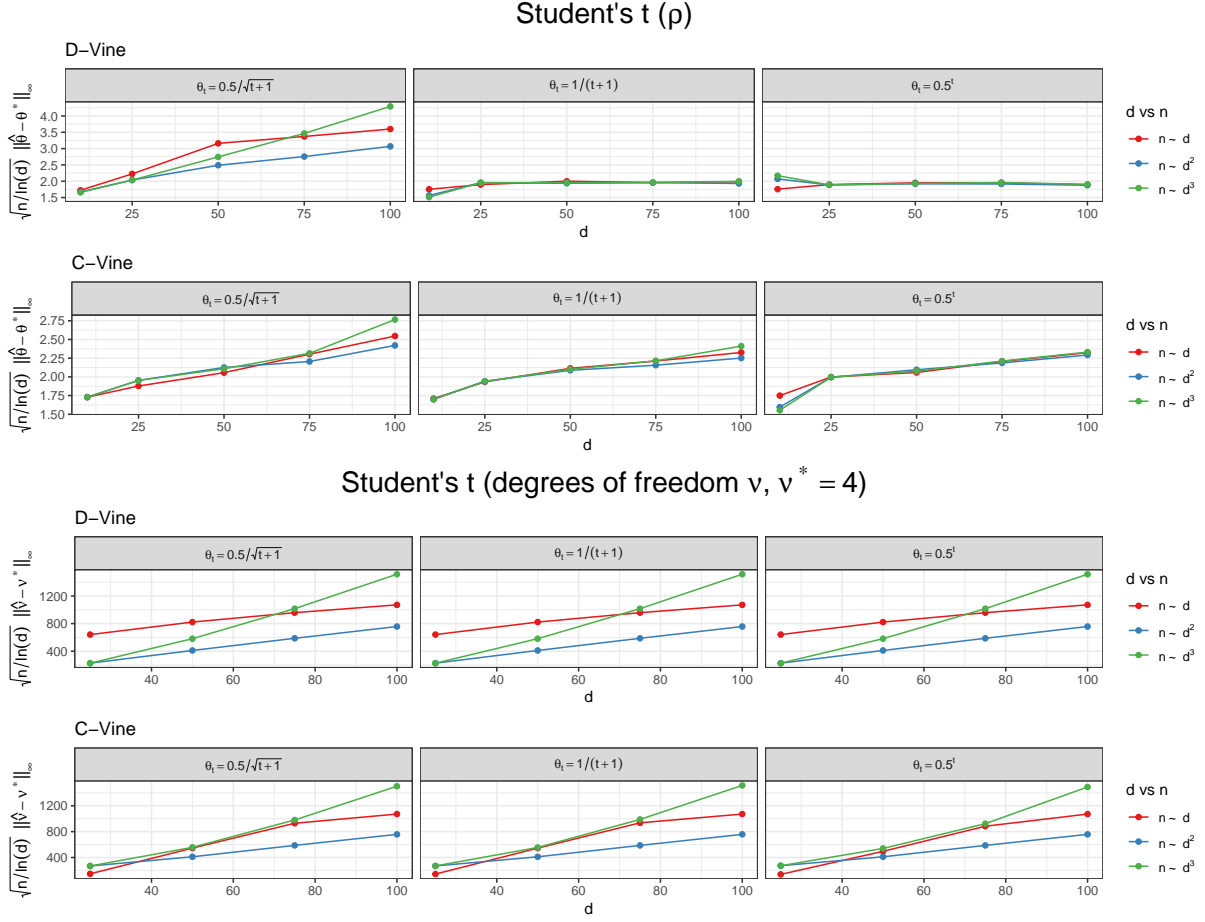


Figure 7: Parameter estimation for Student's t vines, mean maximum norm of estimation error for different proportions of d and n .

For Gaussian vines, $\hat{\theta}_U$ appears to be consistent for all models even when $n \sim d$, except for the D-vine with $\theta_t = 0.5/\sqrt{t+1}$, as already indicated by the results in [Section 4.1](#). Estimation in Gumbel vines mostly exhibits similar behavior, except for the C-vine with $\theta_t = 0.5/\sqrt{t+1}$, which seems to require $n \sim d^2$ or even $n \sim d^3$ for consistency.

For the Gaussian D-vine with $\theta_t = 0.5/\sqrt{t+1}$, $\|\hat{\theta}_U - \theta^*\|_\infty$ is shown with different normalizations in [Fig. 5](#). [Fig. 4](#) indicates that $\hat{\theta}_U$ does not converge at rate $\sqrt{\log(d_n)/n}$ (in maximum norm), however, [Theorem 1](#) also allows for slower convergence rates. While the results in [Section 4.1](#) ($\alpha(t) = t$, i.e., $\alpha_n = d_n$) imply the rate $\sqrt{d_n^2 \log(d_n)/n}$, [Fig. 5](#) suggests that already with rate $\sqrt{d_n \log(d_n)/n}$, the mean maximum norm is bounded.

The normalized sum of errors is shown in [Fig. 6](#) to assess the asymptotic normality of $\hat{\theta}_U$. Recall from [Section 3.1.1](#) that this requires stronger assumptions on ϕ and d_n than consistency. For Gaussian vines, $\hat{\theta}_U$ is asymptotically normal with mean zero in all models except D-vine, $\theta_t = 0.5/\sqrt{t+1}$. For Gumbel vines however, $\hat{\theta}_U$ is heavily biased for both C- and D-vines with $\theta_t = 0.5/\sqrt{t+1}$ and $\theta_t = 1/(t+1)$, with larger biases for D-vines. Even for $\theta_t = 0.5^t$, the results indicate a negative bias. While $\hat{\theta}_U$ is a consistent estimator, a closer analysis shows that, while the magnitudes of positive and negative biases are approximately the same, negative biases (i.e., $\hat{\theta}_k < \theta_k^*$) occur far more often than positive ones.

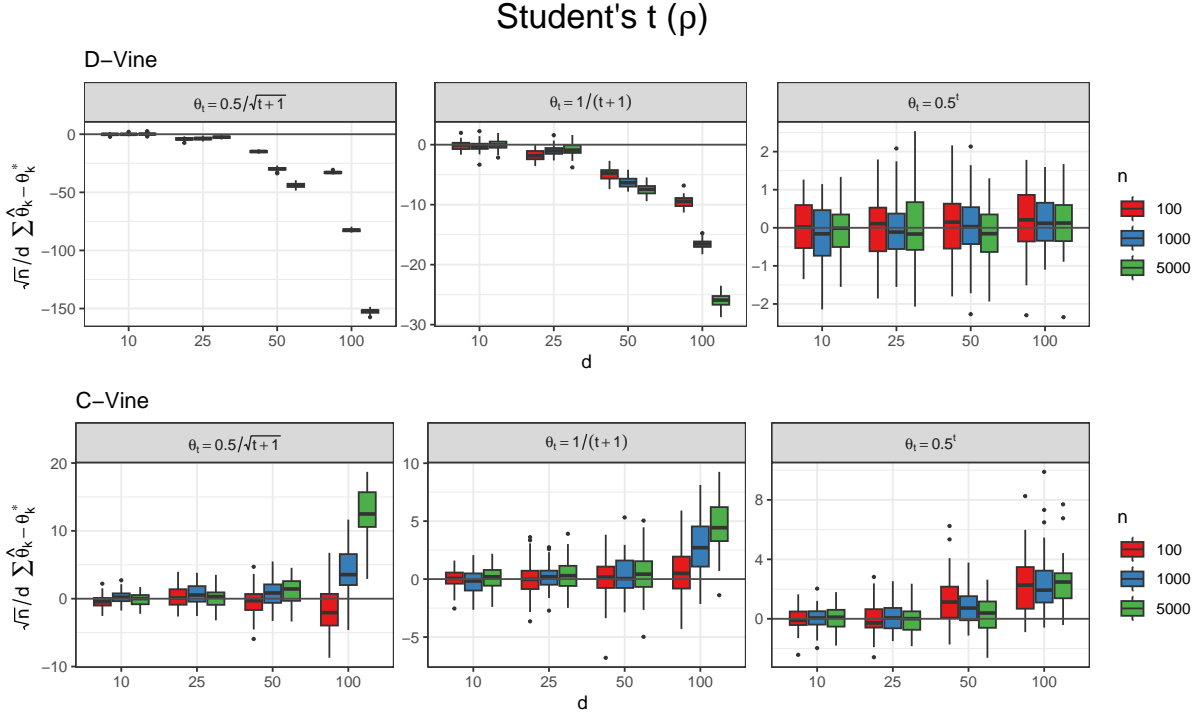


Figure 8: Parameter estimation for Student's t vines, sum of estimation errors. Each boxplot represents $N = 100$ replications.

Student's t Vines We now turn to the estimation of Student's t vines, which is shown in Fig. 7 and 8. Fig. 7 indicates that consistent estimation of ρ is possible in certain settings (D-vine, $\rho_t = 1/(t+1)$ and $\rho_t = 0.5^t$), whereas the degrees of freedom parameter ν cannot be estimated consistently. Fig. 8 shows that $\hat{\rho}$ is biased in almost all settings, with even larger biases for $\hat{\nu}$ (not shown here). Unlike for Gumbel and Gaussian vines, the bivariate copulas do not converge to independence copulas in this setting, which probably causes the observed biases. Nevertheless, the upper right panel (D-vine, $\rho_t = 0.5^t$) suggests that there are settings in which the estimation of the correlation parameter ρ can still be asymptotically normal with mean zero.

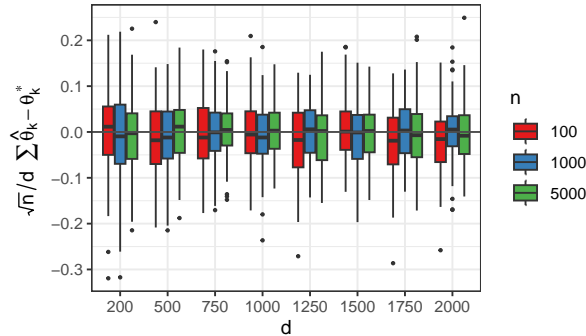


Figure 9: 2-truncated C-vine with $\theta_t^* = 1/(t+1)$, sum of estimation errors. Each boxplot represents $N = 100$ replications.

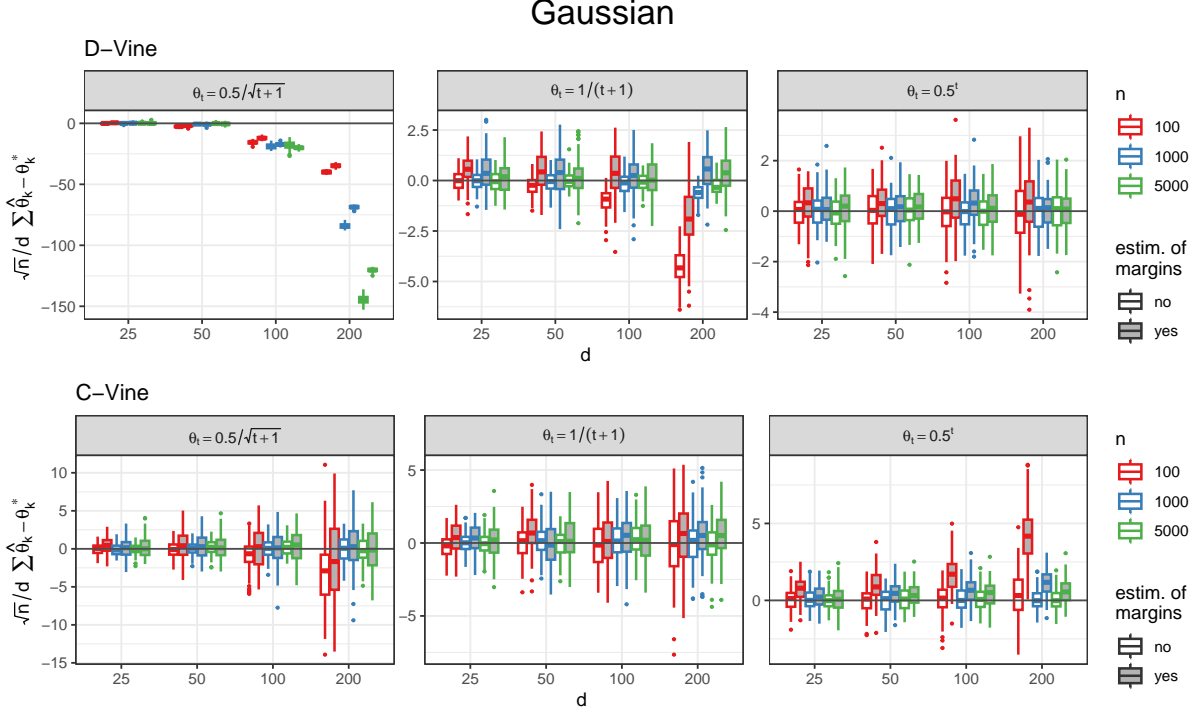


Figure 10: Parameter estimation for Gaussian vines with nonparametric estimation of margins, sum of estimation errors. Each boxplot represents $N = 100$ replications.

Number of Parameters and Truncated Vines In many of the investigated models, we encounter settings in which the number of parameters p exceeds the sample size n , yet consistent estimation of θ^* remains possible. This is in line with [Theorem 1](#), which does not explicitly require $p_n/n \rightarrow 0$. Instead, restrictions on the rate of growth of p_n are implicitly imposed, mainly by [\(A4\)](#). While the results in [Section 4](#) indicate that $p_n^2/n \rightarrow 0$ is sufficient for [\(A4\)](#), the simulation study suggests that this restriction can be relaxed in practice.

Another interesting application of the theoretical results arises in truncated vines with copula dimension $d_n \rightarrow \infty$ and a fixed truncation level. In this setting, [Theorem 1](#) only requires $\ln d_n/n \rightarrow 0$ under standard conditions. [Fig. 9](#) shows results of a simulation from a Gaussian C-vine with $\theta_t = 1/(t+1)$ ($t = 1, 2$) that is truncated after the second tree. The truncation allows computations in very high dimensions, here up to $d = 2000$. Even in such high-dimensional settings, the sum of errors stabilizes at a comparatively small sample size.

5.2 Nonparametric Estimation of Margins

We now compare $\hat{\theta}_U$, the estimator based on known margins, to $\hat{\theta}_X$, where the copula data is estimated from the empirical distribution functions $F_n(\mathbf{X}_i)$, which induces additional uncertainty. All univariate margins are standard normal, and all remaining parameters are the same as above.

[Fig. 10](#) shows the sum of errors for Gaussian vines for both $\hat{\theta}_U$ and $\hat{\theta}_X$. For D-vines, we only observe a small increase in variance, whereas the estimation based on nonparametric margins in high-dimensional C-vines is slightly biased even for large n . The corresponding results for Gumbel and Student's t vines as well as truncated Gaussian vines exhibit similar patterns and are therefore not shown here.

6 Discussion

In this work, we provide asymptotic results for stepwise parameter estimation in high-dimensional vine copulas. The theory covers both parametric and nonparametric estimation of margins as well as truncated vines, and is not restricted to specific R-vine structures or parametric families. Moreover, it is applicable to any parametric estimation problem whose solution can be expressed as the root of a system of estimating equations.

The theoretical results are supported by an empirical validation of the main assumptions. The findings suggest that the derived conditions hold if the pair copulas in higher trees converge sufficiently fast to independence copulas and $p_n = o(\sqrt{n})$, which may be too restrictive for many practical applications. The simulation study indicates that, in practice, inference on vine copula parameters is often feasible even when the dimension and therefore the number of parameters is relatively high compared to the sample size, e.g., $p_n \sim n$.

In applications, estimation of vine copula parameters often comes with the challenge of selecting an appropriate vine tree structure and pair copula families. Choosing between different copula families is typically done by minimizing AIC or BIC, with a modified version tailored for high-dimensional vines developed by Nagler et al. (2019). As the number of possible vine structures grows superexponentially (Morales-Nápoles, 2010), maximizing such criteria to also determine the vine structure is infeasible. The popular algorithm by Dißmann et al. (2013) greedily builds the tree structure by fitting the strongest dependencies first, typically resulting in (almost) independence in higher trees. The validity of this approach is supported by our derived conditions for consistent estimation, which require decreasing dependence in higher trees.

While we only cover parametric estimation of vine copulas, our results are also applicable to nonparametric approaches such as B-splines (Kauermann and Schellhase, 2014) or Bernstein copulas (Scheffer and Weiß (2017), see Nagler et al. (2017) for an overview and comparison of nonparametric methods). In such cases, it is natural to increase the number of basis functions as n increases. Bias-variance tradeoff considerations typically yield rules of the form n^δ parameters per pair copula for some small δ . If M_n in (A4) is sufficiently small, our theory ensures consistent estimation of $O(d^2 n^\delta)$ parameters. The smaller M_n , the more we can relax the assumptions on d_n . Another use case with $p \rightarrow \infty$ as $n \rightarrow \infty$, even when $d = O(1)$, arises in non-simplified vines. Here, our results can be extended to approaches such as the nonparametric method by Schellhase and Spanhel (2018), as well as models in which the parameters $\theta_{a_e, b_e; D_e}(\mathbf{u}_{D_e})$ are estimated using GLMs (Han et al., 2017) or splines (Vatter and Nagler, 2018). Working out the assumptions in detail in these special cases is beyond the scope of this paper.

Many existing methods for high-dimensional vines aim to induce sparsity, such as truncated or thresholded vines (see Section 1), which could also be achieved by adding a suitable penalty term to the estimating function. To the best of our knowledge, penalized estimation of parametric vines has not yet been studied, and extending the results on high-dimensional penalized estimation in Gauss and Nagler (2025) remains an open problem for future research.

Code availability The code and data to reproduce the results can be found on Github: https://github.com/JanaGauss/high_dimensional_vines

References

- Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4).
- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.
- Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1/4):245–268.
- Bedford, T. and Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4).
- Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 40(1):68–85.
- Brechmann, E. C. and Joe, H. (2015). Truncation of vine copulas using fit indices. *Journal of Multivariate Analysis*, 138:19–33.
- Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*. Springer International Publishing.
- Czado, C. and Nagler, T. (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(1):453–477.
- Dißmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.
- Erhardt, V. and Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal*, 2012(2):106–129.
- Gauss, J. and Nagler, T. (2025). Asymptotics for estimating a diverging number of parameters – with and without sparsity.
- Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:347–368.
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.
- Han, D., Tan, K. S., and Weng, C. (2017). Vine copula models with glm and sparsity. *Communications in Statistics - Theory and Methods*, 46(13):6358–6381.
- Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli*, 19(2):462 – 491.

- Hobæk Haff, I., Frigessi, A., and Maraun, D. (2015). How well do regional climate models simulate the spatial dependence of precipitation? an application of pair-copula constructions. *Journal of Geophysical Research: Atmospheres*, 120(7):2624–2646.
- Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325.
- Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141.
- Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.
- Kauermann, G. and Schellhase, C. (2014). Flexible pair-copula estimation in d-vines using bivariate penalized splines. *Statistics and Computing*, 24(6):1081–1100.
- Kiriliouk, A., Lee, J., and Segers, J. (2024). X-vine models for multivariate extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(3):579–602.
- Kulpa, W. (1997). The poincaré-miranda theorem. *The American Mathematical Monthly*, 104(6):545–550.
- Kurowicka, D. (2010). *Optimal Truncation of Vines*, page 233–247. WORLD SCIENTIFIC.
- Kurowicka, D. and Cooke, R. (2003). A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251.
- Morales-Nápoles, O. (2010). *Counting Vines*, pages 189–218.
- Möller, A., Spazzini, L., Kraus, D., Nagler, T., and Czado, C. (2018). Vine copula based post-processing of ensemble forecasts for temperature.
- Müller, D. and Czado, C. (2018). Representing sparse gaussian dags as sparse r-vines allowing for non-gaussian dependence. *Journal of Computational and Graphical Statistics*, 27(2):334–344.
- Müller, D. and Czado, C. (2019a). Dependence modelling in ultra high dimensions with vine copulas and the graphical lasso. *Computational Statistics & Data Analysis*, 137:211–232.
- Müller, D. and Czado, C. (2019b). Selection of sparse vine copulas in high dimensions with the lasso. *Statistics and Computing*, 29(2):269–287.
- Nagler, T., Bumann, C., and Czado, C. (2019). Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *Journal of Multivariate Analysis*, 172:180–192.
- Nagler, T., Krüger, D., and Min, A. (2022). Stationary vine copula models for multivariate time series. *Journal of Econometrics*, 227(2):305–324.
- Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods. *Dependence Modeling*, 5(1):99–120.
- Nelsen, R. (2007). *An Introduction to Copulas*. Springer Series in Statistics. Springer New York.
- Scheffer, M. and Weiß, G. N. F. (2017). Smooth nonparametric bernstein vine copulas. *Quantitative Finance*, 17(1):139–156.

- Schellhase, C. and Spanhel, F. (2018). Estimating non-simplified vine copulas using penalized splines. *Statistics and Computing*, 28(2):387–409.
- Schepsmeier, U. and Stöber, J. (2014). Derivatives and fisher information of bivariate copulas. *Statistical Papers*, 55(2):525–542.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP*, VIII(3):229–231.
- Stöber, J. and Schepsmeier, U. (2013). Estimating standard errors in regular vine copula models. *Computational Statistics*, 28(6):2679–2707.
- Stöber, J., Hong, H. G., Czado, C., and Ghosh, P. (2015). Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses. *Computational Statistics & Data Analysis*, 88:28–39.
- Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33(3):357–375.
- Van der Vaart, A. and Wellner, J. A. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Nature.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- Vatter, T. and Nagler, T. (2018). Generalized additive models for pair-copula constructions. *Journal of Computational and Graphical Statistics*, 27(4):715–727.
- Xue, L. and Zou, H. (2014). Optimal estimation of sparse correlation matrices of semiparametric gaussian copulas. *Statistics and Its Interface*, 7(2):201–209.

A $J(\theta)$ for Gaussian C-Vines

The following proposition provides a condition necessary to ensure that all eigenvalues of $J(\theta^*) + J(\theta^*)^\top$ are negative for a Gaussian C-vine. For Gaussian vines, the parameters coincide with the partial correlations $\rho_{a,b;D}$ (Kurowicka and Cooke, 2003), so we will use partial correlations instead of $\theta_{a,b;D}$ for convenience.

Both our theoretical results (Section 3) and the simulation study (Section 5) indicate that negative definiteness of $J(\theta^*) + J(\theta^*)^\top$ is not required for consistent estimation of θ^* . This proposition thus shows that the results in Gauss and Nagler (2025) are not suitable for vine copulas, as they rely on the eigenvalues of $J(\theta^*) + J(\theta^*)^\top$ being negative and bounded away from 0.

Proposition 1. *Consider a d_n -dimensional Gaussian C-vine with root nodes 1 and (1, 2) in the first and second tree. Assume that the true parameters in the second tree satisfy $|\rho_{23;1}^*| = |\rho_{24;1}^*| = \dots = |\rho_{2d;1}^*| \neq 0$. For the condition*

$$\limsup_{d \rightarrow \infty} \lambda_{\max} (J(\theta^*) + J(\theta^*)^\top) \leq -c < 0$$

to hold, the true parameters ρ_{12}^ or $\rho_{2i;1}^*$ must decay with d_n . Specifically, it is necessary that*

$$\frac{\rho_{12}^* \rho_{2i;1}^*}{(1 - \rho_{2i;1}^{*2})(1 - \rho_{12}^{*2})} = O\left(\frac{1}{\sqrt{d_n}}\right).$$

For instance, this condition is satisfied if $\rho_{12}^* = O(d_n^{-c})$ and $\rho_{2i;1}^* = O(d_n^{-c})$ with $c \geq 0.25$, or if $\rho_{12}^* = O(1)$ and $\rho_{2i;1}^* = O(d_n^{-c})$ with $c \geq 0.5$.

Proof. We parameterize the density using the transformed variables $X_{a|D} = \Phi^{-1}(U_{a|D})$. Let 1 and (1, 2) be the root nodes in the first and second tree, i.e., the first tree contains the copulas $c_{1i}, i = 2, \dots, d_n$, and the second tree contains the copulas $c_{2i;1}, i = 3, \dots, d_n$. To simplify notation, we first assume that $\rho_{12}^* > 0$ and that all true parameters $\rho_{2i;1}^*, i = 3, \dots, d_n$ in the second tree are equal and positive, and then extend the proof to $|\rho_{23;1}^*| = |\rho_{24;1}^*| = \dots = |\rho_{2d;1}^*| \neq 0$.

Since

$$\frac{1}{2} \lambda_{\max} (J(\theta^*) + J(\theta^*)^\top) = \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top J(\theta^*) \mathbf{u},$$

we derive conditions necessary for $\sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top J(\theta^*) \mathbf{u}$ not to diverge to $+\infty$. Let \mathbf{u} be such that the entry corresponding to the estimating equation of ρ_{12} is $u_1 > 0$, and the $d_n - 2$ entries corresponding to the equations of $\rho_{2i;1}$ are all $u_2 < 0$. All other entries of \mathbf{u} are set to zero. Denote

$$s(u_1, u_2; \theta) := \frac{\partial \log c(u_1, u_2; \theta)}{\partial \theta}$$

and define

$$a_1 := \mathbb{E} \left[\frac{\partial s_{12}(X_1, X_2; \rho_{12})}{\partial \rho_{12}} \right], \quad a_2 := \mathbb{E} \left[\frac{\partial s_{2i;1}(X_{2|1}, X_{i|1}; \rho_{2i;1})}{\partial \rho_{2i;1}} \right], \quad b := \mathbb{E} \left[\frac{\partial s_{2i;1}(X_{2|1}, X_{i|1}; \rho_{2i;1})}{\partial \rho_{12}} \right].$$

Since all $\rho_{2i;1}$ are the same, a_2 and b do not depend on i . As $\partial s_{2i;1}(X_{2|1}, X_{i|1}; \rho_{2i;1}) / \partial \rho_{2j;1} = 0$ for $i \neq j$, we have

$$\begin{aligned} \mathbf{u}^\top J(\theta^*) \mathbf{u} &= u_1^2 a_1 + (d_n - 2) u_2^2 a_2 + (d_n - 2) u_1 u_2 b \\ &= u_1^2 a_1 + (1 - u_1^2) a_2 - \sqrt{d_n - 2} u_1 \sqrt{1 - u_1^2} b, \end{aligned}$$

where the second equality uses that $\|\mathbf{u}\| = 1$ implies $u_2 = -\sqrt{(1 - u_1^2)/(d_n - 2)}$. The first two terms are negative (since $a_1, a_2 < 0$, see [Lemma 1](#)) and do not depend on d_n if u_1 is fixed. Since

$$b = -\frac{\rho_{12}^* \rho_{2i;1}^*}{(1 - \rho_{2i;1}^{*2})(1 - \rho_{12}^{*2})} < 0$$

by [Lemma 1](#) and $\rho_{12}^*, \rho_{2i;1}^* > 0$, the term $-\sqrt{d_n - 2}b$ diverges to $+\infty$ for $d_n \rightarrow \infty$ unless $b = O(1/\sqrt{d_n})$. The proof can be easily extended to the case where no assumptions are made on the sign of ρ_{12}^* and where $|\rho_{23;1}^*| = |\rho_{24;1}^*| = \dots = |\rho_{2d_n;1}^*| \neq 0$. \square

B Proofs of the Main Results

To simplify notation, we use the following abbreviations:

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(\mathbf{X}_i), \quad Pf = \mathbb{E}[f(\mathbf{X})]$$

and

$$\mathbb{P}_n \phi_X(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \phi(F_n(\mathbf{X}_i); \boldsymbol{\theta}) \quad \text{and} \quad \mathbb{P}_n \phi_U(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \phi(F(\mathbf{X}_i); \boldsymbol{\theta}),$$

wheres

$$F(\mathbf{X}_i) = (F_1(X_{i,1}), \dots, F_{d_n}(X_{i,d_n})), \quad F_n(\mathbf{X}_i) = (F_{n1}(X_{i,1}), \dots, F_{nd_n}(X_{i,d_n}))$$

and $F_{nk}(x) = (n+1)^{-1} \sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x)$.

$\|\cdot\|$ denotes the Euclidean norm for vectors and the spectral norm $\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$ for matrices. Define $r_n := \sqrt{\ln p_n/n}$.

B.1 Proof of [Theorem 1](#)

Denote $\Theta_n^\Delta = \{\boldsymbol{\Delta} : |\Delta_j| \leq \alpha_{j,n} \forall j = 1, \dots, p_n\} \subset \mathbb{R}^{p_n}$. The *Poincaré-Miranda theorem* (see for example [Kulpa, 1997](#)) implies that if

$$\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} \text{sign}(\Delta_j) \mathbb{P}_n \phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j \leq 0$$

holds, there is a solution $\hat{\boldsymbol{\theta}}_U$ of $\mathbb{P}_n \phi_U(\boldsymbol{\theta}) = \mathbf{0}$ with $|\hat{\theta}_{j,U} - \theta_j^*| \leq r_n C \alpha_{j,n}$ for all $j = 1, \dots, p_n$. We show that, by choosing C large enough, the probability of

$$\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n \phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j < 0 \quad (8)$$

gets arbitrarily close to 1. This implies that, with probability tending to 1, the sets Θ_n contain a solution $\hat{\boldsymbol{\theta}}_U$ that satisfies $\max_{1 \leq j \leq p_n} |\hat{\theta}_{j,U} - \theta_j^*|/\alpha_{j,n} = O_p(r_n)$. Denote $\underline{\alpha}_n = \min_{1 \leq j \leq p_n} \alpha_{j,n}$. We have

$$\begin{aligned} & \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n \phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j \\ & \leq \max_{1 \leq j \leq p_n} (r_n C \alpha_{j,n})^{-1} |\mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)_j| \\ & \quad + \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n [\phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j]. \end{aligned} \quad (9)$$

The first term is of order $O_p((\underline{\alpha}_n C)^{-1})$ since $\|\mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_\infty = O_p(\sqrt{\ln p_n/n})$ by (A1), (A2) and Lemma 1 in Gauss and Nagler (2025). For the second term, we have, with some $c > 0$ and k_n as defined in (A4),

$$\begin{aligned} & \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n [\phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j] \\ & \leq \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) P [\phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j] + \\ & \quad \max_{1 \leq j \leq p_n} \sup_{\|\boldsymbol{\Delta}\|_\infty \leq \alpha} (r_n C \alpha_{j,n})^{-1} |(\mathbb{P}_n - P) [\phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j]| \\ & \leq -c + O_p \left(\frac{M_n \sqrt{k_n + \ln p_n}}{\sqrt{n}} + \frac{D_n(k_n + \ln p_n)}{n} \right) = -c + o_p(1) \end{aligned}$$

by (A3), (A4) and Lemma 2. Choosing C large enough, the second term in (9) dominates the first since $\underline{\alpha}_n$ is bounded away from 0. This implies the claim.

B.2 Proof of Theorem 2

Note that this is the same as the proof of Theorem 3 in Gauss and Nagler (2025).

Define $\tilde{r}_n = \alpha_n \sqrt{p_n \ln p_n/n}$. We have

$$\begin{aligned} \mathbf{0} &= \mathbb{P}_n \phi(\hat{\boldsymbol{\theta}}_U) = \mathbb{P}_n \phi(\boldsymbol{\theta}^*) + \mathbb{P}_n [\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)] \\ &= \mathbb{P}_n \phi(\boldsymbol{\theta}^*) + P[\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)] + (\mathbb{P}_n - P)[\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)] \\ &= \mathbb{P}_n \phi(\boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} P \phi(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) + (\mathbb{P}_n - P)[\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)], \end{aligned}$$

with some $\tilde{\boldsymbol{\theta}}$ on the segment between $\hat{\boldsymbol{\theta}}_U$ and $\boldsymbol{\theta}^*$. We have $\nabla_{\boldsymbol{\theta}} P \phi(\tilde{\boldsymbol{\theta}}) = J(\tilde{\boldsymbol{\theta}})$, so

$$-J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) = \mathbb{P}_n \phi(\boldsymbol{\theta}^*) + [J(\tilde{\boldsymbol{\theta}}) - J(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) + (\mathbb{P}_n - P)[\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)]$$

and

$$\begin{aligned} -\sqrt{n} A_n J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) &= \sqrt{n} A_n \mathbb{P}_n \phi(\boldsymbol{\theta}^*) \\ &\quad + \sqrt{n} A_n [J(\tilde{\boldsymbol{\theta}}) - J(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) \\ &\quad + \sqrt{n} (\mathbb{P}_n - P) A_n [\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)]. \end{aligned}$$

The second and the third term are negligible, since

$$\sqrt{n} A_n [J(\tilde{\boldsymbol{\theta}}) - J(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) = o_p \left(\sqrt{n} \frac{1}{\sqrt{n} \tilde{r}_n} \tilde{r}_n \right) = o_p(1)$$

by assumption (A5), and Lemma 10 in Gauss and Nagler (2025) together with (A5) yields

$$\sqrt{n} (\mathbb{P}_n - P) A_n [\phi(\hat{\boldsymbol{\theta}}_U) - \phi(\boldsymbol{\theta}^*)] = o_p(1).$$

It remains to prove a central limit theorem for

$$\sqrt{n} A_n \mathbb{P}_n \phi(\boldsymbol{\theta}^*) = \sum_{i=1}^n \frac{1}{\sqrt{n}} A_n \phi_i(\boldsymbol{\theta}^*) := \sum_{i=1}^n \mathbf{Y}_i.$$

Since

$$\sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^2 \mathbb{1}\{\|\mathbf{Y}_i\| > \varepsilon\}] \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^2 \mathbb{1}\{\|\mathbf{Y}_i\| > \varepsilon\} \|\mathbf{Y}_i\|^2 / \varepsilon^2] \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^4] / \varepsilon^2,$$

and $\mathbb{E}[\|\mathbf{Y}_i\|^4] = n^{-2}\mathbb{E}[\|A_n\phi_i(\boldsymbol{\theta}^*)\|^4] = o(n^{-1})$ for all $i = 1, \dots, n$, by (A6), we have

$$\sum_{i=1}^n \mathbb{E}[\|\mathbf{Y}_i\|^2 \mathbb{1}\{\|\mathbf{Y}_i\| > \varepsilon\}] \rightarrow 0 \text{ for every } \varepsilon > 0.$$

Since $\mathbb{E}[\mathbf{Y}_i] = \mathbf{0}$ for all $i = 1, \dots, n$ and

$$\sum_{i=1}^n \text{Cov}(\mathbf{Y}_i) = \frac{1}{n} \sum_{i=1}^n \text{Cov}[A_n\phi_i(\boldsymbol{\theta}^*)] = \frac{1}{n} A_n \sum_{i=1}^n \text{Cov}[\phi_i(\boldsymbol{\theta}^*)] A_n^\top = A_n I(\boldsymbol{\theta}^*) A_n^\top \rightarrow \Sigma,$$

the conditions of the Lindeberg-Feller central limit theorem (van der Vaart, 1998, Section 2.8) are satisfied, and we obtain

$$\sqrt{n} A_n J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_U - \boldsymbol{\theta}^*) \rightarrow_d \mathcal{N}(\mathbf{0}, \Sigma).$$

B.3 Proof of Theorem 5

Similar to the proof of Theorem 1, we show that, by choosing C large enough, the probability of

$$\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n \phi_X(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j < 0 \quad (10)$$

gets arbitrarily close to 1. This implies that, with probability tending to 1, the sets Θ_n contain a solution $\hat{\boldsymbol{\theta}}_X$ that satisfies $\max_{1 \leq j \leq p_n} |\hat{\theta}_{X,j} - \theta_j^*| / \alpha_{j,n} = O_p(r_n)$.

It holds that

$$\begin{aligned} & \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n \phi_X(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j \\ & \leq \max_{1 \leq j \leq p_n} (r_n C \alpha_{j,n})^{-1} |\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*)_j| \\ & \quad + \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n [\phi_X(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_X(\boldsymbol{\theta}^*)_j]. \end{aligned} \quad (11)$$

For the first term, we have

$$(r_n C \alpha_{j,n})^{-1} \|\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*)\|_\infty \leq (r_n C \alpha_{j,n})^{-1} \|\mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_\infty + (r_n C \alpha_{j,n})^{-1} \|\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) - \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_\infty.$$

The first term is of order $O_p((\alpha_n C)^{-1})$ by (A1) and (A2), see the proof of Theorem 1. The second term is also of order $O_p((\alpha_n C)^{-1})$ since $\|\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) - \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_\infty = O_p(\sqrt{\ln p_n / n})$ by Lemma 3, (A7), (A8), (A9) and $d_n^3/n \rightarrow 0$.

For the second term in (11), we have

$$\begin{aligned} & \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n [\phi_X(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_X(\boldsymbol{\theta}^*)_j] \\ & \leq \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} \text{sign}(\Delta_j) \mathbb{P}_n [\phi_U(\boldsymbol{\theta}^* + r_n C \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j] \\ & \quad + \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\Delta} \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} |\mathbb{P}_n f(r_n C \boldsymbol{\Delta}, j)|, \end{aligned}$$

where

$$f_j(\boldsymbol{\Delta}) := \phi_X(\boldsymbol{\theta}^* + \boldsymbol{\Delta})_j - \phi_X(\boldsymbol{\theta}^*)_j - (\phi_U(\boldsymbol{\theta}^* + \boldsymbol{\Delta})_j - \phi_U(\boldsymbol{\theta}^*)_j). \quad (12)$$

The first term remains below some $-c < 0$ with probability tending to 1 by (A3) and (A4), see the proof of Theorem 1. The second term is $o_p(1)$ by Lemma 6, (A10), (A11) and $d_n^2/n \rightarrow 0$. This implies the claim.

B.4 Proof of Theorem 6

It holds that

$$\begin{aligned} \mathbf{0} &= \mathbb{P}_n \phi_X(\hat{\boldsymbol{\theta}}_X) = \mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) + \mathbb{P}_n [\phi_X(\hat{\boldsymbol{\theta}}_X) - \phi_X(\boldsymbol{\theta}^*)] \\ &= \mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) + \mathbb{P}_n [\phi_U(\hat{\boldsymbol{\theta}}_X) - \phi_U(\boldsymbol{\theta}^*)] + \mathbb{P}_n [\phi_X(\hat{\boldsymbol{\theta}}_X) - \phi_X(\boldsymbol{\theta}^*) - (\phi_U(\hat{\boldsymbol{\theta}}_X) - \phi_U(\boldsymbol{\theta}^*))]. \end{aligned}$$

Similar to the proof of Theorem 2, we obtain

$$\begin{aligned} -\sqrt{n}A_n J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_X - \boldsymbol{\theta}^*) &= \sqrt{n}A_n \mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) \\ &\quad + \sqrt{n}A_n [J(\tilde{\boldsymbol{\theta}}) - J(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}}_X - \boldsymbol{\theta}^*) \\ &\quad + \sqrt{n}(\mathbb{P}_n - P)A_n [\phi_U(\hat{\boldsymbol{\theta}}_X) - \phi_U(\boldsymbol{\theta}^*)] \\ &\quad + \sqrt{n}A_n \mathbb{P}_n [\phi_X(\hat{\boldsymbol{\theta}}_X) - \phi_X(\boldsymbol{\theta}^*) - (\phi_U(\hat{\boldsymbol{\theta}}_X) - \phi_U(\boldsymbol{\theta}^*))], \end{aligned}$$

with some $\tilde{\boldsymbol{\theta}}$ on the segment between $\boldsymbol{\theta}^*$, $\hat{\boldsymbol{\theta}}_X$. The second and third term are $o_p(1)$ by (A5), see the proof of Theorem 2. For the last term, Lemma 6, $\|\hat{\boldsymbol{\theta}}_X - \boldsymbol{\theta}^*\|_\infty = O_p(\alpha_n \sqrt{\ln p_n/n})$ and $\|\mathbf{x}\|_2 \leq \sqrt{p}\|\mathbf{x}\|_\infty$ for $\mathbf{x} \in \mathbb{R}^p$ yield

$$\sqrt{n}A_n \mathbb{P}_n [\phi_X(\hat{\boldsymbol{\theta}}_X) - \phi_X(\boldsymbol{\theta}^*) - (\phi_U(\hat{\boldsymbol{\theta}}_X) - \phi_U(\boldsymbol{\theta}^*))] = O_p \left(\alpha_n \sqrt{\frac{p_n d_n^2 \ln p_n}{n}} \right) = o_p(1)$$

since $\alpha_n^2 p_n d_n^2 \ln p_n/n \rightarrow 0$. It remains to show a central limit theorem for $\sqrt{n}A_n \mathbb{P}_n \phi_X(\boldsymbol{\theta}^*)$. It holds that

$$\begin{aligned} \mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) &= \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*) + [\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) - \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)] \\ &= \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*) + \mathbb{P}_n \nabla_{\mathbf{u}} \phi(F(\cdot); \boldsymbol{\theta}^*)(F_n(\cdot) - F(\cdot)) + I_2 \\ &= \tilde{I}_1 + I_2 \end{aligned}$$

with I_2 as defined in (16). Lemma 7 and (A12) give $\|I_2\| = O_p(d_n \sqrt{p_n}/n)$, so $\sqrt{n}A_n I_2 = o_p(1)$ since $\|A_n\| = O(1)$ and $d_n^2 p_n/n \rightarrow 0$.

\tilde{I}_1 can be treated as a U-statistics (see for example van der Vaart, 1998). Define

$$\tilde{h}(\mathbf{X}_i, \mathbf{X}_{i'}) := \phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*) + \sum_{l=1}^{d_n} \frac{\partial}{\partial u_l} \phi(F(\mathbf{X}_{i'}); \boldsymbol{\theta}^*) (\mathbb{1}(X_{il} \leq X_{i'l}) - F_l(X_{i'l})) \in \mathbb{R}^{p_n}, \quad (13)$$

where $\frac{\partial}{\partial u_l} \phi(F(\mathbf{X}_{i'}); \boldsymbol{\theta}^*)$ denotes the \mathbb{R}^{p_n} dimensional vector with entries $\frac{\partial}{\partial u_l} \phi(F(\mathbf{X}_{i'}); \boldsymbol{\theta}^*)_k, k = 1, \dots, p_n$. Note that the second term in (13) is the same as $\mathbf{Y}_{i'i}$ as defined in (5). Then

$$\frac{n+1}{n-1} \tilde{I}_1 = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n \tilde{h}(\mathbf{X}_i, \mathbf{X}_{i'}) + \frac{1}{n(n-1)} \sum_{i=1}^n \mathbf{Y}_{ii}.$$

The second term is of order $O_p(\sqrt{p_n}/n)$ by (A8), see Lemma 3. Multiplying this term with $\sqrt{n}A_n$ yields an $o_p(1)$ term since $p_n/n \rightarrow 0$ and $\|A_n\| = O(1)$.

Now define the symmetric kernel

$$\bar{h}(\mathbf{X}_i, \mathbf{X}_{i'}) = \frac{1}{2} \tilde{h}(\mathbf{X}_i, \mathbf{X}_{i'}) + \frac{1}{2} \tilde{h}(\mathbf{X}_{i'}, \mathbf{X}_i) \quad (14)$$

and

$$\bar{h}_1(\mathbf{x}) = \mathbb{E}[\bar{h}(\mathbf{x}, \mathbf{X})] = \frac{1}{2} \phi(F(\mathbf{x}); \boldsymbol{\theta}^*) + \frac{1}{2} \mathbb{E} \left[\sum_{l=1}^{d_n} \frac{\partial}{\partial u_l} \phi(F(\mathbf{X}); \boldsymbol{\theta}^*) (\mathbb{1}(x_l \leq X_l) - F_l(X_l)) \right], \quad (15)$$

since $\mathbb{E}[\tilde{h}(\mathbf{X}, \mathbf{x})] = \mathbf{0}$, as $\mathbb{E}[\mathbb{1}(X_{il} \leq X_{i'l}) - F_1(X_{i'l}) | X_{i'l}] = 0$ for all $l = 1, \dots, d_n$, $i, i' = 1, \dots, n$. Note that also $\mathbb{E}[\bar{h}_1(\mathbf{X})] = \mathbf{0}$. Then, *Hoeffding's decomposition* (Hoeffding, 1948) yields

$$\begin{aligned} \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n \tilde{h}(\mathbf{X}_i, \mathbf{X}_{i'}) &= \frac{2}{n} \sum_{i=1}^n \bar{h}_1(\mathbf{X}_i) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n \bar{h}(\mathbf{X}_i, \mathbf{X}_{i'}) - \bar{h}_1(\mathbf{X}_i) - \bar{h}_1(\mathbf{X}_{i'}) \\ &= I + II. \end{aligned}$$

Lemma 4, (A7) and (A8) yield $\sqrt{n}A_nII = o_p(1)$ since $p_n/n \rightarrow 0$ and $\|A_n\| = O(1)$. We can apply Lemma 4 here despite the difference between $\bar{h}(\mathbf{X}_i, \mathbf{X}_{i'})$ and $h(\mathbf{X}_i, \mathbf{X}_{i'})$ as defined in (6), since the terms $\phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*), \phi(F(\mathbf{X}_{i'}); \boldsymbol{\theta}^*)$ in $\bar{h}_1(\mathbf{X}_i), \bar{h}_1(\mathbf{X}_{i'})$ cancel with the same terms in $\bar{h}(\mathbf{X}_i, \mathbf{X}_{i'})$.

It remains to show a central limit theorem for $\sum_{i=1}^n 2/\sqrt{n} A_n \bar{h}_1(\mathbf{X}_i) =: \sum_{i=1}^n \mathbf{Y}_i$. We write

$$\bar{h}_1(\mathbf{x}) = \frac{1}{2} \phi(F(\mathbf{x}); \boldsymbol{\theta}^*) + \frac{1}{2} \int \sum_{l=1}^{d_n} \frac{\partial}{\partial u_l} \phi(\mathbf{u}; \boldsymbol{\theta}^*) (\mathbb{1}(F(x_l) \leq u_l) - u_l) dC(\mathbf{u}),$$

where C is the copula distribution of \mathbf{X} , i.e., the distribution of $\mathbf{U} = F(\mathbf{X})$. It holds that

$$\text{Cov}(2\bar{h}_1(\mathbf{X})) = \text{Cov} \left(\phi(\boldsymbol{\xi}; \boldsymbol{\theta}^*) + \int \sum_{l=1}^{d_n} \frac{\partial}{\partial u_l} \phi(\mathbf{u}; \boldsymbol{\theta}^*) (\mathbb{1}(\xi_l \leq u_l) - u_l) dC(\mathbf{u}) \right),$$

where $\boldsymbol{\xi}$ is a random variable with distribution C . Therefore

$$\lim_{n \rightarrow \infty} \text{Cov} \left(\sum_{i=1}^n \mathbf{Y}_i \right) = \lim_{n \rightarrow \infty} A_n \text{Cov}(2\bar{h}_1(\mathbf{X})) A_n^\top = \Sigma_X$$

with Σ_X as defined in Theorem 6. Since $\mathbb{E}[\mathbf{Y}_i] = \mathbf{0}$ as $\mathbb{E}[\bar{h}_1(\mathbf{X}_i)] = \mathbf{0}$ and

$$\sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^2 \mathbb{1}\{\|\mathbf{Y}_i\| > \varepsilon\}] \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^2 \mathbb{1}\{\|\mathbf{Y}_i\| > \varepsilon\} \|\mathbf{Y}_i\|^2 / \varepsilon^2] \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{Y}_i\|^4] / \varepsilon^2 = o(1)$$

for every $\varepsilon > 0$ by (A6) and (A13), the conditions of the Lindeberg-Feller central limit theorem (van der Vaart, 1998, Section 2.8) are satisfied, and we obtain

$$\sqrt{n}A_n J(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_X - \boldsymbol{\theta}^*) \rightarrow_d \mathcal{N}(\mathbf{0}, \Sigma_X).$$

C Lemmas

Lemma 1. For a Gaussian vine with $X_{a|D} = \phi^{-1}(U_{a|D}), X_{b|D} = \phi^{-1}(U_{b|D})$, it holds

$$\mathbb{E} \left[\frac{\partial s_{ab;D}(X_{a|D}, X_{b|D}; \rho_{ab;D})}{\partial \rho_{ab;D}} \right] = -\frac{1 + \rho_{ab;D}^2}{(1 - \rho_{ab;D}^2)^2}$$

and

$$\mathbb{E} \left[\frac{\partial s_{2i;1}(X_{2|1}, X_{i|1}; \rho_{2i;1})}{\partial \rho_{12}} \right] = -\frac{\rho_{12}\rho_{2i;1}}{(1 - \rho_{2i;1}^2)(1 - \rho_{12}^2)}, \quad i = 3, \dots, d.$$

Proof. In a Gaussian bivariate copula, the log-likelihood of a parameter ρ is given by

$$\log c(u_1, u_2; \rho) = -\frac{1}{2} \log(1 - \rho^2) - \frac{\rho^2(x_1^2 + x_2^2) - 2\rho x_1 x_2}{2(1 - \rho^2)}$$

with $x_i = \Phi^{-1}(u_i)$, see for example [Czado \(2019\)](#) for the copula density. We will write all functions with arguments x_1, x_2 from now on. The derivative w.r.t. ρ is

$$s(x_1, x_2; \rho) := \frac{\partial \log c(x_1, x_2; \rho)}{\partial \rho} = \frac{\rho}{1 - \rho^2} - \frac{\rho}{(1 - \rho^2)^2} (x_1^2 + x_2^2) + \frac{1 + \rho^2}{(1 - \rho^2)^2} x_1 x_2.$$

The derivative of $s(x_1, x_2; \rho)$ w.r.t. ρ is

$$\frac{\partial s(x_1, x_2; \rho)}{\partial \rho} = \frac{1 + \rho^2}{(1 - \rho^2)} - \frac{1 + 3\rho^2}{(1 - \rho^2)^3} (x_1^2 + x_2^2) + 2 \frac{3\rho + \rho^3}{(1 - \rho^2)^3} x_1 x_2$$

and the derivative w.r.t. x_1 is

$$\frac{\partial s(x_1, x_2; \rho)}{\partial x_1} = \frac{(1 + \rho^2)x_2 - 2\rho x_1}{(1 - \rho^2)^2}.$$

The h -function is given by $h(x_1, x_2; \rho) = (x_1 - \rho x_2)/\sqrt{1 - \rho^2}$ (see for example [Schepsmeier and Stöber, 2014](#)), so the derivative of h w.r.t. ρ is

$$\begin{aligned} \frac{\partial h(x_1, x_2; \rho)}{\partial \rho} &= \frac{\rho x_1}{(1 - \rho^2)^{3/2}} - \frac{\rho^2 x_2}{(1 - \rho^2)^{3/2}} - \frac{x_2}{\sqrt{1 - \rho^2}} \\ &= \frac{\rho x_1 - \rho^2 x_2 - (1 - \rho^2)x_2}{(1 - \rho^2)^{3/2}} = \frac{\rho x_1 - x_2}{(1 - \rho^2)^{3/2}} \end{aligned}$$

For the expectation of $\partial s(x_1, x_2; \rho)/\partial \rho$, we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{\partial s(X_1, X_2; \rho)}{\partial \rho} \right] &= \frac{1 + \rho^2}{(1 - \rho^2)} - 2 \frac{1 + 3\rho^2}{(1 - \rho^2)^3} + 2 \frac{3\rho^2 + \rho^4}{(1 - \rho^2)^3} \\ &= \frac{(1 - \rho^2)(1 + \rho^2) - 2 - 6\rho^2 + 6\rho^2 + 2\rho^4}{(1 - \rho^2)^3} = \frac{-1 + \rho^4}{(1 - \rho^2)^3} = -\frac{1 + \rho^2}{(1 - \rho^2)^2} \end{aligned}$$

using $\mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = 1$ and $\mathbb{E}(X_1 X_2) = \rho$ since X_1, X_2 are standard normally distributed with correlation ρ . The statement for pair copulas $c_{ab;D}$ involving conditioning variables D follows immediately since $\mathbb{E}(X_{a|D} X_{b|D}) = \rho_{ab;D}$.

For the second statement, the chain rule yields

$$\begin{aligned} \frac{\partial s_{2i;1}(x_{2|1}, x_{i|1}; \rho_{2i;1})}{\partial \rho_{12}} &= \frac{\partial s_{2i;1}(x_{2|1}, x_{i|1}; \rho_{2i;1})}{\partial x_{2|1}} \cdot \frac{\partial h(x_2, x_1; \rho_{12})}{\partial \rho_{12}} \\ &= \frac{(1 + \rho_{2i;1}^2)x_{i|1} - 2\rho_{2i;1}x_{2|1}}{(1 - \rho_{2i;1}^2)^2} \cdot \frac{\rho_{12}x_2 - x_1}{(1 - \rho_{12}^2)^{3/2}}. \end{aligned}$$

To compute its expectation, we need $\mathbb{E}[X_{i|1} X_2]$, $\mathbb{E}[X_{i|1} X_1]$, $\mathbb{E}[X_{2|1} X_1]$ and $\mathbb{E}[X_{2|1} X_2]$. Plugging

in the definition $x_{a|b} = h(x_a, x_b; \rho_{ab})$ yields

$$\begin{aligned}\mathbb{E}[X_{i|1}X_1] &= \mathbb{E}\left[\frac{X_iX_1 - \rho_{1i}X_1^2}{\sqrt{1 - \rho_{1i}^2}}\right] = \frac{\rho_{1i} - \rho_{1i}}{\sqrt{1 - \rho_{1i}^2}} = 0 = \mathbb{E}[X_{2|1}X_1], \\ \mathbb{E}[X_{2|1}X_2] &= \mathbb{E}\left[\frac{X_2^2 - \rho_{12}X_1X_2}{\sqrt{1 - \rho_{12}^2}}\right] = \frac{1 - \rho_{12}^2}{\sqrt{1 - \rho_{12}^2}} = \sqrt{1 - \rho_{12}^2}, \\ \mathbb{E}[X_{i|1}X_2] &= \mathbb{E}\left[\frac{X_iX_2 - \rho_{1i}X_1X_2}{\sqrt{1 - \rho_{1i}^2}}\right] = \frac{\rho_{2i} - \rho_{1i}\rho_{12}}{\sqrt{1 - \rho_{1i}^2}} = \rho_{2i;1}\sqrt{1 - \rho_{12}^2},\end{aligned}$$

where we used the definition of the partial correlation ([Kurowicka and Cooke, 2003](#)) in the last step:

$$\rho_{2i;1} = \frac{\rho_{2i} - \rho_{1i}\rho_{12}}{\sqrt{(1 - \rho_{12}^2)(1 - \rho_{1i}^2)}}.$$

We now obtain

$$\begin{aligned}\mathbb{E}\left[\frac{\partial s_{2i;1}(X_{2|1}, X_{i|1}; \rho_{2i;1})}{\partial \rho_{12}}\right] &= \mathbb{E}\left[\frac{(1 + \rho_{2i;1}^2)X_{i|1} - 2\rho_{2i;1}X_{2|1}}{(1 - \rho_{2i;1}^2)^2} \cdot \frac{\rho_{12}X_2 - X_1}{(1 - \rho_{12}^2)^{3/2}}\right] \\ &= \frac{(1 + \rho_{2i;1}^2)\rho_{12}\rho_{2i;1}\sqrt{1 - \rho_{12}^2} - 2\rho_{2i;1}\rho_{12}\sqrt{1 - \rho_{12}^2}}{(1 - \rho_{2i;1}^2)^2(1 - \rho_{12}^2)^{3/2}} \\ &= \frac{\rho_{12}\rho_{2i;1}(\rho_{2i;1}^2 - 1)}{(1 - \rho_{2i;1}^2)^2(1 - \rho_{12}^2)} = -\frac{\rho_{12}\rho_{2i;1}}{(1 - \rho_{2i;1}^2)(1 - \rho_{12}^2)}.\end{aligned}$$

□

Lemma 2. *It holds that*

$$\sup_{f \in \mathcal{F}_n} |(\mathbb{P}_n - P)f| = O_p\left(\frac{M_n c_n \sqrt{k_n + \ln p_n}}{\sqrt{n}} + \frac{D_n c_n (k_n + \ln p_n)}{n}\right),$$

where

$$\mathcal{F}_n = \{f_{\Delta,j}(\mathbf{u}) = \alpha_{j,n}^{-1}[\phi(\mathbf{u}; \boldsymbol{\theta}^* + \Delta)_j - \phi(\mathbf{u}; \boldsymbol{\theta}^*)_j] : \forall j' = 1, \dots, p_n : |\Delta_{j'}| \leq c_n \alpha_{j',n}, j = 1, \dots, p_n\}$$

with $c_n = r_n C$ for some $C < \infty$ and sequences D_n, M_n such that

$$\begin{aligned}\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} \mathbb{E}\left[\left(\sum_{k=1}^j \left|\frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{U}; \boldsymbol{\theta})_j\right|\right)^2\right] &\leq M_n^2, \\ \mathbb{P}\left(\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} \sum_{k=1}^j \left|\frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{U}; \boldsymbol{\theta})_j\right| > D_n\right) &= o(1/n),\end{aligned}$$

where the sequence k_n denotes the maximum number of non-zero entries in $\nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j$, i.e.,

$$\max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n, \mathbf{u} \in \mathbb{R}^{d_n}} \|\nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j\|_0 = k_n.$$

Proof. Define

$$F_n(\mathbf{u}) := \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} \sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{u}; \boldsymbol{\theta})_j \right|.$$

F_n is an envelope for the functions in $c_n^{-1} \mathcal{F}_n$, i.e., $\sup_{f \in \mathcal{F}_n} c_n^{-1} |f(\mathbf{u})| \leq F_n(\mathbf{u})$, since a Taylor expansion and Hölder's inequality with $p = 1, q = \infty$ give

$$|f_{\Delta,j}(\mathbf{u})| = \alpha_{j,n}^{-1} |\Delta^T \nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j| \leq c_n \sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{u}; \boldsymbol{\theta})_j \right|$$

with some $\boldsymbol{\theta}$ on the segment between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + \Delta$. Lemma 6 (ii) in [Gauss and Nagler \(2025\)](#) gives

$$\sup_{f \in \mathcal{F}_n} |(\mathbb{P}_n - P)f| \leq \sup_{f \in \mathcal{F}_n} |(\mathbb{P}_n - P)f \mathbf{1}_{F_n \leq D_n}| + o_p \left(\sqrt{\frac{\sup_{f \in \mathcal{F}_n} P f^2}{n}} \right)$$

with D_n as defined above. The second term is $o_p(M_n c_n / \sqrt{n})$ since

$$\sup_{f \in \mathcal{F}_n} P f^2 \leq \max_{1 \leq j \leq p_n} \sup_{\boldsymbol{\theta} \in \Theta_n} c_n^2 \mathbb{E} \left[\left(\sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{U}; \boldsymbol{\theta})_j \right| \right)^2 \right] \leq c_n^2 M_n^2$$

by the definition of M_n .

The following arguments are similar to Lemma 9 in [Gauss and Nagler \(2025\)](#). Fix j and consider $\mathcal{F}_n^{(j)} = \{f_{\Delta,j} \mathbf{1}_{F_n \leq D_n} : \forall j' = 1, \dots, p_n : |\Delta_{j'}| \leq c_n \alpha_{j',n}\}$. We need to bound the covering numbers $N(\varepsilon, \mathcal{F}_n^{(j)}, L_2(P))$ and $N(\varepsilon, \mathcal{F}_n^{(j)}, \|\cdot\|_{\infty})$ (see [Van der Vaart and Wellner, 2023](#), Chapter 2.1.1 for a definition), where $\|f\|_{L_2(P)}^2 = P f^2$ and $\|f\|_{\infty} = \sup_{\mathbf{u}} |f(\mathbf{u})|$.

Denote $\tilde{I}_{n,j} \subseteq \{1, \dots, p_n\}$ the set of indices k such that $\partial \nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j / \partial \theta_k = 0$ for all $\boldsymbol{\theta}$ and \mathbf{u} and denote $I_{n,j} = \{1, \dots, p_n\} \setminus \tilde{I}_{n,j}$. By the above definition of k_n , it holds $|I_{n,j}| \leq k_n$ for each j , i.e., $\nabla_{\boldsymbol{\theta}} \phi(\mathbf{u}; \boldsymbol{\theta})_j$ has at most k_n non-zero entries. It holds

$$\|f_{\Delta,j} - f_{\Delta',j}\|_{L_2(P)}^2 \leq \sup_{\boldsymbol{\theta} \in \Theta_n} \mathbb{E} \left[(\alpha_{j,n}^{-1} (\Delta - \Delta')^T \nabla_{\boldsymbol{\theta}} \phi(\mathbf{U}; \boldsymbol{\theta})_j)^2 \right] \leq M_n^2 \max_{k \in I_{n,j}} \left| \frac{\Delta_k}{\alpha_{k,n}} - \frac{\Delta'_k}{\alpha_{k,n}} \right|^2$$

and

$$\begin{aligned} \|(f_{\Delta,j} - f_{\Delta',j}) \mathbf{1}_{F_n \leq D_n}\|_{\infty} &= \sup_{\mathbf{u}: F_n(\mathbf{u}) \leq D_n} |f_{\Delta,j}(\mathbf{u}) - f_{\Delta',j}(\mathbf{u})| \\ &\leq \max_{k \in I_{n,j}} \left| \frac{\Delta_k}{\alpha_{k,n}} - \frac{\Delta'_k}{\alpha_{k,n}} \right| \sup_{\mathbf{u}: F_n(\mathbf{u}) \leq D_n, \boldsymbol{\theta} \in \Theta_n} \sum_{k=1}^j \left| \frac{\alpha_{k,n}}{\alpha_{j,n}} \frac{\partial}{\partial \theta_k} \phi(\mathbf{u}; \boldsymbol{\theta})_j \right| \\ &\leq \max_{k \in I_{n,j}} \left| \frac{\Delta_k}{\alpha_{k,n}} - \frac{\Delta'_k}{\alpha_{k,n}} \right| D_n. \end{aligned}$$

Note that $|\Delta_k / \alpha_{k,n}| \leq c_n$ for all k by the definition of Δ and therefore $|\Delta_k / \alpha_{k,n} - \Delta'_k / \alpha_{k,n}| \leq 2c_n$. Now let $\Delta_1, \dots, \Delta_N$ be the centers of an η -covering of $\{\Delta \in \mathbb{R}^{p_n} : \|\Delta\|_{\infty} \leq 2c_n, \Delta_{j'} = 0 \forall j' \notin I_{j,n}\}$ w.r.t. $\|\cdot\|_{\infty}$, which we can find with $N = (2c_n/\eta)^{k_n}$. Then, $f_{\Delta_1,j}, \dots, f_{\Delta_N,j}$ are the centers of an $M_n \eta$ -covering of $\mathcal{F}_n^{(j)}$ w.r.t. $L_2(P)$ and a $D_n \eta$ -covering of $\mathcal{F}_n^{(j)}$ w.r.t. $\|\cdot\|_{\infty}$. Choosing $\eta = \varepsilon / M_n$ and $\eta = \varepsilon / D_n$, respectively, gives

$$N(\varepsilon, \mathcal{F}_n^{(j)}, L_2(P)) \leq (2M_n c_n / \varepsilon)^{k_n}, \quad N(\varepsilon, \mathcal{F}_n^{(j)}, \|\cdot\|_{\infty}) \leq (2D_n c_n / \varepsilon)^{k_n}.$$

Taking the union over the coverings of $\mathcal{F}_n^{(j)}, j = 1, \dots, p_n$, we obtain a covering of \mathcal{F}_n and

$$\ln N(\varepsilon, \mathcal{F}_n, L_2(P)) \leq \ln p_n + k_n \ln(2M_n c_n / \varepsilon), \quad \ln N(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq \ln p_n + k_n \ln(2D_n c_n / \varepsilon).$$

Theorem 2.14.21 in [Van der Vaart and Wellner \(2023\)](#) yields

$$\mathbb{E} \left[\sup_{f \in \mathcal{F}_n} |(\mathbb{P}_n - P)f| \right] \lesssim \frac{M_n c_n \sqrt{k_n + \ln p_n}}{\sqrt{n}} + \frac{D_n c_n (k_n + \ln p_n)}{n},$$

where \lesssim means “bounded up to a universal constant”. The claim now follows from Markov’s inequality. \square

Lemma 3. *Under assumptions (A7), (A8), (A9) and $d_n^3/n = O(1)$, it holds that*

$$\|\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) - \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_\infty = O_p(\sqrt{\ln p_n / n}).$$

Proof. A Taylor expansion gives

$$\begin{aligned} \mathbb{P}_n[\phi_X(\boldsymbol{\theta}^*) - \phi_U(\boldsymbol{\theta}^*)] &= \frac{1}{n} \sum_{i=1}^n \phi(F_n(\mathbf{X}_i); \boldsymbol{\theta}^*) - \phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*) \\ &= \frac{1}{n} \sum_{i=1}^n \underbrace{\nabla_{\mathbf{u}} \phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*)}_{\in \mathbb{R}^{p_n \times d_n}} (F_n(\mathbf{X}_i) - F(\mathbf{X}_i)) + I_2 \\ &= \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1}^n \nabla_{\mathbf{u}} \phi(F(\mathbf{X}_i); \boldsymbol{\theta}^*) \begin{pmatrix} \mathbb{1}(X_{i'1} \leq X_{i1}) - F_1(X_{i1}) \\ \vdots \\ \mathbb{1}(X_{i'd_n} \leq X_{id_n}) - F_{d_n}(X_{id_n}) \end{pmatrix} + I_2 \\ &= I_1 + I_2, \end{aligned}$$

where

$$I_2 = \frac{1}{n} \sum_{i=1}^n I_{2,i} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} (F_n(\mathbf{X}_i) - F(\mathbf{X}_i))^\top \nabla_{\mathbf{u}}^2 \phi(\tilde{\mathbf{u}}_i; \boldsymbol{\theta}^*)_1 (F_n(\mathbf{X}_i) - F(\mathbf{X}_i)) \\ \vdots \\ (F_n(\mathbf{X}_i) - F(\mathbf{X}_i))^\top \nabla_{\mathbf{u}}^2 \phi(\tilde{\mathbf{u}}_i; \boldsymbol{\theta}^*)_{p_n} (F_n(\mathbf{X}_i) - F(\mathbf{X}_i)) \end{pmatrix} \quad (16)$$

with some $\tilde{\mathbf{u}}_i$ on the segment between $F_n(\mathbf{X}_i)$ and $F(\mathbf{X}_i)$, and $\phi(\mathbf{u}; \boldsymbol{\theta})_k$ denotes the k -th entry of $\phi(\mathbf{u}; \boldsymbol{\theta})$. Denote

$$I_1 = \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1}^n \mathbf{Y}_{ii'}, \quad h(\mathbf{X}_i, \mathbf{X}_{i'}) := \frac{1}{2} \mathbf{Y}_{ii'} + \frac{1}{2} \mathbf{Y}_{i'i} \quad \text{and} \quad h_1(\mathbf{x}) := \mathbb{E}[h(\mathbf{x}, \mathbf{X})]. \quad (17)$$

Hoeffding’s decomposition yields

$$\begin{aligned} \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1}^n \mathbf{Y}_{ii'} &= \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1}^n h(\mathbf{X}_i, \mathbf{X}_{i'}) \\ &= \frac{1}{n(n+1)} \sum_{i=1}^n h(\mathbf{X}_i, \mathbf{X}_i) + \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n h(\mathbf{X}_i, \mathbf{X}_{i'}) \\ &= \frac{1}{n(n+1)} \sum_{i=1}^n h(\mathbf{X}_i, \mathbf{X}_i) + \frac{2}{n} \sum_{i=1}^n h_1(\mathbf{X}_i) \\ &\quad + \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n h(\mathbf{X}_i, \mathbf{X}_{i'}) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_{i'}). \end{aligned}$$

Since $|h(\mathbf{X}_i, \mathbf{X}_i)_k| = O_p(1)$ for all $k = 1, \dots, p_n$ by (A8), we have

$$\left\| \frac{1}{n(n+1)} \sum_{i=1}^n h(\mathbf{X}_i, \mathbf{X}_i) \right\|_{\infty} \leq \left\| \frac{1}{n(n+1)} \sum_{i=1}^n h(\mathbf{X}_i, \mathbf{X}_i) \right\|_2 = O_p(\sqrt{p_n/n}) = o_p(\sqrt{\ln p_n/n}).$$

Lemma 13 in Gauss and Nagler (2025) and (A7) yield $\|\frac{2}{n} \sum_{i=1}^n h_1(\mathbf{X}_i)\|_{\infty} = O_p(\sqrt{\ln p_n/n})$. Lemma 4, (A7) and (A8) imply

$$\left\| \frac{1}{n(n+1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n h(\mathbf{X}_i, \mathbf{X}_{i'}) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_{i'}) \right\|_{\infty} = o_p(\sqrt{\ln p_n/n}),$$

so $\|I_1\|_{\infty} = O_p(\sqrt{\ln p_n/n})$. Lemma 5, (A9) and $d_n^3/n = O(1)$ yield $\|I_2\|_{\infty} = O_p(\sqrt{\ln p_n/n})$. Since

$$\|\mathbb{P}_n \phi_X(\boldsymbol{\theta}^*) - \mathbb{P}_n \phi_U(\boldsymbol{\theta}^*)\|_{\infty} \leq \|I_1\|_{\infty} + \|I_2\|_{\infty},$$

this implies the claim. \square

Lemma 4. Under assumption (A7) and (A8), with $h(\mathbf{X}_i, \mathbf{X}_{i'})$ and $h_1(\mathbf{x})$ as defined in (6), it holds that

$$\|II\|_2 = \left\| \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i'=1, i' \neq i}^n h(\mathbf{X}_i, \mathbf{X}_{i'}) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_{i'}) \right\|_2 = O_p(\sqrt{p_n/n}).$$

Proof. Define

$$\eta(\mathbf{X}_i, \mathbf{X}_{i'}) := h(\mathbf{X}_i, \mathbf{X}_{i'}) - h_1(\mathbf{X}_i) - h_1(\mathbf{X}_{i'}). \quad (18)$$

It holds that

$$\begin{aligned} \mathbb{E}[\|II\|_2^2] &= \frac{1}{n^2(n-1)^2} \sum_{k=1}^{p_n} \mathbb{E} \left[\left(\sum_{i=1}^n \sum_{i'=1, i' \neq i}^n \eta(\mathbf{X}_i, \mathbf{X}_{i'})_k \right)^2 \right] \\ &= \frac{1}{n^2(n-1)^2} \sum_{k=1}^{p_n} \sum_{i_1=1}^n \sum_{i'_1=1, i'_1 \neq i_1}^n \sum_{i_2=1}^n \sum_{i'_2=1, i'_2 \neq i_2}^n \mathbb{E} \left[\eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_1})_k \eta(\mathbf{X}_{i_2}, \mathbf{X}_{i'_2})_k \right]. \end{aligned} \quad (19)$$

In this sum, all terms with $\{i_1, i'_1\} \neq \{i_2, i'_2\}$ are zero. To see this, first note that

$$\begin{aligned} \mathbb{E}[\eta(\mathbf{X}_i, \mathbf{X}_{i'})_k | \mathbf{X}_i] &= \mathbb{E}[h(\mathbf{X}_i, \mathbf{X}_{i'})_k | \mathbf{X}_i] - \mathbb{E}[h_1(\mathbf{X}_i)_k | \mathbf{X}_i] - \mathbb{E}[h_1(\mathbf{X}_{i'})_k | \mathbf{X}_i] \\ &= h_1(\mathbf{X}_i)_k - h_1(\mathbf{X}_i)_k - \mathbb{E}[h_1(\mathbf{X}_{i'})_k] = h_1(\mathbf{X}_i)_k - h_1(\mathbf{X}_i)_k - 0 = 0, \end{aligned}$$

where $\mathbb{E}[h_1(\mathbf{X}_{i'})_k] = 0$ follows from $\mathbb{E}[\mathbb{1}(X_{i'l} \leq X_{il}) - F_1(X_{il}) | X_{il}] = 0$ for all $l = 1, \dots, d_n$, $i, i' = 1, \dots, n$. Now, with $i_1 = i_2$ but $i'_1 \neq i'_2$ (the case $i_1 \neq i_2, i'_1 = i'_2$ works the same), we have

$$\begin{aligned} \mathbb{E}[\eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_1})_k \eta(\mathbf{X}_{i_2}, \mathbf{X}_{i'_2})_k] &= \mathbb{E} \left[\mathbb{E}[\eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_1})_k \eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_2})_k | \mathbf{X}_{i_1}, \mathbf{X}_{i'_2}] \right] \\ &= \mathbb{E} \left[\mathbb{E}[\eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_1})_k | \mathbf{X}_{i_1}, \mathbf{X}_{i'_2}] \eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_2})_k \right] \\ &= \mathbb{E} \left[\mathbb{E}[\eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_1})_k | \mathbf{X}_{i_1}] \eta(\mathbf{X}_{i_1}, \mathbf{X}_{i'_2})_k \right] = 0. \end{aligned}$$

Each of the remaining $O(p_n n^2)$ terms in (19) is $O(1)$ since (A7) and (A8) imply $\mathbb{E}[\eta(\mathbf{X}_i, \mathbf{X}_{i'})_k^2] = O(1)$. Therefore $\mathbb{E}[\|II\|_2^2] = O(p_n/n^2)$, so $\|II\|_2 = O_p(\sqrt{p_n/n})$. \square

Lemma 5. Under assumptions (A9), with I_2 as defined in (16), it holds that

$$\|I_2\|_\infty = O_p\left(\sqrt{\frac{d_n^3 \ln p_n}{n^2}}\right).$$

Proof. We use $\langle A, B \rangle_F \leq \|A\|_F \|B\|_F$ for matrices A, B with $\langle A, B \rangle_F = \sum_{ml} a_{ml} b_{ml}$ and $\|A\|_F = \sqrt{\langle A, A \rangle_F}$. The inequality follows directly from Cauchy-Schwarz inequality when treating a matrix $A \in \mathbb{R}^{l \times m}$ as a vector in \mathbb{R}^{lm} . Denote

$$\partial_{ml}\phi(\mathbf{u}; \boldsymbol{\theta})_k = \frac{\partial^2}{\partial u_m \partial u_l} \phi(\mathbf{u}; \boldsymbol{\theta})_k.$$

I_2 depends on $\tilde{\mathbf{u}}_i, i = 1, \dots, n$ on segments between $F_n(\mathbf{X}_i)$ and $F(\mathbf{X}_i)$. Lemma 8 implies that we can bound terms involving these $\tilde{\mathbf{u}}_i$ by taking the supremum over $G \in \mathcal{G}_n$ as defined in (A9).

Using $\|I_2\|_\infty \leq 1/n \sum_{i=1}^n \|I_{2,i}\|_\infty$, $\|I_{2,i}\|_\infty = O_p(\mathbb{E}[\|I_{2,i}\|_\infty])$ and $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$ for a function $f: \mathbb{R} \rightarrow \mathbb{R}$, we obtain

$$\begin{aligned} & \mathbb{E}[\|I_{2,i}\|_\infty] \\ & \leq \mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \partial_{ml}\phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k (F_{nm}(X_{im}) - F_m(X_{im}))(F_{nl}(X_{il}) - F_l(X_{il})) \right] \\ & \leq \mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \underbrace{|\partial_{ml}\phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k|}_{a_{ml,k}(G(\mathbf{X}_i))} \underbrace{\|F_{nm} - F_m\|_\infty \|F_{nl} - F_l\|_\infty}_{b_{ml}} \right] \\ & \leq \mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \|A_{G,k}\|_F \|B\|_F \right] \leq \sqrt{\mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \|A_{G,k}\|_F^2 \right] \mathbb{E}[\|B\|_F^2]} \end{aligned}$$

with matrices $A_{G,k}, B \in \mathbb{R}^{d_n \times d_n}$. Now

$$\begin{aligned} \mathbb{E}[\|B\|_F^2] &= \mathbb{E} \left[\sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \|F_{nm} - F_m\|_\infty^2 \|F_{nl} - F_l\|_\infty^2 \right] \\ &= \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \mathbb{E}[\|F_{nm} - F_m\|_\infty^2 \|F_{nl} - F_l\|_\infty^2] = O(d_n^2/n^2), \end{aligned}$$

since $\mathbb{E}[(\sqrt{n}\|F_{nm} - F_m\|_\infty)^4] = O(1)$ for all m . To see this, define the function class $\mathcal{F} = \{f_t(x) = \mathbf{1}(x \leq t) \mid t \in \mathbb{R}\}$. Then $\|\sqrt{n}(F_n - F)\|_\infty = \sup_{f_t \in \mathcal{F}} |\sqrt{n}(\mathbb{P}_n - P)f_t|$. One can show that the covering numbers $\mathcal{N}(\varepsilon, \mathcal{G}, L_2(Q))$ are of order $O(1/\varepsilon^2)$ since $\|f_t - f_{t'}\|_{L_2(Q)} = \max\{\sqrt{F_Q(t) - F_Q(t')}, \sqrt{F_Q(t') - F_Q(t)}\}$, where F_Q is the c.d.f. of the probability measure Q . The uniform entropy integral of \mathcal{F} (see Van der Vaart and Wellner, 2023, Chapter 2.14.1) is therefore bounded, so Van der Vaart and Wellner (2023, Theorem 2.14.1) with $p = 4$ and envelope $F(x) = 1$ of \mathcal{F} yields the above claim.

For $A_{G,k}$, we have

$$\mathbb{E} \left[\sup_{1 \leq k \leq p_n, G \in \mathcal{G}_n} \|A_{G,k}\|_F^2 \right] = O(d_n \ln p_n)$$

by (A9). Altogether, we have

$$\mathbb{E}[\|I_2\|_\infty] = O\left(\sqrt{\frac{d_n^3 \ln p_n}{n^2}}\right).$$

□

Lemma 6. Under assumptions (A10) and (A11), with $f(\Delta, j)$ as defined in (12), it holds that

$$\max_{1 \leq j \leq p_n} \sup_{\Delta \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} |\mathbb{P}_n f_j(r_n C \Delta)| = O_p(\sqrt{d_n^2/n}).$$

Proof. With some \tilde{U}_i on the segment between $F_n(\mathbf{X}_i)$ and $F(\mathbf{X}_i)$, it holds that

$$\begin{aligned} f_j(r_n C \Delta) &= \nabla_{\mathbf{u}} \left[\phi(\tilde{U}_i; \boldsymbol{\theta}^* + r_n C \Delta)_j - \phi(\tilde{U}_i; \boldsymbol{\theta}^*)_j \right] (F_n(\mathbf{X}_i) - F(\mathbf{X}_i)) \\ &\leq \left\| \nabla_{\mathbf{u}} \left[\phi(\tilde{U}_i; \boldsymbol{\theta}^* + r_n C \Delta)_j - \phi(\tilde{U}_i; \boldsymbol{\theta}^*)_j \right] \right\|_2 \|F_n(\mathbf{X}_i) - F(\mathbf{X}_i)\|_2. \end{aligned}$$

We have

$$\|F_n(\mathbf{X}_i) - F(\mathbf{X}_i)\|_2 \leq \|(\|F_{nl} - F_l\|_\infty)_{l=1, \dots, d_n}\|_2 = O_p(\sqrt{d_n/n})$$

since $\mathbb{E}[(\sqrt{n}\|F_{nl} - F_l\|_\infty)^2] = O(1)$ for all $l = 1, \dots, d_n$, see Lemma 5. By (A10), there exist non-negative, real-valued functions $\psi_l(\mathbf{u})$, $l = 1, \dots, d_n$, such that

$$(r_n C \alpha_{j,n})^{-1} \left| \frac{\partial}{\partial u_l} \left[\phi(\tilde{U}_i; \boldsymbol{\theta}^* + r_n C \Delta)_j - \phi(\tilde{U}_i; \boldsymbol{\theta}^*)_j \right] \right| \leq \frac{\max_{1 \leq k \leq j} \alpha_{k,n}}{\alpha_{j,n}} \psi_l(\tilde{U}_i).$$

Denote $\boldsymbol{\psi}(\mathbf{u}) = (\psi_l(\mathbf{u}))_{l=1, \dots, d_n}$. Since $\max_{1 \leq k \leq j} \alpha_{k,n}/\alpha_{j,n} = O(1)$ for all $j = 1, \dots, p_n$ by (A11), we have, for all $i = 1, \dots, n$,

$$\max_{1 \leq j \leq p_n} (r_n C \alpha_{j,n})^{-1} \left\| \nabla_{\mathbf{u}} \left[\phi(\tilde{U}_i; \boldsymbol{\theta}^* + r_n C \Delta)_j - \phi(\tilde{U}_i; \boldsymbol{\theta}^*)_j \right] \right\|_2 = O_p(\|\boldsymbol{\psi}(\tilde{U}_i)\|_2) = O_p(\sqrt{d_n})$$

by (A10). Thus

$$\max_{1 \leq j \leq p_n} \sup_{\Delta \in \Theta_n^\Delta, |\Delta_j| = \alpha_{j,n}} (r_n C \alpha_{j,n})^{-1} |\mathbb{P}_n f_j(r_n C \Delta)| = O_p(\sqrt{d_n^2/n}).$$

□

Lemma 7. Under assumption (A12), for I_2 as defined in (16), it holds that

$$\|I_2\| = O_p\left(\frac{d_n \sqrt{p_n}}{n}\right).$$

Proof. We slightly adapt Lemma 5 to obtain tight bounds in $\|\cdot\|_2$ norm:

Using $\|\mathbf{x}\| = \sup_{\|\Delta\|=1} \langle \mathbf{x}, \Delta \rangle$, $\|I_2\| \leq 1/n \sum_{i=1}^n \|I_{2,i}\|$ and $\|I_{2,i}\| = O_p(\mathbb{E}[\|I_{2,i}\|])$, we have

$$\begin{aligned}
& \mathbb{E}[\|I_{2,i}\|] \\
& \leq \mathbb{E} \left[\sup_{\|\Delta\|=1, G \in \mathcal{G}_n} \sum_{k=1}^{p_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \partial_{ml} \phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k (F_{nm}(X_{im}) - F_m(X_{im}))(F_{nl}(X_{il}) - F_l(X_{il})) \Delta_k \right] \\
& \leq \mathbb{E} \left[\sup_{\|\Delta\|=1, G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \|F_{nm} - F_m\|_\infty \|F_{nl} - F_l\|_\infty \sum_{k=1}^{p_n} |\partial_{ml} \phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k| \Delta_k \right] \\
& \leq \mathbb{E} \left[\sup_{\|\Delta\|=1, G \in \mathcal{G}_n} \sum_{m=1}^{d_n} \sum_{l=1}^{d_n} \underbrace{\|F_{nm} - F_m\|_\infty \|F_{nl} - F_l\|_\infty}_{b_{ml}} \underbrace{\|\mathbf{a}_{ml}(G(\mathbf{X}_i))\|}_{a_{ml}(G(\mathbf{X}_i))} \|\Delta\| \right] \\
& \leq \mathbb{E} \left[\sup_{G \in \mathcal{G}_n} \|A_G\|_F \|B\|_F \right] \leq \sqrt{\mathbb{E} \left[\sup_{G \in \mathcal{G}_n} \|A_G\|_F^2 \right] \mathbb{E} [\|B\|_F^2]}
\end{aligned}$$

with \mathcal{G}_n as defined in (A9), matrices $A_G, B \in \mathbb{R}^{d_n \times d_n}$ and

$$\mathbf{a}_{ml}(G(\mathbf{x})) = (\partial_{ml} \phi(G(\mathbf{X}_i); \boldsymbol{\theta}^*)_k)_{k=1, \dots, p_n} \in \mathbb{R}^{p_n}.$$

Now $\mathbb{E}[\|B\|_F^2] = O(d_n^2/n^2)$, see Lemma 5. For A_G , we have

$$\mathbb{E} \left[\sup_{G \in \mathcal{G}_n} \|A_G\|_F^2 \right] = O(p_n)$$

by (A12). This yields the claim. \square

Lemma 8. For any random variable $\mathbf{X} \in \mathbb{R}^{d_n}$ with continuous c.d.f.s $F_m(x)$ and empirical c.d.f.s $F_{nm}(x)$, $m = 1, \dots, d_n$, it holds

$$\sup_{x \in \mathbb{R}, 1 \leq m \leq d_n} \frac{|F_{nm}(x) - F_m(x)|}{w(F_m(x))} = O_p \left(d_n^b \sqrt{\frac{\ln d_n}{n}} \right)$$

with $w(s) = s^\gamma(1-s)^\gamma$ with some $\gamma \in (0, 1)$ and $b > \gamma$.

Proof. Define the function class

$$\mathcal{G} = \{f_{m,t}(\mathbf{x}) = \mathbb{1}(x_m \leq t)/w(F_m(x_m)) : t \in \mathbb{R}, m = 1, \dots, d_n\}.$$

Now

$$\sup_{x \in \mathbb{R}, 1 \leq m \leq d_n} \frac{|F_{nm}(x) - F_m(x)|}{w(F_m(x))} = \sup_{f \in \mathcal{G}} |(\mathbb{P}_n - P)f|.$$

We need to derive the bracketing number $\mathcal{N}_{[]}(\varepsilon, \mathcal{G}, L_2(P))$ (see Van der Vaart and Wellner, 2023, Chapter 2.1.1 for a definition) of this function class. Denote $\mathcal{G}_m = \{f_t(\mathbf{x}) = \mathbb{1}(x_m \leq t)/w(F_m(x_m)) : t \in \mathbb{R}\}$. It holds $\mathcal{N}_{[]}(\varepsilon, \mathcal{G}_m, L_2(P)) = O(\varepsilon^{-k})$ for some $k < \infty$. This follows from Nagler et al. (2022, proof of Lemma A.1), since $\|\cdot\|_\beta = \|\cdot\|_{L_2(P)}$ for independent random variables. Since $\mathcal{G} = \bigcup_m \mathcal{G}_m$, a union of ε -bracketings of \mathcal{G}_m , $m = 1, \dots, d_n$, is an ε -bracketing of \mathcal{G} and therefore $\mathcal{N}_{[]}(\varepsilon, \mathcal{G}, L_2(P)) = O(d_n \varepsilon^{-k})$ for some $0 < k < \infty$.

Now we derive $\|G\|_{L_2(P)} = O(d_n^b)$ with some $b > \gamma$ for the envelope $G(\mathbf{x}) = \max_{1 \leq m \leq d_n} 1/w(F_m(x_m))$ of \mathcal{G} . With $U \sim U[0, 1]$ and some a such that $2a < 1/\gamma$, it holds

$$\mathbb{P}(G(\mathbf{X})^2 > t) \leq d_n \mathbb{P}\left(\frac{1}{w(U)^2} > t\right) \leq d_n \frac{\mathbb{E}[w(U)^{-2a}]}{t^a} \lesssim d_n t^{-a},$$

since $F_m(X_m) \sim U[0, 1]$ and $\mathbb{E}[(U^{-c}(1-U)^{-c})] < \infty$ for $c < 1$. It holds that

$$\mathbb{E}[G(\mathbf{X})^2] = \eta + \int_{\eta}^{\infty} \mathbb{P}(G(\mathbf{X})^2 > t) dt \lesssim \eta + \int_{\eta}^{\infty} d_n t^{-a} dt \lesssim \eta + d_n \eta^{1-a}.$$

With $\eta = d_n^{1/a}$, we obtain

$$\mathbb{E}[G(\mathbf{X})^2] \lesssim d_n^{1/a} + d_n \cdot d_n^{(1-a)/a} = d_n^{1/a} + d_n^{1/a} = O(d_n^{1/a})$$

and therefore $\|G\|_{L_2(P)} = O(d_n^b)$ with $b = 1/(2a) > \gamma$.

For the bracketing integral ([Van der Vaart and Wellner, 2023](#), Chapter 2.14.2), we obtain with $\|G\|_{L_2(P)} \geq 1$

$$J_{[]}^*(1, \mathcal{G} \mid G, L_2(P)) \leq \int_0^1 \sqrt{1 + \ln \mathcal{N}_{[]}(\varepsilon, \mathcal{G}, L_2(P))} d\varepsilon = O(\sqrt{\ln d_n}).$$

Now, Theorem 2.14.16 in [Van der Vaart and Wellner \(2023, Chapter 2.14.2\)](#) yields

$$\sup_{f \in \mathcal{G}} |(\mathbb{P}_n - P)f| = O_p(\sqrt{\ln d_n} d_n^b / \sqrt{n}).$$

□