## 4.1 Introduction

In this chapter, we briefly explain how one can retrieve the underlying tree structure of dependencies for a multivariate Pareto distribution which is a graphical model with respect to a forest $\mathcal{F}$. In Section (2.7), we discussed that using the conventional parametric learning approach for extremal graphical models would negatively affect the estimation. Engelke and Volgushev, 2020 suggest an ideal approach to learning the structure of extremal tree models. This approach is able to recover the true underlying tree structure of extremal tree models using the summary statistics, i.e., extremal correlation and extremal variogram. These summary statistics are similar to the correlation coefficients $r_{ij}$ in the Gaussian case [see Section (2.3)] that can be estimated empirically. Theorem (2.7.1) shows that using their corresponding weights [see Section (2.6)] as edge weights and under some mild assumptions, the minimum spanning tree uniquely recovers the true underlying tree.

## 4.2 Extremal forest learning using weight-restricted minimum spanning forest, population level

This thesis extends the notion of structure learning for extremal trees provided by Engelke and Volgushev, 2020 to learn the true underlying forest of an extremal forest model. Particularly, we will show that the weight-restricted minimum spanning forest, introduced in Subsection (2.2.5), uniquely recovers the true underlying forest model, using the weights corresponding to the extremal correlation. Note that, as it is mentioned in Remark 2 of Engelke and Volgushev, 2020, the extremal variogram does not exist if $\mathbf{Y}$ lies on lower-dimensional sub-faces of $\mathcal{L}$. Hence, the extremal forest model $\mathbf{Y}$ has mass on lower-dimensional sub-faces, we disregard the extremal variogram for the structure learning of the extremal forest, and we only use the weights corresponding to the extremal correlation. In the following, we introduce the useful lemma, that is Lemma (4.2.1), then we introduce our approach for structure learning of the extremal forest model in Proposition (4.2.4).

**Lemma 4.2.1.** *Let the $d$-dimensional random vector $\mathbf{X} = (X_1, ..., X_d)$ be in the max-domain of attraction of a multivariate Pareto distribution $\mathbf{Y}$ (defined as in Theorem (2.1.27)) which is an extremal graphical model with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$ that consists of the trees $\mathcal{T}_1, ..., \mathcal{T}_c$ where for any $i \in \{1, ..., c\}$, we have that $\mathcal{T}_i = (V_i, E_i)$ and $E_i$ is restricted the pairs in $V_i$. Then, under the additional assumption of the faithfulness of $\mathbf{Y}$ with respect to $\mathcal{F}$, for any $i, j \in V$, $\chi_{ij} = 0$ if and only if $i$ and $j$ are disconnected nodes in $\mathcal{F} = (V, E_{\mathcal{F}})$.*

*Proof.* ($\Rightarrow$) Suppose that for $i \in V_p$ and $j \in V_q$ where $p, q \in \{1, ..., c\}$, we have that

$$\chi_{ij} = 0.$$

By Equation (2.6.2),

$$\chi_{ij} = \mathbb{P}\left(Y_i > 1 \middle| Y_j > 1\right).$$

Hence, we have that,

$$\begin{aligned}\chi_{ij} &= \mathbb{P}\left(Y_i > 1 \middle| Y_j > 1\right) \\ &= \frac{\mathbb{P}(Y_i > 1, Y_j > 1)}{\mathbb{P}(Y_j > 1)}.\end{aligned}$$

Consequently,

$$\mathbb{P}(Y_i > 1, Y_j > 1) = 0.$$

as we showed in the Proposition (3.4.2) $\mathbb{P}(Y_i > 1, Y_j > 1) = 0$ implies that $Y_i$ and $Y_j$ are asymptotically independent, i.e., $Y_i \perp_e Y_j$. Here, we require the additional assumption about the faithfulness of $\mathbf{Y}$, to say that $Y_i \perp_e Y_j$ leads to that $i$ and $j$ are disconnected nodes in $\mathcal{F} = (V, E_{\mathcal{F}})$.

($\Leftarrow$) Suppose that $i$ and $j$ are disconnected nodes in the forest $\mathcal{F} = (V, E_{\mathcal{F}})$. Therefore, from Markov properties $Y_i \perp_e Y_j$. By Lemma (2.4.2)(3), $\Lambda([1,\infty) \times [1,\infty)) = 0$ and by Lemma(3.3.1) we have that

$$\Lambda([1,\infty) \times [1,\infty)) = \mathbb{P}(Y_i > 1, Y_j > 1) = 0.$$

Thus,

$$\begin{aligned} \chi_{ij} &= \mathbb{P}\left(Y_i > 1 \middle| Y_j > 1\right) \\ &= \frac{\mathbb{P}(Y_i > 1, Y_j > 1)}{\mathbb{P}(Y_j > 1)} = 0. \end{aligned}$$

Consequently,

$$\chi_{ij} = 0.$$

$\square$

*Remark* 4.2.2. From now on, when we state that $\mathbf{Y}$ is a graphical model with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$, we assume that $\mathbf{Y}$ satisfies both global Markov property and faithfulness with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$. This is because of the required assumption for Lemma (4.2.1).

In the following, we introduce a structure learning method for the extremal forest. Particularly, Proposition (4.2.4) shows that the introduced method is able to recover the true underlying extremal forest at the population level.

**Theorem 4.2.3.** *(Proposition 5 of Engelke and Volgushev, 2020) Let $\mathbf{Y}$ be an extremal graphical model on the tree $\mathcal{T} = (V, E_{\mathcal{T}})$. Then the extremal correlation coefficients satisfy for any $h, l \in V$ with $h \neq l$ that*

$$\chi_{hl} \le \chi_{ij}, \quad \forall (i,j) \in ph(hl; \mathcal{T}), \tag{4.2.1}$$

*where $ph(hl; \mathcal{T})$ denotes the path between nodes $h$ and $l$.*

**Proposition 4.2.4.** *Let the multivariate Pareto distribution $\mathbf{Y} = (Y_1, ..., Y_d)$ be an extremal graphical model with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$ that consists of trees $\mathcal{T}_1, ..., \mathcal{T}_c$. Let $\chi_{\min}$ be the smallest $\chi_{ij}$ for all $i, j \in \{1, ..., d\}$ that is strictly larger than zero. Under the additional assumption that Inequality (4.2.1) is strict for any $h$ and $l$ (with $h \neq l$) in a tree $\mathcal{T}_b$ for all $b \in \{1, ..., c\}$, as soon as $(i,j) \neq (h,l)$. The weight-restricted minimum spanning forest, defined as in Subsection (2.2.5), corresponding to the weights $w_{ij} = -\log(\chi_{ij})$ and the threshold $\tau \in (-\log(\chi_{\min}), \infty)$ is unique and satisfies*

$$\mathcal{F}_w = \mathcal{F}, \tag{4.2.2}$$

*where $\mathcal{F}_w$ is the weight-restricted minimum spanning forest.*

*Proof.* The proof contains two steps:

- We must show that the sets of connected components of $\mathcal{F}_w$ and $\mathcal{F}$ are the same.

  $$\mathfrak{C}(\mathcal{F}_w) = \mathfrak{C}(\mathcal{F}),$$

  where $\mathfrak{C}(\mathcal{F})$ is the set of connected components of $\mathcal{F}$. The latter Equation is equivalent to saying that,

  $$\begin{aligned} &\{\text{for any connected nodes } i, j \in \mathcal{F} : i, j \text{ are connected nodes in } \mathcal{F}_w\} \wedge \\ &\{\text{for any unconnected nodes } i, j \in \mathcal{F} : i, j \text{ are unconnected nodes in } \mathcal{F}_w\} \Leftrightarrow \mathfrak{C}(\mathcal{F}_w) = \mathfrak{C}(\mathcal{F}) \end{aligned}$$

1. Any two vertices $i$ and $j$ in $\{1, ..., d\}$, that are connected in the $\mathcal{F}$, must be connected in $\mathcal{F}_w$.

   By Lemma (4.2.1), for connected nodes $i$ and $j$ in $\mathcal{F}$, $\chi_{ij}$ must be strictly larger than zero. Therefore, $\chi_{ij} \geq \chi_{\min}$ and $w_{ij} < \tau$. Now by contradiction, suppose that $i$ and $j$ are not connected in $\mathcal{F}_w$. If we add edge $e_{ij}$ to the $\mathcal{F}_w$, the created spanning subgraph, call it $\mathcal{H}$, remains acyclic, and all its edges are lighter than $\tau$. However, $\mathcal{H}$ is a forest which has more edges than $\mathcal{F}_w$ which contradicts the maximality of $|\mathcal{F}_w|$ (see the Remark (2.2.9)). Therefore, $i$ and $j$ are connected in $\mathcal{F}_w$.

2. Any two vertices $i$ and $j$ in $\{1, ..., d\}$, that are unconnected in the $\mathcal{F}$, must be unconnected in $\mathcal{F}_w$.

   By contradiction, suppose that $i$ and $j$ are connected in $\mathcal{F}_w$. If $i$ and $j$ are adjacent in $\mathcal{F}_w$, then by definition of weight-restricted minimum spanning forest in Subsection (2.2.5), $w_{ij} < \tau$. The latter results in $\chi_{ij} \geq \chi_{\min}$ and $\chi_{ij} \neq 0$. By Lemma (4.2.1), $i$ and $j$ must be connected in $\mathcal{F}$, which is a contradiction. If $i$ and $j$ are connected but not adjacent in $\mathcal{F}_w$, then there is a path $p$ in $\mathcal{F}_w$ between $i$ and $j$, where $p = (s_1, s_2, ..., s_n)$ such that $s_1 = i$ and $s_n = j$. Let $l \in \{1, ..., n-1\}$, by definition of weight-restricted minimum spanning forest in Subsection (2.2.5), for any two adjacent nodes $s_l$ and $s_{l+1}$ in path $p$, we have $w_{s_l, s_{l+1}} < \tau$. Hence, $\chi_{s_l, s_{l+1}} \geq \chi_{\min}$ and $\chi_{s_l, s_{l+1}} \neq 0$. Consequently, by Lemma (4.2.1), $s_l$ and $s_{l+1}$ must be connected in $\mathcal{F}$. So by transitive property of connectivity $i$ and $j$ must be connected in $\mathcal{F}$, which is a contradiction. Therefore, any two unconnected vertices $i$ and $j$ in the $\mathcal{F}$, must be unconnected in $\mathcal{F}_w$.

- We must show that each connected component of $\mathcal{F}_w$, which is a tree, is exactly the same as a tree in $\mathcal{F}$.

  $$\mathcal{T}_{w,i} = \mathcal{T}_i, \quad \text{for all } i \in \{1, ..., c\}$$

  First, we show that any tree in $\mathcal{F}_w$, say $\mathcal{T}_{w,i} = (V_i, E_{w,i})$, is a minimum spanning tree of the subgraph $\mathcal{H} = (V_i, E_{\mathcal{H}})$, in which $E_{\mathcal{H}}$ contains the edges corresponding to the pairs in $V_i$.

  By contradiction, suppose that $\mathcal{T}_{w,i}$ is heavier than the minimum spanning tree of $\mathcal{H}$, called $\mathcal{T}_{mst,\mathcal{H}}$. We claim that $\mathcal{T}_{mst,\mathcal{H}}$ has edges that are strictly lighter than $\tau$. By contradiction, suppose there is an edge, say $e$, in $\mathcal{T}_{mst,\mathcal{H}}$ for which $w_e \geq \tau$. We know that $e$ is not in the $\mathcal{T}_{w,i}$ since it is heavier than $\tau$. Thus, by adding $e$ to $\mathcal{T}_{w,i}$ we create a cycle $C$ in which all edges except $e$ are lighter than $\tau$. Since $\mathcal{T}_{mst,\mathcal{H}}$ does not contain the cycle $C$ (tree is acyclic), there is an $e'$ in $\mathcal{T}_{w,i}$, which is not in the $\mathcal{T}_{mst,\mathcal{H}}$ and is lighter than $\tau$ and $e$. Hence, by replacing $e$ and $e'$, the resulting graph remains a tree, and it is lighter than the minimum spanning tree, $\mathcal{T}_{mst,\mathcal{H}}$, which is a contradiction. So, $\mathcal{T}_{mst,\mathcal{H}}$ has edges that are strictly lighter than $\tau$.

  If we replace $\mathcal{T}_{w,i}$ by the minimum spanning tree of $\mathcal{H}$, then the created graph remains acyclic; all its edges are strictly lighter than $\tau$; it has the same number of edges as $\mathcal{F}_w$, and it is lighter than $\mathcal{F}_w$, which is in contradiction with the optimality of $\mathcal{F}_w$. Therefore, any tree in $\mathcal{F}_w$, say $\mathcal{T}_{w,i} = (V_i, E_{w,i})$, is a minimum spanning tree of subgraph $\mathcal{H} = (V_i, E_{\mathcal{H}})$.

  Therefore, inspired by the proof of Proposition 5 Engelke and Volgushev, 2020 one can introduce an 'E-saturating matching' for any tree in $\mathcal{F}$, $\mathcal{T}_i$ for $i \in \{1, ..., c\}$, that guarantees

  $$\mathcal{T}_{w,i} = \mathcal{T}_i,$$

  concerning that the additional assumption that Inequality (4.2.1) is strict as soon as $(i, j) \neq (h, l)$. Therefore, $\mathcal{F}_w$ consists of $\mathcal{T}_{w,1}, ..., \mathcal{T}_{w,c}$ which are is exactly same as $\mathcal{T}_1, ..., \mathcal{T}_c$ in $\mathcal{F}$. This implies Equation (4.2.2).

□

## 4.3 Consistency of the extremal forest learning method, sample level

Suppose that we observe independent copies $\mathbf{X}_1, ..., \mathbf{X}_n$ of the $d$-dimensional random vector $\mathbf{X} = (X_1, ..., X_d)$, that is in the max-domain of attraction of a multivariate Pareto distribution $\mathbf{Y}$ (defined as in Theorem (2.1.27)) which is an extremal graphical model with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$. We would like to estimate $\mathcal{F}$ from the observations $\mathbf{X}_1, ..., \mathbf{X}_n$. To do so, we first estimate extremal correlations using the estimator $\hat{\chi}$ in Equation (2.6.5), then based on the theoretical fundamental that is provided by Proposition (4.2.4), we find the weight-restricted minimal spanning forest concerning the empirical weights obtained by $\hat{w}_\chi = -\log(\hat{\chi})$, using the Algorithm (3). This forest is denoted by $\hat{\mathcal{F}}_{\chi,\tau}$ where

$$\hat{\mathcal{F}}_{\chi,\tau} := \underset{\mathcal{F}=(V,E_{\mathcal{F}})}{\arg\min} \sum_{(i,j)\in E_{\mathcal{F}}} -\log(\hat{\chi}_{ij}),$$

regarding that all edges in $\hat{\mathcal{F}}_{\chi,\tau}$ are strictly lighter than the $\tau$, and $|\hat{\mathcal{F}}_{\chi,\tau}|$ is maximal.

In what follows, we first show the consistency of the structure learning method for extremal forest under the assumption that $\chi_{\min}$, which is the smallest **non-zero** extremal correlation, is known by the information provided by an omniscient oracle. Subsequently, we show that the structure learning method for the extremal forest is consistent even without knowing the $\chi_{\min}$.

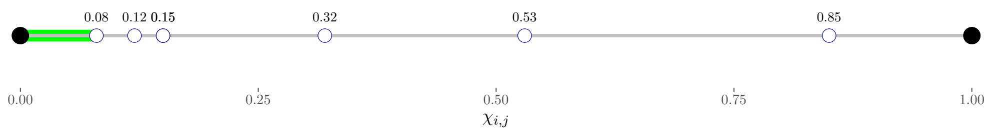


Figure 4.1: Green interval shows the extremal correlations for which negative log transformations are suitable choices for the threshold $\tau$ at the population level.

Let us introduce a helpful lemma for the proof of subsequent propositions.

**Lemma 4.3.1.** *Suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ are two graphs with same set of vertices. If we know that any unconnected nodes $i$ and $j$ in $\mathcal{G}_1$ are not adjacent in $\mathcal{G}_2$ then $i$ and $j$ are unconnected in $\mathcal{G}_2$.*

*Proof.* Assume that $i$ and $j$ be in two different connected components of $\mathcal{G}_1$, denoted by $\mathfrak{C}_i$ and $\mathfrak{C}_j$, respectively. By contradiction, suppose that $i$ and $j$ are connected in $\mathcal{G}_2$. Let $p = (s_1, ..., s_n)_{n\geq 3}$ be the shortest path $p$ between $\mathfrak{C}_i$ and $\mathfrak{C}_j$ in $\mathcal{G}_2$, where $s_1 \in \mathfrak{C}_i$ and $s_n \in \mathfrak{C}_j$. Then $s_1$ and $s_2$ must be unconnected in $\mathcal{G}_1$, otherwise the path $q = (s_2, ..., s_n)$ is the shortest path between $\mathfrak{C}_i$ and $\mathfrak{C}_j$ which contradicts minimality of $p$. Therefore, $s_1$ and $s_2$ are unconnected in $\mathcal{G}_1$ but adjacent in $\mathcal{G}_2$ which is a contradiction. Therefore, any unconnected nodes $i$ and $j$ in $\mathcal{G}_1$ are unconnected in $\mathcal{G}_2$ as well. $\square$

**Proposition 4.3.2.** *Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be independent copies of the $d$-dimensional random vector $\mathbf{X} = (X_1, ..., X_d)$, that is in the max-domain of attraction of a multivariate Pareto distribution $\mathbf{Y}$ (defined as in Theorem (2.1.27)) which is an extremal graphical model with respect to the forest $\mathcal{F} = (V, E_{\mathcal{F}})$. Suppose that $\chi_{\min}$ is the smallest* ***non-zero*** *extremal correlation among all extremal correlations of the pairs in $V$, and we know it from an omniscient oracle. Let us denote the number of extreme observations by $k$. If $k \to \infty$ and $n \to \infty$ such that $k/n \to 0$, then $\hat{\mathcal{F}}_{\chi,\tau}$ consistently recovers the true underlying forest using a threshold $\tau \in (-\log(\chi_{\min}), +\infty)$.*

$$\mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau} = \mathcal{F}\right) \to 1.$$

*Remark* 4.3.3. Henceforth, without loss of generality, we assume that $k = o(n^{2\alpha/(2\alpha+1)})$. This assumption is in line with the assumption in Theorem 2.2 Einmahl et al., 2006. With this assumption, $k \to \infty$ implies $n \to \infty$ and $k/n \to 0$.

Let us denote the set of connected components of $\hat{\mathcal{F}}_{\chi,\tau}$ and $\mathcal{F}$ by $\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau})$ and $\mathfrak{C}(\mathcal{F})$ respectively. A tree in $\mathcal{F}$ is denoted by $\mathcal{T}$ and a tree in $\hat{\mathcal{F}}_{\chi,\tau}$ is denoted by $\hat{\mathcal{T}}_\chi$. We also denote $\chi_\tau := \exp(-\tau)$.

*Proof.* We must show that **for the known threshold** $\tau$,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau} = \mathcal{F}\right) = 1.$$

For the first step of the proof, we show that

$$\lim_{k\to\infty} \mathbb{P}\left(\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\right) = 1,$$

meaning that the connected components of the estimated forest consistently recover the connected components of the true underlying forest. Therefore, we must show that as $k$ goes to $\infty$, any two unconnected nodes, say $p$ and $q$, in the forest $\mathcal{F}$ are unconnected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau}$. Moreover, as $k$ goes to $\infty$, any two connected nodes, say $i$ and $j$, in the forest $\mathcal{F}$ are connected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau}$. Therefore, Equation (4.3) is equivalent to

$$\begin{aligned}\lim_{k\to\infty} \mathbb{P}\Big(&\{\text{for any unconnected nodes } p, q \in \mathcal{F} : p, q \text{ are unconnected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\} \wedge \\ &\{\text{for any connected nodes } i, j \in \mathcal{F} : i, j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\}\Big) = 1. \qquad (4.3.1)\end{aligned}$$

We start with the first event in Equation (4.3.1). By the consistency of the estimator of the extremal correlation $\hat{\chi}$,

$$\lim_{k\to\infty} \mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}| < \varepsilon\right) = 1, \qquad \forall \varepsilon > 0. \tag{4.3.2}$$

Moreover, by Lemma (4.2.1), for any unconnected nodes $p$ and $q$ in $\mathcal{F}$, we have that

$$\chi_{pq} = 0.$$

Hence, Equation (4.3.2) is equivalent to

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\chi}_{pq} < \varepsilon\right) = 1, \qquad \forall \varepsilon > 0. \tag{4.3.3}$$

From the fact that $\chi_\tau$ is between 0 and $\chi_{\min}$, i.e., the **smallest non-zero** extremal correlation among all extremal correlations, there exists an $\varepsilon^* > 0$ where

$$\varepsilon^* < \chi_\tau.$$

Thus by Equation (4.3.3), there exists an $\varepsilon^*$ for which

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\chi}_{pq} < \varepsilon^* < \chi_\tau\right) = 1.$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\chi}_{pq} < \chi_\tau\right) = 1,$$

and,

$$\lim_{k\to\infty} \mathbb{P}(\hat{w}_{pq} > \tau) = 1.$$

The latter shows that as $k$ goes to $\infty$, any two unconnected nodes $p$ and $q$ in $\mathcal{F}$ are not adjacent in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau}$ with probability equals to one; since the output of Algorithm (3) is restricted to have edges that are strictly lighter than the threshold $\tau$. By Lemma (4.3.1), the latter implies that any two unconnected nodes $p$ and $q$ in $\mathcal{F}$ are not connected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau}$ with probability equals to one. Therefore,

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any connected nodes } p,q \in \mathcal{F} : p,q \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\}\Big) = 1. \quad (4.3.4)$$

We consider the second event of Equation (4.3.1) in the next step. By the consistency of the estimator of the extremal correlation $\hat{\chi}$,

$$\lim_{k\to\infty} \mathbb{P}(|\hat{\chi}_{ij} - \chi_{ij}| < \varepsilon) = 1, \qquad \forall \varepsilon > 0,$$

which implies,

$$\lim_{k\to\infty} \mathbb{P}(-\varepsilon < \hat{\chi}_{ij} - \chi_{ij} < \varepsilon) = 1, \qquad \forall \varepsilon > 0,$$

and,

$$\lim_{k\to\infty} \mathbb{P}(\chi_{ij} - \varepsilon < \hat{\chi}_{ij} < \chi_{ij} + \varepsilon) = 1, \qquad \forall \varepsilon > 0. \quad (4.3.5)$$

Equation (4.3.5) implies that

$$\lim_{k\to\infty} \mathbb{P}(\chi_{ij} - \varepsilon < \hat{\chi}_{ij}) = 1, \qquad \forall \varepsilon > 0. \quad (4.3.6)$$

Moreover, by Lemma (4.2.1), for any connected nodes $i$ and $j$ in $\mathcal{F}$, we have that

$$\chi_{ij} \neq 0,$$

which by the fact that $\chi_\tau$ is between 0 and $\chi_{\min}$, i.e., the **smallest non-zero** extremal correlation among all extremal correlations, implies

$$\chi_{ij} > \chi_\tau.$$

as an immediate consequence of the latter, there exists $\varepsilon^* > 0$ for which

$$\chi_{ij} - \varepsilon^* = \chi_\tau.$$

Hence, Equation (4.3.6) is equivalent to

$$\lim_{k\to\infty} \mathbb{P}(\chi_\tau = \chi_{ij} - \varepsilon^* < \hat{\chi}_{ij}) = 1.$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}(\chi_\tau < \hat{\chi}_{ij}) = 1,$$

and,

$$\lim_{k\to\infty} \mathbb{P}(\tau > \hat{w}_{ij}) = 1. \quad (4.3.7)$$

The latter shows that as $k$ goes to $\infty$, nodes $i$ and $j$ are connected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau}$. By contradiction, suppose that $i$ and $j$ are unconnected; therefore, if we add $e_{ij}$ to the forest $\hat{\mathcal{F}}_{\chi,\tau}$, the created graph does not have a cycle and $\tau > \hat{w}_{ij}$ which guarantees that all edges in the created graph are lighter than $\tau$. However, the created graph is a forest that has one edge more than $\hat{\mathcal{F}}_{\chi,\tau}$ which contradicts the maximality of the number of edges of the estimated forest by Algorithm (3). Consequently, Equation (4.3.7) implies that $i$ and $j$ are connected in $\hat{\mathcal{F}}_{\chi,\tau}$. Therefore,

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any connected nodes } i,j \in \mathcal{F} : i,j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\}\Big) = 1. \tag{4.3.8}$$

By Equations (4.3.4) and (4.3.8) for two events in Equation (4.3.1),

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any unconnected nodes } p,q \in \mathcal{F} : p,q \text{ are unconnected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\}\wedge$$
$$\{\text{for any connected nodes } i,j \in \mathcal{F} : i,j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau}\}\Big) = 1.$$

As an immediate consequence,

$$\lim_{k\to\infty} \mathbb{P}\Big(\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\Big) = 1. \tag{4.3.9}$$

By Theorem 2 of Engelke and Volgushev, 2020 we have that,

$$\lim_{k\to\infty} \mathbb{P}\Big(\hat{\mathcal{T}}_{\chi} = \mathcal{T}\Big|\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\Big) = 1, \quad \text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}. \tag{4.3.10}$$

Hence, by Equations (4.3.10) and (4.3.9),

$$\begin{aligned}
\lim_{k\to\infty} \mathbb{P}\Big(\hat{\mathcal{F}}_{\chi,\tau} = \mathcal{F}\Big) &= \lim_{k\to\infty} \mathbb{P}\Big(\{\text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}\text{: } \hat{\mathcal{T}}_{\chi} = \mathcal{T}\} \wedge \{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\}\Big) \\
&= \lim_{k\to\infty} \mathbb{P}\Big(\{\text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}\text{: } \hat{\mathcal{T}}_{\chi} = \mathcal{T}\}\Big|\{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\}\Big) \cdot \\
&\qquad \mathbb{P}\Big(\{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau}) = \mathfrak{C}(\mathcal{F})\}\Big) \\
&= 1 \cdot 1.
\end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}\Big(\hat{\mathcal{F}}_{\chi,\tau} = \mathcal{F}\Big) = 1.$$

□

The following shows the consistency of the structure learning method for the extremal forest model even without knowing the $\chi_{\min}$. In this case, we do not have a known interval for selecting the threshold required for Algorithm (3) as former case. Even in this case, the weight-restricted algorithm is able to recover the true underlying forest structure consistently. Before going further, we introduce some definitions and a lemma that will be useful in the following.

### 4.3.1 Rate of convergence

Rates of convergence quantify the stochastic order of magnitude of an estimation error in dependence of the sample size $n$. This order of magnitude is usually represented using the symbols: $O_{\mathbb{P}}$ and $o_{\mathbb{P}}$ (for more details, see Liebl et al., 2022).

**Definition 4.3.4.** *Let $\{Z_n\}_{n=1,2,3,\dots}$ be a sequence of random variables, and let $\{c_n\}_{n=1,2,3,\dots}$ be a sequence of positive deterministic numbers. We will write $Z_n = O_{\mathbb{P}}(c_n)$ if for any $\varepsilon > 0$ there exist numbers $M$ and $m$ such that*

$$\mathbb{P}(|\, Z_n| \geq M \cdot c_n) \leq \varepsilon, \quad \textit{for all} \quad n \geq m.$$

*We write $Z_n = o_{\mathbb{P}}(c_n)$ if*

$$\lim_{n\to\infty} \mathbb{P}(|\, Z_n| \geq M \cdot c_n) \leq \varepsilon, \quad \textit{for all} \quad \varepsilon > 0.$$

*Remark* 4.3.5. With $c_n = 1$ for all $n$, $Z_n = O_{\mathbb{P}}(1)$ means that the sequence $\{Z_n\}$ is stochastically bounded. i.e., for any $\varepsilon > 0$ there exist numbers $M$ and $m$ such that

$$\mathbb{P}(|\,Z_n| \geq M) \leq \varepsilon, \quad \text{for all} \quad n \geq m.$$

With $c_n = 1$ for all $n$, $Z_n = o_{\mathbb{P}}(1)$ is equivalent to $Z_n \xrightarrow{\mathbb{P}} 0$, i.e., $Z_n$ converges in probability to zero. Note that:

- $Z_n = O_{\mathbb{P}}(c_n)$ is equivalent to $Z_n/c_n = O_{\mathbb{P}}(1)$.
- $Z_n = o_{\mathbb{P}}(c_n)$ is equivalent to $Z_n/c_n = o_{\mathbb{P}}(1)$.

**Definition 4.3.6.** *An estimator $\hat{\theta}_n$ of a parameter $\theta$ possesses the* ***rate of convergence*** $n^{-r}$ *if $r$ be a positive number with the property that*

$$|\hat{\theta}_n - \theta| = O_{\mathbb{P}}(n^{-r}).$$

*The rate of convergence quantifies how fast the estimation error decreases when increasing the sample size $n$.*

**Theorem 4.3.7.** *(Proposition 7 Engelke and Volgushev, 2020) There exists a universal constant $K$ such that for all $s > 0$, we have*

$$\mathbb{P}(|\hat{\chi}_{ij} - \chi_{ij}(k/n)| \geq s) \leq 5 \exp\left(-\frac{3k}{10}\left\{\frac{s^2}{K^2} \wedge 1\right\}\right). \tag{4.3.11}$$

**Lemma 4.3.8.** *The estimator $\hat{\chi}_{k/n}$ in Equation (2.6.5), denoted by $\hat{\chi}$, possesses the* ***rate of convergence*** $k^{(-1/2)}$. *That is*

$$|\hat{\chi} - \chi(k/n)| = O_{\mathbb{P}}(k^{(-1/2)})$$

*Proof.* We show that $k^{(-1/2)}$ is the rate of convergence of $\hat{\chi}$ according to the Definitions (4.3.4) and (4.3.6). Since the smaller bound is more desired, we only consider the case that $s < K$ where $s$ is smaller than the universal $K$. Therefore, by Proposition (4.3.7), we have that,

$$\mathbb{P}(|\hat{\chi}_{ij} - \chi_{ij}(k/n)| \geq s) \leq 5 \exp\left(-\frac{3ks^2}{10K^2}\right). \tag{4.3.12}$$

By Equation(4.3.4), we say that $|\hat{\chi}_{ij} - \chi_{ij}(k/n)| = O_{\mathbb{P}}(c_{n,k})$, if for any $\varepsilon > 0$ there exist numbers $M$, $n'$, and $k'$ (where $k < n$) such that

$$\mathbb{P}(|\hat{\chi}_{ij} - \chi_{ij}(k/n)| \geq M \cdot c_{n,k}) \leq \varepsilon, \quad \text{for all} \quad n > n' \text{ and } k > k',$$

where $k < n$ and $c_{n,k} = k^{(-1/2)}$.

Let us denote $M := s \cdot k^{(1/2)}$. Then by Equation (4.3.12),

$$\mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}(k/n)| \geq M \cdot k^{(-1/2)}\right) \leq 5 \exp\left(-\frac{3k}{10}\left\{\frac{\left(M \cdot k^{(-1/2)}\right)^2}{K^2}\right\}\right).$$

Hence,

$$\mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}(k/n)| \geq M \cdot k^{(-1/2)}\right) \leq 5 \exp\left(-\frac{3M^2}{10K^2}\right).$$

Let us denote $\varepsilon := 5\exp\left(-(3M^2)/(10K^2)\right)$. Then, for any $0 < \varepsilon < 5$ there exist numbers $M = \sqrt{(10K^2/3)\log(5/\varepsilon)}$ and $k' = (10/3)\log(5/\varepsilon)$ such that

$$\mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}\,(k/n)\,| \geq M \cdot k^{(-1/2)}\right) \leq \varepsilon, \quad \text{for all } k > k'. \tag{4.3.13}$$

Note that the condition $k > (10/3)\log(5/\varepsilon)$ comes from the fact that $s^2 < K^2$ and $s^2 = M^2k^{-1} = (10K^2/3)\log(5/\varepsilon)k^{-1}$. Consequently,

$$|\hat{\chi}_{ij} - \chi_{ij}(k/n)| = O_{\mathbb{P}}(k^{(-1/2)}).$$

□

**Proposition 4.3.9.** *Let* $\mathbf{X}_1, ..., \mathbf{X}_n$ *be independent copies of the d-dimensional random vector* $\mathbf{X} = (X_1, ..., X_d)$*, that is in the max-domain of attraction of a multivariate Pareto distribution* $\mathbf{Y}$ *(defined as in Theorem (2.1.27)) which is an extremal graphical model with respect to the forest* $\mathcal{F} = (V, E_{\mathcal{F}})$*. Let us denote the number of extreme observations by k. If* $k \to \infty$ *and* $k = o(n^{(2\alpha+1)/2\alpha})$*, then for any diverging sequence* $f_k$*, which satisfies*

$$\begin{aligned} \lim_{k\to\infty} \frac{1}{f_k} &= 0, \\ \lim_{k\to\infty} \frac{f_k}{k} &= 0, \end{aligned} \tag{4.3.14}$$

$\hat{\mathcal{F}}_{\chi,\tau_k}$ *consistently recovers the true underlying forest using a threshold* $\tau_k = -\log(M_k k^{(-1/2)})$*, where* $M_k := \sqrt{K'\log(5/\varepsilon_k)}$*,* $\varepsilon_k = 5\exp(-f_k/K')$*, and* $K' = (10K^2/3)$ *for the universal K in Proposition (4.3.7).*

$$\mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau_k} = \mathcal{F}\right) \to 1.$$

Let us denote the set of connected components of $\hat{\mathcal{F}}_{\chi,\tau_k}$ and $\mathcal{F}$, by $\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k})$ and $\mathfrak{C}(\mathcal{F})$ respectively. A tree in $\mathcal{F}$ is denoted by $\mathcal{T}$, and a tree in $\hat{\mathcal{F}}_{\chi,\tau_k}$ is denoted by $\hat{\mathcal{T}}_{\chi}$. We also denote $\chi_{\tau_k} := \exp(-\tau_k) = M_k k^{(-1/2)}$ that can be written as $\chi_{\tau_k} = \sqrt{\frac{f_k}{k}}$ by doing some calculus.

*Proof.* We must show that for any arbitrary sequence $f_k$ which satisfies conditions in (4.3.14), have the threshold $\tau_k$ for which,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau_k} = \mathcal{F}\right) = 1.$$

In the first step of proof, we show that,

$$\lim_{k\to\infty} \mathbb{P}(\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})) = 1.$$

Same as what we said in the proof of Proposition (4.3.2), the latter is equivalent to have that

$$\begin{aligned} \lim_{k\to\infty} \mathbb{P}\Big(&\{\text{for any unconnected nodes } p,q \in \mathcal{F} : p,q \text{ are unconnected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\} \wedge \\ &\{\text{for any connected nodes } i,j \in \mathcal{F} : i,j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\}\Big) = 1. \end{aligned} \tag{4.3.15}$$

We start with the first event in Equation (4.3.15). By Lemma (4.3.8), we have that

$$\mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}\,(k/n)\,| \geq M_k \cdot k^{(-1/2)}\right) \leq \varepsilon_k, \quad \text{for all } k > \frac{f_k}{K^2},$$

where $k > \dfrac{f_k}{K^2}$ comes from the conditions for Inequality (4.3.13). Therefore,

$$\mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}(k/n)| \geq \sqrt{\frac{f_k}{k}}\right) \leq 5\exp(-f_k/K'), \quad \text{for all } k > \frac{f_k}{K^2}.$$

The latter implies that

$$\mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}(k/n)| - |\chi_{pq}(k/n) - \chi_{pq}| \geq \sqrt{\frac{f_k}{k}} - |\chi_{pq}(k/n) - \chi_{pq}|\right) \leq 5\exp(-f_k/K'), \quad \text{for all } k > \frac{f_k}{K^2}.$$

By triangle inequality, we have that

$$|\hat{\chi}_{pq} - \chi_{pq}(k/n)| \leq |\hat{\chi}_{pq} - \chi_{pq}| + |\chi_{pq}(k/n) - \chi_{pq}|.$$

hence,

$$\mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}| \geq \sqrt{\frac{f_k}{k}} - |\chi_{pq}(k/n) - \chi_{pq}|\right) \leq 5\exp(-f_k/K'), \quad \text{for all } k > \frac{f_k}{K^2}. \tag{4.3.16}$$

As $k$ goes to infinity, the rate of convergence of $|\chi_{pq}(k/n) - \chi_{pq}|$ ((private communication, Lalancette, 2022)) and $5\exp(-f_k/K')$ to zero is faster than the rate of convergence of $f_k/k$ to zero. Therefore, Inequality (4.3.16) implies that

$$\lim_{k\to\infty} \mathbb{P}\left(|\hat{\chi}_{pq} - \chi_{pq}| \geq \sqrt{\frac{f_k}{k}}\right) = 0.$$

Hence,

$$\lim_{k\to\infty} \mathbb{P}(|\hat{\chi}_{pq} - \chi_{pq}| \geq \chi_{\tau_k}) = 0. \tag{4.3.17}$$

Moreover, by Lemma (4.2.1), for any unconnected nodes $p$ and $q$ in $\mathcal{F}$, we have that

$$\chi_{pq} = 0.$$

Thus, as an immediate consequence of Equation (4.3.17),

$$\lim_{k\to\infty} \mathbb{P}(\hat{\chi}_{pq} \geq \chi_{\tau_k}) = 0,$$

and

$$\lim_{k\to\infty} \mathbb{P}(\hat{\chi}_{pq} < \chi_{\tau_k}) = 1,$$

which implies that,

$$\lim_{k\to\infty} \mathbb{P}(\hat{w}_{pq} > \tau_k) = 1, \tag{4.3.18}$$

The latter shows that as $k$ goes to $\infty$, any two unconnected nodes $p$ and $q$ in $\mathcal{F}$ are not adjacent in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau_k}$ with probability equals to one; since the output of Algorithm (3) is restricted to have edges that are strictly lighter than the threshold $\tau_k$. By Lemma (4.3.1), the latter implies that any two unconnected nodes $p$ and $q$ in $\mathcal{F}$ are not connected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau_k}$ with probability equals to one. Therefore,

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any unconnected nodes } p, q \in \mathcal{F} : p, q \text{ are unconnected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\}\Big) = 1. \tag{4.3.19}$$

We consider the second event of Equation (4.3.15) in the next step. Inequality (4.3.16) implies that for any two nodes of the forest $\mathcal{F}$ say $i$ and $j$

$$\mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}| < \sqrt{\frac{f_k}{k}} - |\chi_{ij}(k/n) - \chi_{ij}|\right) > 1 - 5\exp(-f_k/K'), \quad \text{for all } k > \frac{f_k}{K^2}. \tag{4.3.20}$$

As $k$ goes to infinity, the rate of convergence of $|\chi_{pq}(k/n) - \chi_{pq}|$ ((private communication, Lalancette, 2022)) and $5\exp(-f_k/K')$ to zero is faster than the rate of convergence of $f_k/k$ to zero. Therefore, Inequality (4.3.20) implies that

$$\lim_{k\to\infty} \mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}| < \sqrt{\frac{f_k}{k}}\right) = 1.$$

Hence,

$$\lim_{k\to\infty} \mathbb{P}\left(|\hat{\chi}_{ij} - \chi_{ij}| < \chi_{\tau_k}\right) = 1.$$

The latter implies that

$$\lim_{k\to\infty} \mathbb{P}\left(-\chi_{\tau_k} < \hat{\chi}_{ij} - \chi_{ij} < \chi_{\tau_k}\right) = 1,$$

and

$$\lim_{k\to\infty} \mathbb{P}\left(\chi_{ij} - \chi_{\tau_k} < \hat{\chi}_{ij} < \chi_{ij} + \chi_{\tau_k}\right) = 1.$$

As an immediate consequence,

$$\lim_{k\to\infty} \mathbb{P}\left(\chi_{ij} - \chi_{\tau_k} < \hat{\chi}_{ij}\right) = 1. \tag{4.3.21}$$

Moreover, by Lemma (4.2.1), for any connected nodes $i$ and $j$ in $\mathcal{F}$, we have that

$$\chi_{ij} \neq 0.$$

Since as $k$ goes to $\infty$, $\chi_{\tau_k}$ goes to zero, there exists a positive $\varepsilon^*$ for which

$$\lim_{k\to\infty} \chi_{ij} - \chi_{\tau_k} > \varepsilon,$$

Thus,

$$\lim_{k\to\infty} \chi_{ij} - \varepsilon > \chi_{\tau_k},$$

Therefore, by Equation (4.3.21) we have that

$$\lim_{k\to\infty} \mathbb{P}\left(\chi_{\tau_k} < \chi_{ij} - \chi_{\tau_k} < \hat{\chi}_{ij}\right) = 1,$$

which implies that,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\chi}_{ij} > \chi_{\tau_k}\right) = 1.$$

As an immediate consequence,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{w}_{ij} < \tau_k\right) = 1. \tag{4.3.22}$$

The latter shows that as $k$ goes to $\infty$, nodes $i$ and $j$ are connected in the estimated forest $\hat{\mathcal{F}}_{\chi,\tau_k}$. By contradiction, suppose that $i$ and $j$ are unconnected; therefore, if we add $e_{ij}$ to the forest $\hat{\mathcal{F}}_{\chi,\tau_k}$, the created graph does not have a cycle and $\tau_k > \hat{w}_{ij}$ which guarantees that all edges in the created graph are lighter than $\tau_k$. However, the created graph is a forest that has one edge more than $\hat{\mathcal{F}}_{\chi,\tau_k}$ which contradicts the maximality of the number of edges of the estimated forest by Algorithm (3). Consequently, Equation (4.3.22) implies that $i$ and $j$ are connected in $\hat{\mathcal{F}}_{\chi,\tau_k}$. Therefore,

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any connected nodes } i,j \in \mathcal{F} : i,j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\}\Big) = 1. \quad (4.3.23)$$

By Equations (4.3.19) and (4.3.23) and independence of two events in Equation (4.3.15),

$$\lim_{k\to\infty} \mathbb{P}\Big(\{\text{for any unconnected nodes } p,q \in \mathcal{F} : p,q \text{ are unconnected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\}\wedge \\ \{\text{for any connected nodes } i,j \in \mathcal{F} : i,j \text{ are connected nodes in } \hat{\mathcal{F}}_{\chi,\tau_k}\}\Big) = 1.$$

As an immediate consequence,

$$\lim_{k\to\infty} \mathbb{P}\left(\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})\right) = 1. \quad (4.3.24)$$

By Theorem 2 of Engelke and Volgushev, 2020 we have that,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\mathcal{T}}_\chi = \mathcal{T} \middle| \mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})\right) = 1, \quad \text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}. \quad (4.3.25)$$

Hence, by Equations (4.3.25) and (4.3.24),

$$\begin{aligned}\lim_{k\to\infty} \mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau_k} = \mathcal{F}\right) &= \lim_{k\to\infty} \mathbb{P}\left(\{\text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}\text{: } \hat{\mathcal{T}}_\chi = \mathcal{T}\} \wedge \{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})\}\right) \\ &= \lim_{k\to\infty} \mathbb{P}\left(\{\text{for all trees } \mathcal{T} \text{ of the forest } \mathcal{F}\text{: } \hat{\mathcal{T}}_\chi = \mathcal{T}\} \middle| \{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})\}\right) \cdot \\ &\qquad \mathbb{P}\left(\{\mathfrak{C}(\hat{\mathcal{F}}_{\chi,\tau_k}) = \mathfrak{C}(\mathcal{F})\}\right) \\ &= 1 \cdot 1.\end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}\left(\hat{\mathcal{F}}_{\chi,\tau_k} = \mathcal{F}\right) = 1.$$

□

### 4.3.2 Forest recovery simulations

In the following, we study the finite sample behavior of the extremal forest estimator $\hat{\mathcal{F}}_{\chi,\tau}$, which is the weight-restricted minimum spanning forest obtained by Algorithm (3), using the empirical weights corresponding to the extremal correlations and some suitable threshold $\tau$. Let $\mathcal{F} = (V, E_{\mathcal{F}})$ be a random forest structure that is generated with the following instruction. We select two random numbers for the number of vertices and connected components, respectively denoted by $d$ and $c$ where $1 \le c \le d$. Then for each connected component (i.e., a tree), we randomly generate a Prüfer sequence and make its corresponding tree. This procedure generates a random forest structure.

Then, we simulate $n$ samples $\mathbf{X}_1, ..., \mathbf{X}_n$ from a random vector $\mathbf{X} = (X_1, ..., X_d)$ in the domain of attraction of a multivariate Pareto distribution $\mathbf{Y}$ that is an extremal graphical model with respect to the forest $\mathcal{F}$ where $|V| = d$. Precisely, we consider the random vector $\mathbf{X}$ as a perturbed max-stable distribution which is in the max-domain of attraction of $\mathbf{Y}$.

This perturbation is essential. Particularly, Algorithm (3) is based on the comparison between extremal correlations. In each step, the algorithm adds an edge with the largest extremal correlation, which does not create any cycle with the previously selected edges. The estimated forest is based on the empirical pre-asymptotic extremal correlation. By Equation (2.6.6), we know that the bias of the estimator $\hat{\chi}(k/n)$ is controlled by the pre-asymptotic extremal correlation $\chi(k/n)$; therefore, the bias is proportional to the $k/n$. Moreover, Lemma (4.3.8) shows