

Recursive max-linear models with propagating noise

Johannes Buck

*Center for Mathematical Sciences,
Technical University of Munich,
Boltzmanstrasse 3,
85748 Garching, Germany
e-mail: j.buck@tum.de*

and

Claudia Klüppelberg

*Center for Mathematical Sciences,
Technical University of Munich,
Boltzmanstrasse 3,
85748 Garching, Germany
e-mail: cklu@ma.tum.de*

Abstract: Recursive max-linear vectors model causal dependence between node variables by a structural equation model, expressing each node variable as a max-linear function of its parental nodes in a directed acyclic graph (DAG) and some exogenous innovation. For such a model, there exists a unique minimum DAG, represented by the Kleene star matrix of its edge weight matrix, which identifies the model and can be estimated. For a more realistic statistical modeling we introduce some random observational noise. A probabilistic analysis of this new noisy model reveals that the unique minimum DAG representing the distribution of the non-noisy model remains unchanged and identifiable. Moreover, the distribution of the minimum ratio estimators of the model parameters at their left limits are completely determined by the distribution of the noise variables up to a positive constant. Under a regular variation condition on the noise variables we prove that the estimated Kleene star matrix converges to a matrix of independent Weibull entries after proper centering and scaling.

MSC2020 subject classifications: 60G70, 62F12, 62G32, 62H22.

Keywords and phrases: Graphical model, Bayesian network, directed acyclic graph, extreme value analysis, max-linear model, noisy model, regular variation.

Received May 2020.

1. Introduction

Graphical modeling has shown to be a powerful tool for understanding causal dependencies in a multivariate random vector. However, most models are linear and limited to discrete or Gaussian distributions (see e.g. [25] and [27]). Such models lead to severe underestimation of large risks and, therefore, are not suitable in the context of extreme risk assessment. First examples combining

extreme value methods with graphical models include flooding in river networks ([12]), financial risk ([10], [23]), and nutrients ([23]).

We consider the class of recursive max-linear (ML) models, which has been defined in [14]. A recursive ML model is defined by a structural equation model (SEM) of the form

$$X_i = \bigvee_{j \in \text{pa}(i)} c_{ji} X_j \vee Z_i, \quad i = 1, \dots, d \quad (1.1)$$

where the dependence structure between random variables is represented by a DAG $\mathcal{D} := (V, E)$ with node set $V := \{1, \dots, d\}$ and edge set $E = E(\mathcal{D}) \subseteq V \times V$, and each variable X_i for $i \in V$ has a representation in terms of ML functions of its parental nodes $\text{pa}(i) = \{j \in V : (j, i) \in E\}$ and an independent innovation Z_i .

Both, SEMs (e.g. [5], [31]) and directed graphical models (e.g. [25], [27], [35]) are well-established models and widely used to understand causality.

ML models similar to (1.1) have been proposed and studied in a time series context (e.g. [9]), in terms of moving maxima processes (e.g. [16]), or as tropical models in algebra (e.g. [21], [30]) with applications to various optimization problems (e.g. [2], [6], [36]).

As shown in [24], recursive ML models respect the basic Markov properties associated with DAGs (e.g. [28], [29]). Moreover, the equation system (1.1) has the solution

$$X_i = \bigvee_{j \in \text{pa}(i)} b_{ji} Z_j, \quad i = 1, \dots, d, \quad (1.2)$$

with ML coefficient matrix (in tropical algebra called the Kleene star matrix) $\mathbf{B} := (b_{ij})_{d \times d}$, see [6], Corollary 1.6.16. Unlike the edge weight matrix $\mathbf{C} := (c_{ij})_{d \times d}$, \mathbf{B} is identifiable and completely determines the distribution of $\mathbf{X} := (X_1, \dots, X_d)$ (see [15], Theorem 1). Also, \mathbf{B} is idempotent with respect to the tropical matrix multiplication defined in (2.4) below, and defines a graphical model on a DAG with node set V and an edge $j \rightarrow i$ whenever there is a path from j to i in \mathcal{D} . Furthermore, [15] proposes a minimum ratio estimator for \mathbf{B} , which itself is idempotent, and is a generalized maximum likelihood estimator in the sense of [22].

The model (1.1) states that an extreme node observation X_i in the DAG is either the result of a large external innovation Z_i , or the weighted maximum of observations from the parent nodes of i in \mathcal{D} . As we see from the solution (1.2), all past innovations drive this observation. Our aim is to generalise this rather restricted recursive structure as to allow for certain observation errors, independent of this model.

More precisely, we extend the original model (1.1) by allowing for multiplicative observation errors and define

$$U_i = \left(\bigvee_{j \in \text{pa}(i)} c_{ji} U_j \vee Z_i \right) \varepsilon_i, \quad i = 1, \dots, d, \quad (1.3)$$

with $\varepsilon_i \geq 1$ and iid for $i = 1, \dots, d$. By taking advantage of tropical algebra, we

present in Theorem 3.2 a solution of (1.3) which represents each node variable U_i in terms of a ML function of its ancestral nodes and an independent innovation Z_i given by

$$U_i = \bigvee_{j \in \text{an}(i) \cup \{i\}} \bar{b}_{ji} Z_j, \quad i = 1, \dots, d,$$

where $\text{an}(i)$ denotes the ancestors of i and \bar{b}_{ji} are random variables involving the edge weights and the noise variables.

It comes as no surprise that the true DAG and edge weights for a recursive ML model with propagating noise inherit the non-identifiability property from the non-noisy model. However, as we will prove in Section 4, the ML coefficient matrix $\mathbf{B} = (b_{ij})_{d \times d}$ remains identifiable in spite of the observational noise and even if we do not know the underlying DAG.

To link up our new model (1.3) with existing literature, observe that a log-transformation of (1.3) yields

$$\tilde{U}_i = \bigvee_{j \in \text{pa}(i)} (\tilde{c}_{ji} + \tilde{U}_j) \vee \tilde{Z}_i + \tilde{\varepsilon}_i, \quad i = 1, \dots, d \quad (1.4)$$

with $\tilde{\varepsilon}_i \geq 0$. However, due to the maximum operator, the class of max-linear models is highly non-smooth such that most standard methods do not apply. For every $j \in \text{pa}(i)$, the difference $\tilde{U}_i - \tilde{U}_j$ is lower-bounded by \tilde{c}_{ji} and

$$\mathbb{P}(\tilde{U}_i - \tilde{U}_j \leq \tilde{c}_{ji} + x \mid \tilde{U}_i = \tilde{c}_{ji} + \tilde{U}_j + \tilde{\varepsilon}_i) = \mathbb{P}(\tilde{\varepsilon}_i \leq x).$$

Example 1.2.2 of [6] might then serve as motivating example for model (1.4). Assume that i is a priority flight and there are several feeder flights to i . Assume that the departure of a feeder flight j is delayed by X_j , and there are passengers, who have to catch flight i . Moreover, let c_{ji} be the departure time of flight i minus arrival time of flight j minus transit time according to schedule. There may be a further delay Z_i of flight i caused by non-connecting passengers. Assuming that flight i also waits for such non-connecting passengers, the delay of flight i is given by $\tilde{U}_i = \bigvee_{j \in \text{pa}(i)} (\tilde{c}_{ji} + \tilde{U}_j) \vee \tilde{Z}_i$. Some other delays are independent of possible passenger like bad weather conditions, delayed start clearance etc. Then we find exactly model (1.4), the max-linear model with positive noise.

The estimation of (linear) functions with one-sided errors has been considered in the literature before. For instance, in [17] and [20] observations are given by $Y_j = f(X_j) + \varepsilon_j$ for $j = 1, \dots, n$ with observation errors $\varepsilon_j > 0$, with density given conditionally or unconditionally on $X_j = x$, and f describes some frontier or boundary curve, which has to be estimated. To present an archetypical example, consider the linear regression problem stated in [33] and [34] as $Y_i = \beta + \varepsilon_i$ for $i = 1, \dots, n$ and observation errors, which have density $g(x) \sim \alpha c x^{\alpha-1}$ as $x \downarrow 0$ for $\alpha, c > 0$. In these papers, the focus is on the non-regular case, when $\alpha < 2$. Then β can be estimated by the sample minimum $Y_{1,n}$ which has a Weibull limit law:

$$\lim_{n \rightarrow \infty} \mathbb{P}((nc)^{-1/\alpha} (Y_{1,n} - \beta_0) \leq x) = 1 - \exp(-x^{-\alpha}), \quad 0 < x < \infty. \quad (1.5)$$

The work in [33] has been used in [8] to estimate the coefficient ϕ of a first order autoregressive time series with positive innovations. They propose the minimum ratio estimator $\hat{\phi} = \bigwedge_{j=1}^n X_j/X_{j-1}$ and show in their Corollary 2.4 that it also has a Weibull limit law similar to (1.5).

In our model (1.3) we find two interpretations for the noise variables. Firstly, in the log-transformed version (1.4) we consider a ML model as baseline model, which is observed with some additive noise. A second representation is given in Corollary 3.3 below, where the edge and path weights become noisy by the noise variables. This gives rise to the interpretation that we observe the model parameters with noise similarly as in the regression examples above. As a consequence, a path from j to i realizing the ML coefficient b_{ji} is no longer deterministic but depends on the individual realizations of the noise variables. However, in Theorem 3.12 we show that at the left limit of support the distribution of the ratio of two model components is determined by all noise variables along the path between the two nodes. Assuming noise variables with regularly varying distribution in their left limit of support, we propose a minimum ratio estimator and show in Theorem 5.2 that the estimated ML coefficient matrix converges to a matrix of independent Weibull entries after proper centering and rescaling.

The paper is organized as follows. In Section 2, we summarize the properties of recursive ML models as defined in (1.1) and state the most important results relevant for our paper. In Section 3 we consider the extension of the recursive ML model given in (1.3), which we coin the *max-linear model with propagating noise* and present its solution and the main properties of this new model. In Section 4 we address the identifiability of the ML model with propagating noise. Similarly as in (1.5) we suggest minimum ratio estimators for the model parameters \mathbf{B} . In Section 5 we assume regular variation of the noise variables. Under this assumption, we show that the minimum ratios are asymptotically independent and Weibull distributed. Finally, in Section 6, we provide a data example and apply the theory that we have derived in the previous sections. All proofs are postponed to an Appendix.

Throughout we use the following notation. $\mathbb{R}_+ = (0, \infty)$ and $\overline{\mathbb{R}}_+ = [0, \infty)$, $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$ with $\bigwedge_{i \in \emptyset} x_i = \infty$ and $\bigvee_{i \in \emptyset} x_i = 0$ for $x_i \in \mathbb{R}_+$. Bold letters denote vectors and matrices, e.g. \mathbf{I}_d denotes the $d \times d$ identity matrix. Moreover, all vectors are row vectors unless stated otherwise. For two functions f, g we write $f(x) \sim g(x)$ as $x \downarrow c$ if $\lim_{x \downarrow c} f(x)/g(x) = 1$ and $\mathbf{1}$ denotes the indicator function. For a random variable Y with distribution function F_Y , the symbol F_Y^{\leftarrow} denotes its *quantile function*.

2. Preliminaries — recursive max-linear models

2.1. Graph terminology

We use the same graph notation as in [14]. A *directed graph* is a pair (V, E) of a *node set* $V = \{1, \dots, d\}$ and an *edge set* $E = \{j \rightarrow i : i, j \in V, i \neq j\}$. A node j is called a *parent* of i if $j \rightarrow i \in E$ and write $(j, i) \in E$. A (*directed*) *path* from j to

i is a sequence of distinct nodes $[j = k_0, k_1, \dots, k_n = i]$ such that $k_{r-1} \rightarrow k_r$ for each $r \in \{1, \dots, n\}$, and a directed cycle is a path where $j = i$. A node j is called an *ancestor* of i , if there exists a path from j to i , then i is called a *descendant* of j . The node sets $\text{pa}(i)$, $\text{an}(i)$ and $\text{de}(i)$ denote the parents, ancestors, and the descendants of node i , respectively, and we abbreviate $\text{An}(i) := \text{an}(i) \cup \{i\}$.

Finally, for a path $p = [k_0, k_1, \dots, k_n]$ we define the *node set on the path (excluding the initial node)* by $S_p := \{k_1, \dots, k_n\}$ and its *path length* by $|S_p|$.

Throughout this paper $\mathcal{D} = (V, E)$ is a directed acyclic graph (DAG), and we recall that a complete DAG is a complete graph with directed edges.

A matrix $\mathbf{C} \in \overline{\mathbb{R}}_+^{d \times d}$ defines a *weighted directed graph*, where $j \rightarrow i \in \mathcal{D}$ if and only if its *edge weight* c_{ji} is positive. The *path weight* of a path in \mathcal{D} is then the product of its edge weights.

For a DAG \mathcal{D} on V with edge weight matrix \mathbf{C} , its *reachability DAG* is defined as a DAG on V having edge $j \rightarrow i$ if and only if \mathcal{D} has a path from $j \rightarrow i$; moreover, \mathbf{B} represents the edge weight matrix of the reachability DAG. We call \mathbf{B} the *ML coefficient matrix* and remark that it is a *weighted reachability matrix* for \mathcal{D} .

2.2. Recursive max-linear models

We first formally introduce the class of recursive ML models and state their most important results for this paper. Let $\mathcal{D} = (V, E)$ be a DAG. Then a random vector $\mathbf{X} := (X_1, \dots, X_d)$ is a *recursive max-linear vector* or follows a *max-linear Bayesian network* on \mathcal{D} if

$$X_i := \bigvee_{k \in \text{pa}(i)} c_{ki} X_k \vee Z_i, \quad i \in 1, \dots, d, \quad (2.1)$$

with positive edge weights c_{ki} for $i \in V$ and $k \in \text{pa}(i)$, and independent positive random variables Z_1, \dots, Z_d with support \mathbb{R}_+ and atom-free distributions. We shall refer to $\mathbf{Z} := (Z_1, \dots, Z_d)$ as the *vector of innovations*.

For a path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ from j to i we define the *path weight*

$$d_{ji}(p) := \prod_{l=0}^{n-1} c_{k_l k_{l+1}}. \quad (2.2)$$

Denoting the set of all paths from j to i by P_{ji} , we define the ML coefficient matrix $\mathbf{B} = (b_{ij})_{d \times d}$ of \mathbf{X} with entries

$$b_{ij} := \bigvee_{p \in P_{ij}} d_{ij}(p) \quad \text{for } i \in \text{an}(j), \quad b_{ii} = 1, \quad \text{and} \quad b_{ij} = 0 \quad \text{for } i \in V \setminus \text{An}(j).$$

The components of \mathbf{X} can also be expressed as ML functions of their ancestral innovations and an independent one; the corresponding ML coefficients are the

entries of \mathbf{B} :

$$X_i = \bigvee_{k \in \text{An}(i)} b_{ki} Z_k, \quad i \in 1, \dots, d, \quad (2.3)$$

which can be shown by a path analysis as in Theorem 2.2 in [14] or by tropical algebra as in (2.6) below, and as we explain now.

For two non-negative matrices \mathbf{F} and \mathbf{G} , where the number of columns in \mathbf{F} is equal to the number of rows in \mathbf{G} , we define the matrix product $\odot : \overline{\mathbb{R}}_+^{m \times n} \times \overline{\mathbb{R}}_+^{n \times p} \rightarrow \overline{\mathbb{R}}_+^{m \times p}$ by

$$(\mathbf{F} = (f_{ij})_{m \times n}, \mathbf{G} = (g_{ij})_{n \times p}) \mapsto \mathbf{F} \odot \mathbf{G} := \left(\bigvee_{k=1}^n f_{ik} g_{kj} \right)_{m \times p}. \quad (2.4)$$

The triple $(\overline{\mathbb{R}}_+, \vee, \cdot)$, is an idempotent semiring with 0 as 0-element and 1 as 1-element and the operation \odot is therefore a matrix product over this semiring; see for example [6]. Denoting by \mathcal{M} all $d \times d$ matrices with non-negative entries and by \vee the componentwise maximum between two matrices, $(\mathcal{M}, \vee, \odot)$ is also a semiring with the null matrix as 0-element and the $d \times d$ identity matrix \mathbf{I}_d as 1-element.

The matrix product \odot allows us to represent the ML coefficient matrix \mathbf{B} of \mathbf{X} in terms of the edge weight matrix $\mathbf{C} := (c_{ij} \mathbf{1}_{\text{pa}(j)}(i))_{d \times d}$ of \mathcal{D} , since (2.1) can be rewritten as

$$\mathbf{X} = (\mathbf{X} \odot \mathbf{C}) \vee \mathbf{Z} \quad (2.5)$$

with unique solution (equivalent to (2.3)) given by

$$\mathbf{B} = (\mathbf{I}_d \vee \mathbf{C})^{\odot(d-1)} = \bigvee_{k=0}^{d-1} \mathbf{C}^{\odot k}, \quad \mathbf{X} = \mathbf{Z} \odot \mathbf{B}, \quad (2.6)$$

where \mathbf{B} is the *Kleene star matrix* and $\mathbf{A}^{\odot 0} = \mathbf{I}_d$ and $\mathbf{A}^{\odot k} = \mathbf{A}^{\odot(k-1)} \odot \mathbf{A}$ for $\mathbf{A} \in \overline{\mathbb{R}}_+^{d \times d}$ and $k \in \mathbb{N}$; see Proposition 1.6.15 of [6] as well as Theorem 2.4 and Corollary 2.5 of [14]. For more information on the max-times (tropical) algebra in ML models, see Section 2.2 in [1].

We have seen that a recursive ML vector \mathbf{X} has two representations, one in terms of parental nodes X_j and edge weights c_{ji} and another in terms of innovations Z_j and ML coefficients b_{ji} . However, while the ML coefficient matrix \mathbf{B} of \mathbf{X} is identifiable from the distribution of \mathbf{X} , the edge weight matrix \mathbf{C} is generally not, see Theorem 5.4(b) in [14]. Theorem 5.3 in that paper and Theorem 2 in [15] show that an edge with edge weight c_{ji} is identifiable from \mathbf{B} if and only if it is the unique path from j to i with $d_{ji}(p) = b_{ji}$.

For a recursive ML vector \mathbf{X} on a DAG $\mathcal{D} = (V, E)$ and ML coefficient matrix \mathbf{B} this result leads to the following definition.

Definition 2.1. Let $\mathbf{X} \in \mathbb{R}_+^d$ be a recursive ML vector on the DAG $\mathcal{D} = (V, E)$ with ML coefficient matrix \mathbf{B} . We define the minimum ML DAG of \mathbf{X} as

$$\mathcal{D}^B = (V, E^B) := \left(V, \left\{ (j, i) \in E : b_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \right\} \right).$$

Moreover, it has been shown that the support of a ratio of components of a recursive ML vector \mathbf{X} satisfies

$$\text{supp}(X_i/X_j) = \begin{cases} [b_{ji}, \infty) & \text{for } j \in \text{an}(i), \\ [0, 1/b_{ij}] & \text{for } i \in \text{an}(j), \\ \{1\} & \text{for } i = j, \\ \mathbb{R}_+ & \text{otherwise,} \end{cases} \quad (2.7)$$

with $\mathbb{P}(X_i/X_j = b_{ji}) > 0$ for all $j \in \text{an}(i)$; see Lemma 1 of [15]. Hence, for a given iid sample $\mathbf{X}^1, \dots, \mathbf{X}^n$ from \mathbf{X} define a minimum ratio estimator $\hat{\mathbf{B}}$ of \mathbf{B} by $\hat{b}_{ij} := \bigwedge_{k=1}^n (X_i^k/X_j^k)$ for $i, j \in V$. Moreover, when the DAG \mathcal{D} is known, we define \mathbf{B}_0 by

$$\mathbf{B}_0 = (B_0(i, j))_{d \times d} := \left(\bigwedge_{k=1}^n \frac{X_j^k}{X_i^k} \mathbf{1}_{\text{pa}(j)}(i) \right)_{d \times d} \quad \text{and set} \quad \hat{\mathbf{B}} = (\mathbf{I}_d \vee \mathbf{B}_0)^{\odot(d-1)}.$$

Theorem 4 of [15] ensures that $\hat{\mathbf{B}}$ is a generalized maximum likelihood estimate (GMLE) in the sense of [22].

2.3. Minimum domain of attraction and regular variation

Introducing propagating noise into the recursive ML model will smooth out the atoms in (2.7). The minimum ratio estimators will still estimate the left endpoints $\mathbf{B} = (b_{ji})$ and we will be able to provide distributional limit results. These will be based on minimum domain of attraction results and regular variation. Extreme value theory is more focused nowadays on running maxima (e.g. [11]), but results for running minima are obtained by noting that $\bigwedge_{i=1}^n Y_i = -\bigvee_{i=1}^n (-Y_i)$. From this we obtain that the family of Weibull distributions are limit distributions of running minima of i.i.d. random variables, see equation (2.9).

Definition 2.2. A positive random variable Ψ_α is Weibull distributed with left endpoint $x_L > -\infty$, shape $\alpha > 0$ and scale $s > 0$ and we write $Y \sim \text{Weibull}(\alpha, x_L, s)$ if the distribution function of Y is given by

$$\Psi_{\alpha, x_L, s}(x) = 1 - \exp \left(- \left(\frac{x - x_L}{s} \right)^\alpha \right), \quad x \geq x_L.$$

Random variables, whose running minima have such a Weibull limit satisfy certain conditions. Here the following definition is essential and we refer to [4] for details.

Definition 2.3. Let Y be a random variable with distribution function F and left endpoint x_L . Then we call Y or F regularly varying at x_L with exponent $\alpha > 0$, if

$$\lim_{t \downarrow 0} \frac{F(x_L + tx)}{F(x_L + t)} = x^\alpha, \quad x > 0. \quad (2.8)$$

We abbreviate this by $Y \in RV_\alpha^{x_L}$ or $F \in RV_\alpha^{x_L}$, respectively. We also note that $Y \in RV_\alpha^{x_L}$ is equivalent to $Y - x_L \in RV_\alpha^0$.

Then, adapting Theorem 3.3.12 of [11] to the minimum of i.i.d. random variables X_1, \dots, X_d with distribution function F , we obtain

$$\begin{aligned} & \exists (a_n > 0) \text{ s.t. } \frac{1}{a_n} \left(\bigwedge_{i=1}^n X_i - x_L \right) \xrightarrow{d} \Psi_\alpha, \quad n \rightarrow \infty \\ \iff & x_L > -\infty, F(x_L + \cdot) \in RV_\alpha^0. \end{aligned} \quad (2.9)$$

Let ε be a random variable with left endpoint $x_L = 1$, and $\tilde{\varepsilon} := \ln(\varepsilon)$. Then for $x > 0$,

$$\begin{aligned} \lim_{t \downarrow 0} \frac{P(\ln(\varepsilon) \leq tx)}{P(\ln(\varepsilon) \leq t)} &= \lim_{t \downarrow 0} \frac{P(\varepsilon \leq e^{tx})}{P(\varepsilon \leq e^t)} = \lim_{t \downarrow 0} \frac{P(\varepsilon - 1 \leq tx(1 + o(1)))}{P(\varepsilon - 1 \leq t(1 + o(1)))} \\ &= \lim_{t \downarrow 0} \frac{P(\varepsilon - 1 \leq tx)}{P(\varepsilon - 1 \leq t)}, \end{aligned} \quad (2.10)$$

such that $\varepsilon \in RV_\alpha^1$ if and only if $\tilde{\varepsilon} \in RV_\alpha^0$. Two relevant families of distribution functions are given in the next example.

Example 2.4. (a) [Weibull distribution] Let $\tilde{\varepsilon}$ have distribution function as in Definition 2.2 with $x_L = 0$. Then by a l'Hospital argument,

$$\lim_{t \downarrow 0} \frac{\Psi_{\alpha,s}(tx)}{\Psi_{\alpha,s}(t)} = x^\alpha,$$

which implies that $\tilde{\varepsilon} \in RV_\alpha^0$ and $\varepsilon \in RV_\alpha^1$.

(b) [Gamma distribution] Let $\tilde{\varepsilon}$ have density $g(x) = \lambda^\alpha e^{-\lambda x} x^{\alpha-1} / \Gamma(\alpha)$ for $x > 0$ and parameters $\lambda > 0, \alpha > 0$. Then by a l'Hospital argument,

$$\lim_{t \downarrow 0} \frac{G(tx)}{G(t)} = \lim_{t \downarrow 0} \frac{e^{-\lambda tx} t^{\alpha-1} x^\alpha}{e^{-\lambda t} t^{\alpha-1}} = x^\alpha, \quad x > 0,$$

which implies that $\tilde{\varepsilon} \in RV_\alpha^0$ and $\varepsilon \in RV_\alpha^1$. \square

We provide here also some preliminary results.

Proposition 2.5 (Karamata's Tauberian Theorem 1.7.1, [4]). Let U be a non-decreasing function on \mathbb{R} with $U(x) = 0$ for all $x < 0$, and Laplace-Stieltjes transform $\hat{U}(s) = \int_{[0, \infty)} e^{-sx}$

$dU(x) < \infty$ for all large s . For $l \in RV_0^\infty$ and $c \geq 0, \rho \geq 0$, the following are equivalent

$$\begin{aligned} U(x) &\sim cx^\rho l(1/x)/\Gamma(1 + \rho), & x \downarrow 0 \\ \hat{U}(s) &\sim cs^{-\rho} l(s), & s \rightarrow \infty. \end{aligned} \tag{2.11}$$

From this, we obtain the following corollary.

Corollary 2.6. a) Let $X \in RV_{\alpha_1}^0, Y \in RV_{\alpha_2}^0$ be independent, then $X + Y \in RV_{\alpha_1 + \alpha_2}^0$,
 b) Let $X, Y \geq 1$ be independent and such that $\tilde{X} = \ln(X) \in RV_{\alpha_1}^0, \tilde{Y} = \ln(Y) \in RV_{\alpha_2}^0$. Then $(XY - 1) \in RV_{\alpha_1 + \alpha_2}^0$.

PROOF OF COROLLARY 2.6 (a) We use Proposition 2.5 a) for U being F_X or F_Y , the distribution function of X or Y , respectively. Since the Laplace-Stieltjes transforms $\hat{F}_X(s) = \int_{[0, \infty)} e^{-sx} dF_X(x) \leq 1$ and $\hat{F}_Y(s) = \int_{[0, \infty)} e^{-sx} dF_Y(x) \leq 1$ for all $s \geq 0$, by Proposition 2.5, they are both regularly varying at ∞ in the sense of (2.11); i.e., $\hat{F}_X \in RV_{\alpha_1}^\infty, \hat{F}_Y \in RV_{\alpha_2}^\infty$. By independence, the convolution theorem for Laplace-Stieltjes transforms gives $\hat{F}_{X+Y}(s) = \hat{F}_X(s)\hat{F}_Y(s)$ and, therefore, $\hat{F}_{X+Y} \in RV_{\alpha_1 + \alpha_2}^\infty$. Applying again Proposition 2.5 we find that $X_1 + X_2 \in RV_{\alpha_1 + \alpha_2}^0$.
 (b) This follows from a Taylor expansion. □

3. Recursive ML model with propagating noise

In this section we define the recursive ML model with propagating noise, present structural results, investigate which properties of the non-noisy model prevail, and derive distributional results for component ratios of the model in preparation for the structure learning results to follow.

3.1. Definitions and representations

Definition 3.1. A vector $\mathbf{U} \in \mathbb{R}_+^d$ is a recursive ML vector with propagating noise on a DAG $\mathcal{D} = (V, E)$, if

$$U_i := \left(\bigvee_{k \in \text{pa}(i)} c_{ki} U_k \vee Z_i \right) \varepsilon_i, \quad i \in 1, \dots, d, \tag{3.1}$$

with edge weight matrix $\mathbf{C} := (c_{ij} \mathbf{1}_{\text{pa}(j)}(i))_{d \times d}$. The noise variables $\varepsilon_1, \dots, \varepsilon_d$ are iid and atom-free random variables with $\varepsilon_i \geq 1$ and unbounded above for all $i \in V$, and independent of the innovations vector $\mathbf{Z} := (Z_1, \dots, Z_d)$. For simplicity, we denote by ε a generic noise variable and by Z a generic innovation.

Although the noise variables act on the observations, formally we can view them as random scalings of edge weights. More precisely, for a path $p = [j =$

$k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ from j to i we define the *random path weight* \bar{d}_{ji} similarly to the definition of d_{ji} in (2.2) as

$$\bar{d}_{ji}(p) := \varepsilon_j \prod_{l=0}^{n-1} c_{k_l k_{l+1}} \varepsilon_{k_{l+1}} = d_{ji}(p) \varepsilon_j \prod_{l=0}^{n-1} \varepsilon_{k_{l+1}}. \tag{3.2}$$

If we define the *random edge weight matrix*

$$\bar{\mathbf{C}} = (\bar{c}_{ij})_{d \times d} := (c_{ij} \varepsilon_j \mathbf{1}_{\text{pa}(j)}(i))_{d \times d} \tag{3.3}$$

we can rewrite (3.2) as

$$\bar{d}_{ji}(p) := \varepsilon_j \prod_{l=0}^{n-1} \bar{c}_{k_l k_{l+1}}$$

for every path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ from j to i . Hence, we can view the noise variables as random scalings for the edge weights c_{ji} . Since $\varepsilon \geq 1$, the edge weights c_{ji} of the non-noisy model are lower bounds for the random edge-weights \bar{c}_{ji} of the propagating noise model.

Again denoting the set of all paths from j to i by P_{ji} , we define the *random ML coefficient matrix* $\bar{\mathbf{B}} = (\bar{b}_{ij})_{d \times d}$ of \mathbf{U} with entries

$$\bar{b}_{ji} := \bigvee_{p \in P_{ji}} \bar{d}_{ji}(p) \quad \text{for } j \in \text{an}(i), \quad \bar{b}_{jj} = \varepsilon_j, \quad \text{and} \quad \bar{b}_{ji} = 0 \quad \text{for } j \in V \setminus \text{An}(i). \tag{3.4}$$

We next show that there exists a solution of (3.1) in terms of the ancestral innovations \mathbf{Z} and $\bar{\mathbf{B}}$. All proofs of this section are postponed to Appendix A.

Theorem 3.2. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as in (3.1). Define $(\mathbf{E}_d)_{d \times d}$ as the diagonal matrix given by*

$$E_d(i, i) = \varepsilon_i \quad \text{for } i \in V \quad \text{and} \quad E_d(i, j) = 0 \quad \text{for } i, j \in V \text{ and } i \neq j.$$

We rewrite (3.1) in matrix form by means of the matrix multiplication (2.4) as

$$\mathbf{U} = (\mathbf{U} \odot \mathbf{C} \vee \mathbf{Z}) \odot \mathbf{E}_d.$$

Then \mathbf{U} has a unique solution in terms of the tropical matrix multiplication with random matrix $\bar{\mathbf{B}}$ given by

$$\bar{\mathbf{B}} = \mathbf{E}_d \odot (\mathbf{I}_d \vee \bar{\mathbf{C}})^{\odot(d-1)}, \quad \mathbf{U} = \mathbf{Z} \odot \bar{\mathbf{B}}, \tag{3.5}$$

with $\bar{\mathbf{C}}$ as defined in (3.3).

Since $\bar{b}_{ji} = 0$ whenever $j \notin \text{An}(i)$, the representation (3.5) can be rewritten as follows.

Corollary 3.3. *Let \mathbf{U} be as in Theorem 3.2 and \bar{b}_{ji} be the random ML coefficients defined in (3.4). Then (3.5) is equivalent to*

$$U_i = \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} Z_j, \quad i \in 1, \dots, d. \quad (3.6)$$

Note that the definition in (3.1) is equivalent to

$$U_i = \tilde{U}_i \varepsilon_i \quad \text{with} \quad \tilde{U}_i := \bigvee_{k \in \text{pa}(i)} c_{ki} U_k \vee Z_i, \quad i \in 1, \dots, d. \quad (3.7)$$

Since \mathbf{B} is idempotent, the solution of \mathbf{X} as in (2.6) can also be written as $\mathbf{X} = \mathbf{X} \odot \mathbf{B}$. This is no longer the case for the solution \mathbf{U} in (3.5). However, from the above result we can compute the following representation, which is used in Definition 3.5(e) below.

Corollary 3.4. *Let \mathbf{U} and \bar{b}_{ji} be as in Corollary 3.3. Then (3.6) is equivalent to*

$$U_i = \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} \tilde{U}_j, \quad i = 1, \dots, d, \quad (3.8)$$

with \tilde{U}_j as in (3.7).

3.2. Paths classification and graph reduction in the noisy model

We define critical and generic paths which play an essential role for the distributional properties of the model. Similarly as shown for the non-noisy model in Section 5 of [14], the vector \mathbf{U} may also be a recursive ML model on a subgraph of \mathcal{D} . This subgraph depends on the ML coefficients, which are now random. Hence, we start by comparing the ML coefficient matrices \mathbf{B} and $\bar{\mathbf{B}}$ of the non-noisy and noisy models.

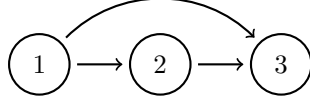
Definition 3.5. *Let \mathcal{D} be a DAG with edge weight matrix \mathbf{C} and let \mathbf{B} be the corresponding ML coefficient matrix (i.e., the Kleene star of \mathbf{C}). Let p be a path from j to i with node set S_p .*

- (a) p is called a (non-random) critical path if $d_{ji}(p) = b_{ji}$.
- (b) p is called a generic path if it is the only path satisfying $d_{ji}(p) = b_{ji}$.
- (c) We call \mathbf{C} generic, if two nodes are connected by at most one critical path.
- (d) For a fixed $\omega \in \Omega$, we call p a random critical path if $\bar{d}_{ji}(p) = \bar{b}_{ji}$.
- (e) p is called a possible critical path realization, if $U_i = U_j d_{ji}(p) \prod_{k \in S_p} \varepsilon_k = \tilde{U}_j \bar{d}_{ji}(p)$ happens with positive probability.

Remark 3.6. *We have defined a non-random critical path and a random critical path. We want to emphasize, however, that while the first path property is simply inherited from \mathbf{C} via \mathbf{B} , the second one is inherited from \mathbf{C} and the noise variables. We also note that by continuity of the innovations and the noise variables, any random critical path between a pair of nodes must be a.s. unique, although it may vary with the realizations of the noise variables.*

We explain the model and the notions of Definition 3.5 in an example.

Example 3.7. Consider the DAG:



Then, \mathbf{C} is generic if and only if $c_{13} \neq c_{12}c_{23}$. Moreover, we have

$$U_3 = (\bar{c}_{13} \vee \bar{c}_{12}\bar{c}_{23})\varepsilon_1 Z_1 \vee \bar{c}_{23}\varepsilon_2 Z_2 \vee \varepsilon_3 Z_3,$$

with $\bar{c}_{ji} = c_{ji}\varepsilon_i$ as defined in (3.3).

Now assume that $c_{13} > c_{12}c_{23}$. In that case, $[1 \rightarrow 3]$ is the critical path, while the path $[1 \rightarrow 2 \rightarrow 3]$ is not critical. However, $\mathbb{P}(\bar{c}_{13} < \bar{c}_{12}\bar{c}_{23}) = \mathbb{P}(\varepsilon_2 > c_{13}/(c_{12}c_{23})) > 0$. If $\mathbb{P}(\bar{c}_{13} < \bar{c}_{12}\bar{c}_{23})$, then the edge $1 \rightarrow 3$ is random critical, otherwise $1 \rightarrow 2 \rightarrow 3$ is random critical. Since both paths can be random critical with positive probability, all paths in \mathcal{D} can be possible critical path realizations.

In contrast, if $c_{13} < c_{12}c_{23}$ we have $\mathbb{P}(\bar{c}_{13} > \bar{c}_{12}\bar{c}_{23}) = \mathbb{P}(\varepsilon_2 < c_{13}/(c_{12}c_{23})) = 0$. In this case, the path $[1 \rightarrow 3]$ can be random critical only on a null set and therefore $[1 \rightarrow 3]$ is not a possible critical path realization.

This illustrates that a path p from j to i with path weight $d_{ji}(p) < b_{ji}$ may as well contribute to the distribution of U_i . However, an edge $p = [j \rightarrow i]$ with $d_{ji}(p) < b_{ji}$ is still not identifiable and does not change the distribution of \mathbf{U} .

Recall from (2.6) and (3.5) that

$$\mathbf{X} = \mathbf{Z} \odot \mathbf{B} \quad \text{and} \quad \mathbf{U} = \mathbf{Z} \odot \bar{\mathbf{B}}.$$

We present some useful properties of \mathbf{B} and $\bar{\mathbf{B}}$ providing a link between the noisy and non-noisy model as defined in (2.1) and (3.1), respectively. Such properties have been shown for \mathbf{B} in [13, 14, 15], and we investigate here which of them remain valid for $\bar{\mathbf{B}}$.

Lemma 3.8. Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1) with \mathbf{B} and $\bar{\mathbf{B}}$ defined in (2.6) and (3.5), respectively. Then the following assertions hold:

- a) $\bar{b}_{ji} = \bigvee_{k \in V} \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}} \geq \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$, where the inequality is strict, whenever the random critical path from j to i is the edge $j \rightarrow i$, or $j = i$.
- b) There exists some path $p := [j \rightarrow \dots \rightarrow k \rightarrow \dots \rightarrow i]$ from j to i that passes through k such that

$$\bar{d}_{ji}(p) = \bar{b}_{ji} \quad \text{if and only if} \quad \bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}.$$

- c) $\frac{U_i}{U_j} \geq \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \geq b_{ji}$ with $b_{ji} = 0$ for $j \notin \text{An}(i)$

$$d) \text{ supp}(U_i/U_j) = \begin{cases} [b_{ji}, \infty) & \text{for } j \in \text{an}(i), \\ [0, 1/b_{ij}] & \text{for } i \in \text{an}(j), \\ \{1\} & \text{for } i = j, \\ \mathbb{R}_+ & \text{otherwise.} \end{cases}$$

Moreover, for $j \neq i$, neither the distribution of U_i/U_j nor the distribution of U_j/U_i have any atoms.

$$e) \text{ If } b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}, \text{ then } \bar{b}_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}.$$

$$f) \text{ If } b_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \text{ and } \text{de}(j) \cap \text{an}(i) \neq \emptyset, \text{ then}$$

$$\mathbb{P}\left(\bar{b}_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}\right) > 0 \quad \text{and} \quad \mathbb{P}\left(\bar{b}_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}\right) > 0.$$

Definition 5.1 of [14] presents the smallest subgraph of \mathcal{D} such that \mathbf{X} is a recursive ML model on this DAG as

$$\mathcal{D}^B = (V, E) := \left(V, \left\{ (j, i) \in E : b_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \right\} \right).$$

This means that \mathcal{D}^B contains an edge $j \rightarrow i$ of \mathcal{D} if and only if this edge is the only critical path from $j \rightarrow i$ in \mathcal{D} .

Lemma 3.8 b) and f) motivate the following definition as the random analog of \mathcal{D}^B .

Definition 3.9. Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on the DAG $\mathcal{D} = (V, E)$ as defined in (3.1). Then we define the minimum ML DAG $\bar{\mathcal{D}}^B$ as

$$\bar{\mathcal{D}}^B = (V, \bar{E}) := \left(V, \left\{ (j, i) \in E : \mathbb{P}\left(\bar{b}_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{pa}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}\right) > 0 \right\} \right).$$

This means that $\bar{\mathcal{D}}^B$ contains an edge $j \rightarrow i$ of \mathcal{D} if and only if this edge is a possible critical path realization from $j \rightarrow i$ in \mathcal{D} .

In addition, applying first Lemma 3.8 e) and f), and in the second part Lemma 3.8 b) yields the following result.

Corollary 3.10. Let $\mathbf{X} \in \mathbb{R}_+^d$ be a recursive ML vector on a DAG $\mathcal{D} = (V, E)$ as defined in (2.1) and $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise as defined in (3.1) on the same DAG \mathcal{D} with the same edge weight matrix \mathbf{C} . Then

$$\mathcal{D}^B = \bar{\mathcal{D}}^B,$$

which is the smallest DAG that preserves the distributions of \mathbf{X} and of \mathbf{U} .

We will henceforth only use the term \mathcal{D}^B .

The next lemma summarizes properties of possible critical path realizations from Definition 3.5(e).

Lemma 3.11. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1). Then the following assertions hold:*

- a) *A path $p = [j = k_0 \rightarrow \dots \rightarrow k_n = i]$ in \mathcal{D} is a possible critical path realization from j to i if and only if all edges of p belong to the minimum ML DAG \mathcal{D}^B .*
- b) *Let p_1 and p_2 be two possible critical path realizations from j to i and from l to m , respectively. Then*

$$\left\{ U_i = U_j d_{ji}(p_1) \prod_{k \in S_{p_1}} \varepsilon_k, U_m = U_l d_{lm}(p_2) \prod_{k \in S_{p_2}} \varepsilon_k \right\} \quad (3.9)$$

has positive probability if and only if $S_{p_1} \cap S_{p_2} = \emptyset$, or for every $r \in S_{p_1} \cap S_{p_2}$ the sub-path of p_1 from j to r is a sub-path of p_2 or the sub-path of p_2 from l to r is a sub-path of p_1 .

We illustrate part b) with Figure 1 and Figure 2.

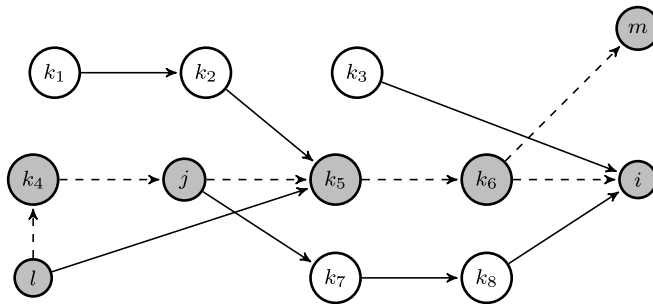


FIG 1. Both dashed paths $p_1 := [j \rightarrow k_5 \rightarrow k_6 \rightarrow i]$ and $p_2 := [l \rightarrow k_4 \rightarrow j \rightarrow k_5 \rightarrow k_6 \rightarrow m]$ can be possible critical path realizations from the same realized noise variables along the nodes.

3.3. Distributions of component ratios of the noisy model

The next result is important as it not only helps us to understand the model better, but is also an important step for learning the model.

Theorem 3.12. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1). Suppose that $p_{\max} := [j = k_0 \rightarrow \dots \rightarrow k_n = i]$*

is generic. Let $S_{p_{\max}} = \{k_1, \dots, k_n\}$ be the set of nodes on p_{\max} . Then

$$\mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x\right) \sim \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x, \frac{U_i}{U_j} = b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k\right) \sim c \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right), \quad x \downarrow 1,$$

for some constant $c \in (0, 1)$.

Remark 3.13. If the distributions of the noise variables and the innovations as well as the path weights of the underlying DAG \mathcal{D} are given, the constant c in Theorem 3.12 can be calculated explicitly.

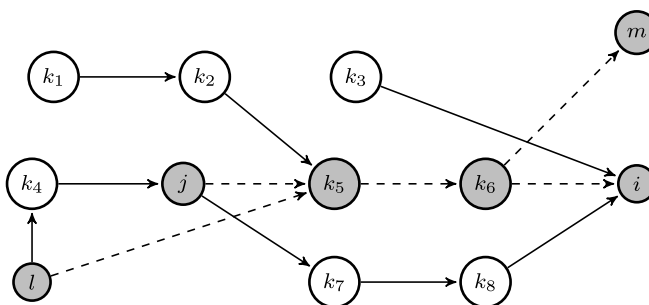


FIG 2. Both dashed paths $p_1 := [j \rightarrow k_5 \rightarrow k_6 \rightarrow i]$ and $p_2 := [l \rightarrow k_5 \rightarrow k_6 \rightarrow m]$ can only on a null-set be possible critical path realizations from the same realized noise variables along the nodes.

Theorem 3.12 also shows that, while a path p from j to i with $d_{ji}(p) < b_{ji}$ can contribute to the distribution of U (as we have seen in Example 3.7), they influence the distribution of U_i/U_j at their left limit of support only by the constant $c \in (0, 1)$.

We now extend the result to situations with several critical paths.

Corollary 3.14. Let U be as in Theorem 3.12. Suppose that exactly the paths p_1, \dots, p_n from j to i are critical; i.e., $d_{ji}(p_1) = \dots = d_{ji}(p_n) = b_{ji}$. Then

$$\mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x\right) \sim c \mathbb{P}\left(\bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\}\right), \quad x \downarrow 1,$$

for some constant $c \in (0, 1)$.

For simplicity, we assume from now on that \mathcal{C} is generic in the sense of Definition 3.5. However, we want to remark that all such results can be extended to the case of several non-random critical paths between two nodes. The proofs of such results work similarly as the proof of Corollary 3.14.

We continue with another consequence of Theorem 3.12.

Corollary 3.15. *Let \mathbf{U} be as in Theorem 3.12 and suppose that $p := [j = k_0 \rightarrow \dots \rightarrow k_n = i]$ is generic. Let $\mathbf{U}^1, \dots, \mathbf{U}^n$ for $n \in \mathbb{N}$ be an iid sample from \mathbf{U} . Then, for the same constant $c \in (0, 1)$ as in Theorem 3.12, we have*

$$\mathbb{P}\left(\bigwedge_{k=0}^n \frac{U_i^k}{U_j^k} \leq b_{ji}x\right) \sim cn \mathbb{P}\left(\prod_{i=1}^n \varepsilon_{k_i} \leq x\right), \quad x \downarrow 1.$$

We conclude this section by extending Theorem 3.12 to multivariate distributions, which is an important structural result of the new model. We only formulate and prove the bivariate case, the general case is then obvious. Recall that in Lemma 3.11 we gave a necessary and sufficient condition for (3.10) below.

Theorem 3.16. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1). Suppose generic paths p_1 from j to i and p_2 from l to m . Assume that*

$$\mathbb{P}\left(U_i = U_j b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, U_m = U_l b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k\right) > 0. \tag{3.10}$$

Then

$$\begin{aligned} &\mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x_1, \frac{U_m}{U_l} \leq b_{lm}x_2\right) \sim c \mathbb{P}\left(\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1, \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2\right) \\ &\sim \mathbb{P}\left(\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1, \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2, \frac{U_i}{U_j} = b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, \frac{U_m}{U_l} = b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k\right), \end{aligned}$$

for $x_1, x_2 \downarrow 1$ and some constant $c \in (0, 1)$.

4. Identification and estimation

We first address the question of identifiability of \mathbf{B} from the distribution of \mathbf{U} . In particular, we are going to show that even though innovations and noise variables are generally not identifiable, \mathbf{B} remains identifiable also in the propagating noise model.

We discuss three settings (1)-(3) below. For each setting, we propose an appropriate minimum ratio estimator for \mathbf{B} . Afterwards, we will show the almost sure convergence of each of the estimators.

4.1. Identifiability of the model

In this section we discuss the question of identifiability of the DAG \mathcal{D} and the edge weights \mathbf{C} of a ML model with recursive noise from the distribution of \mathbf{U} . As we have already seen in Example 3.7, the true DAG \mathcal{D} and the edge weight matrix \mathbf{C} underlying \mathbf{U} in representation (3.1) are generally not identifiable

from the distribution of \mathbf{U} . The smallest DAG with a chance to be identified from the distribution of \mathbf{U} is the minimum ML DAG $\bar{\mathcal{D}}^B$ of Definition 3.9, which in turn can be identified from \mathbf{B} by Corollary 3.10.

By the equivalence of \mathcal{D}^B and $\bar{\mathcal{D}}^B$, Theorem 2 in [15] also holds for the propagating noise model defined in (3.1), i.e., the theorem defines the class of DAGs that preserve the distribution of \mathbf{U} . As by Lemma 3.8 d) the ML coefficients are limits of supports of component ratios of \mathbf{U} , the following is immediate.

Corollary 4.1. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML model with propagating noise on a DAG \mathcal{D} as defined in (3.1). Then the ML coefficient matrix \mathbf{B} is identifiable from the distribution of \mathbf{U} .*

Since we can identify \mathbf{B} from the distribution of \mathbf{U} , we can also identify the minimum ML DAG \mathcal{D}^B from Definition 2.1 (which by Definition 3.9 and Corollary 3.10 is the minimum DAG preserving the distribution of \mathbf{U}). Therefore, since $\varepsilon \geq 1$, Theorem 2 of [15] also holds for the propagating noise model as defined in (3.1). Therefore, as exemplified in Example 3.7, we can identify the class of all DAGs and edge weights that could have generated \mathbf{U} .

However, unlike for the non-noisy model, we can generally not identify innovations or noise variables. To see this assume a source node U_i in a DAG \mathcal{D} such that $\text{an}(i) = \emptyset$. If \mathbf{U} follows a recursive ML model with propagating noise, then $U_i := Z_i \varepsilon_i$. In particular, we can not identify Z_i or ε_i .

When estimating a recursive ML model with propagating noise, we distinguish between three settings. We point out that all algorithms below work equally, if the data is generated from a noise-free max-linear model as defined in (2.1).

- (1) All ancestral relations are known; i.e., we know the set of edges E , hence the DAG. This might be the case when modeling networks that contain natural information about edges. The problem then reduces to finding appropriate estimates \hat{b}_{ji} for $j \in \text{an}(i)$.
- (2) The ancestral relations are unknown; however, we know a topological order of the nodes. Then, in contrast to setting 1, we need to decide if a path from j to i with $j < i$ exists.
- (3) Neither the underlying DAG nor a topological order of the nodes is known. Then we need to find a topological order of the nodes and proceed then as in setting 2.

We next want to estimate \mathbf{B} for each of the three settings (1)-(3).

4.2. Known DAG structure with unknown edge weights

Given an iid sample $\mathbf{U}^1, \dots, \mathbf{U}^n$ from a recursive ML model with propagating noise on a known DAG \mathcal{D} as defined in (3.1) and knowing all ancestral relations

of \mathcal{D} , we could choose the simple estimate

$$\tilde{\mathbf{B}} := (\tilde{b}_{ij})_{d \times d} = \left(\bigwedge_{k=1}^n \frac{U_j^k}{U_i^k} \mathbf{1}_{\text{An}(j)}(i) \right)_{d \times d}. \tag{4.1}$$

However, as in the non-noisy model, the estimate (4.1) may not define any recursive ML model on the given DAG \mathcal{D} , cf. Example 3 of [15].

We use instead

$$\mathbf{B}_0 = (B_0(i, j))_{d \times d} := \left(\bigwedge_{k=1}^n \frac{U_j^k}{U_i^k} \mathbf{1}_{\text{pa}(j)}(i) \right)_{d \times d} \quad \text{and set} \quad \hat{\mathbf{B}} = (\mathbf{I}_d \vee \mathbf{B}_0)^{\odot(d-1)}. \tag{4.2}$$

Applying Lemma 2 in [15] to \mathbf{B}_0 , the estimator $\hat{\mathbf{B}}$ yields a valid estimate of the given DAG in the sense that $\hat{\mathbf{B}}$ defines a recursive ML model and for any pair $(j, i) \notin E(\mathcal{D})$ we have $\hat{b}_{ji} = \bigvee_{k \in \{1, \dots, d\} \setminus \{j, i\}} \hat{b}_{jk} \hat{b}_{ki}$. Moreover, by the idempotency of $\hat{\mathbf{B}}$ and Lemma 3.8 c), similarly to the non-noisy model, it also holds that

$$b_{ji} \leq \hat{b}_{ji} \leq \check{b}_{ji}, \quad j \in \text{an}(i). \tag{4.3}$$

4.3. Known topological order

Given an iid sample $U^1, \dots, U^n \in \mathbb{R}_+^d$ from a recursive ML model with propagating noise without knowing \mathcal{D} , but knowing the topological order of nodes, we adapt the estimator (4.1) to this situation and define

$$\hat{\mathbf{B}} := (\hat{b}_{ij})_{d \times d} = \left(\bigwedge_{k=1}^n \frac{U_j^k}{U_i^k} \mathbf{1}_{(i < j)} \right)_{d \times d}. \tag{4.4}$$

4.4. Unknown DAG and unknown topological order

Given an iid sample $U^1, \dots, U^n \in \mathbb{R}_+^d$ from a recursive ML model with propagating noise without knowing \mathcal{D} or the topological order, we will recover a topological order first and then proceed as in Section 4.3.

Estimating the topological order of an underlying DAG is often done by learning algorithms that successively identify source nodes and succeeding generations. For additive models, usually regression techniques are applied (see e.g. [7] or [32]). In the recursive ML model, the noise is not additive and the model is highly non-linear. Hence, such regression methods cannot be applied. However, under the condition of multivariate regular variation, the paper [23] suggests a learning algorithm for the model without noise as given in (1.1). We propose a different approach, which to the best of our knowledge has not been considered in the literature before. It applies to the propagating noise model without any

distributional assumptions on the innovations and noise variables and learns the DAG by using minimum ratios. We first consider the matrix of all minimum ratios given by

$$\check{\mathbf{B}} := (\check{b}_{ij})_{d \times d} = \left(\bigwedge_{k=1}^n \frac{U_j^k}{U_i^k} \right)_{d \times d}. \quad (4.5)$$

Let Π denote the set of all topological orders of V . Furthermore, denote an equivalence class of topological orders induced by the underlying (unknown) DAG $\mathcal{D} = (V, E)$ by

$$R_{\mathcal{D}} := \{\pi \in \Pi : \pi(j) < \pi(i) \text{ for all } (j, i) \in E\}. \quad (4.6)$$

By Lemma 3.8 d), \check{b}_{ji} is lower bounded by b_{ji} for $j \in \text{an}(i)$ and $\check{b}_{ji} \rightarrow 0$ a.s. as $n \rightarrow \infty$ for $j \notin \text{An}(i)$. This is a direct result from Lemma 3.8 c) and the fact that the minimum is non-increasing. Hence, for any $\pi \in R_{\mathcal{D}}$ it holds that $\check{b}_{ji} \rightarrow 0$ a.s. as $n \rightarrow \infty$ whenever $\pi(j) > \pi(i)$. Therefore, also

$$\max_{\substack{(j,i) \in V \times V: \\ \pi(j) > \pi(i)}} \check{b}_{ji} \rightarrow 0 \quad \text{a.s. for } n \rightarrow \infty. \quad (4.7)$$

In contrast, for any $\pi \notin R_{\mathcal{D}}$, there is a pair of nodes (j, i) such that $b_{ji} > 0$ although $\pi(j) > \pi(i)$. For this reason,

$$\max_{\substack{(j,i) \in V \times V: \\ \pi(j) > \pi(i)}} \check{b}_{ji} \rightarrow c_{\pi} > 0 \quad \text{a.s. for } n \rightarrow \infty. \quad (4.8)$$

As a consequence, for a given topological order π , by (4.7) and (4.8), the maximum converges almost surely to zero if and only if $\pi \in R_{\mathcal{D}}$. Hence we propose a topological order that minimizes this expression, i.e.,

$$\arg \min_{\pi \in \Pi} \max_{\substack{(j,i) \in V \times V: \\ \pi(j) > \pi(i)}} \check{b}_{ji}. \quad (4.9)$$

A topological order found by (4.9) generally is not unique. Algorithm 1 returns a unique topological order for any fixed estimated matrix $\check{\mathbf{B}}$.

Proposition 4.2. *Algorithm 1 solves the optimization problem in equation (4.9).*

Proof. Let π be the topological order from Algorithm 1 and denote by $S(\pi)$ the objective function in (4.9). Algorithm 1 sorts all pairs (j, i) by the size of \check{b}_{ji} and draws an edge from j to i whenever there is no path from i to j . Now consider the first pair (j, i) in the algorithm where we do not draw an edge since $i \in \text{an}(j)$. Consider a permutation π' that is obtained by exchanging the order of nodes i and j in π . Since there exists already a path from i to j , it follows that $S(\pi) \leq S(\pi')$. \square

Algorithm 1 Estimating a topological order**Input:** A matrix of minimum ratios $\tilde{\mathbf{B}}$ as in (4.5)**Output:** An estimated topological order $\hat{\pi}$

- 1: Set $\tilde{\mathcal{D}} = (V, E)$ with $V = \{1, \dots, d\}$ and $E = \emptyset$.
- 2: Set $S := \{(j, i) \in V \times V : j \neq i\}$ and sort the elements (j, i) of S by the size of \tilde{b}_{ji} from large to small.
- 3: **for** (j, i) in S **do**
- 4: **if** $i \notin \text{an}(j)$ in $\tilde{\mathcal{D}}$ **then**
- 5: $E = E \cup (j, i)$
- 6: **end if**
- 7: **end for**
- 8: **return** the topological order $\hat{\pi}$ of the DAG $\tilde{\mathcal{D}}$

The DAG $\tilde{\mathcal{D}}$ constructed in Algorithm 1 works as an auxiliary instrument to infer a topological order. Observe that $\tilde{\mathcal{D}}$ is a complete DAG, with directed edges between every node pair in V and, hence, there is a unique topological order representing $\tilde{\mathcal{D}}$. Moreover, since we sort the weights by size, the algorithm solves (4.9) in an optimal way for given $\tilde{\mathbf{B}}$. At first sight the algorithm bears some similarity to Kruskal's classical algorithm for finding a minimum spanning tree; see [26]. However, Algorithm 1 works with directed edges and, of course, the optimization problem itself is very different.

Adding an edge and checking the presence of a path between any pair of nodes both can be implemented in $O(d)$ amortized complexity (see [18]). Hence, since S as computed in line 2 of Algorithm 1 contains $d(d-1)$ pairs of nodes, we have an overall amortized complexity of $O(d^3)$. After Algorithm 1 we can again use the minimum ratio estimator

$$\hat{\mathbf{B}} := (\hat{b}_{ij})_{d \times d} = \left(\bigwedge_{k=1}^n \frac{U_j^k}{U_i^k} \mathbf{1}_{(\hat{\pi}(i) < \hat{\pi}(j))} \right)_{d \times d}. \quad (4.10)$$

4.5. Strong consistence of $\hat{\mathbf{B}}$ and learning the minimum ML DAG $\mathcal{D}^{\mathbf{B}}$

We first want to formally state the a.s. convergence of the proposed estimators for the ML coefficient matrix \mathbf{B} . Afterwards, we discuss how to learn the minimum ML DAG $\mathcal{D}^{\mathbf{B}}$. The proofs of Proposition 4.3 and Lemma 4.4 can be found in Appendix B.

Proposition 4.3. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise as defined in (3.1) and let $\mathbf{U}^1, \dots, \mathbf{U}^n \in \mathbb{R}_+^d$ be an iid sample from \mathbf{U} . Then the estimates (4.2), (4.4) and (4.10) of \mathbf{B} are strongly consistent, i.e., it holds a.s. for $n \rightarrow \infty$ that*

$$\hat{b}_{ji} \longrightarrow b_{ji} \quad \text{for } j \in \text{an}(i), \quad \hat{b}_{ii} = 1, \quad \text{and} \quad \hat{b}_{ji} \longrightarrow 0 \quad \text{for } j \in V \setminus \text{An}(i).$$

In Sections 4.2-4.4 we have been discussing how to estimate \mathbf{B} under the settings (1)-(3). However, as we know from Corollary 3.10, only critical edges

of \mathcal{D} contribute to the distribution of \mathbf{U} . Asymptotically, we can almost surely identify \mathcal{D}^B since there is an edge $j \rightarrow i$ in \mathcal{D}^B if and only if $b_{ji} > b_{jl}b_{li}$ for all $l \in \text{de}(j) \cap \text{an}(i)$.

However, in real life we estimate the edges of \mathcal{D}^B for a finite data set. Since $\bigwedge_{k=1}^n (U_i^k/U_j^k) > 0$ holds for all $n \in \mathbb{N}$ and all $i, j \in V$, the estimators (4.4) or (4.10) would result in a matrix representing a complete DAG.

Since small estimated values \hat{b}_{ji} may well be 0 in the true model, we use a threshold $\delta_1 > 0$ with the aim to set an estimator $\hat{b}_{ji} < \delta_1$ equal to 0. However, setting single values $\hat{b}_{ji} := 0$ may destroy the idempotency of $\hat{\mathbf{B}}$ since idempotency requires for any triple of nodes (j, l, i) ,

$$\hat{b}_{ji}\hat{b}_{li} = \bigwedge_{k=1}^n \frac{U_l^k}{U_j^k} \bigwedge_{k=1}^n \frac{U_i^k}{U_l^k} \leq \bigwedge_{k=1}^n \frac{U_l^k}{U_j^k} \frac{U_i^k}{U_l^k} = \hat{b}_{ji}. \quad (4.11)$$

For the estimates however, it might be possible that $\hat{b}_{ji} < \delta_1$, while $\hat{b}_{jl} > \delta_1$ and $\hat{b}_{li} > \delta_1$. In this case, setting $\hat{b}_{ji} = 0$ would result in $\hat{b}_{ji} < \hat{b}_{jl}\hat{b}_{li}$ violating (4.11). To preserve the idempotency of $\hat{\mathbf{B}}$ while setting some small values to 0, we propose a simple adapted thresholding algorithm.

Algorithm 2 Thresholding while maintaining idempotency

Input: A (known or estimated) topological order $\pi : 1, \dots, d$ and an idempotent estimate $\hat{\mathbf{B}}$ as in (4.4) or (4.10) and a threshold value $\delta_1 > 0$

Output: An idempotent estimate $\hat{\mathbf{B}}$

```

1:  $E := \{(j, i) \in V \times V : \text{sgn}(\hat{b}_{ji}) = 1 \text{ and } i \neq j\}$ 
2:  $\mathcal{D} := (V, E)$ 
3:  $S := \{(j, i) \in E : 0 < \hat{b}_{ji} < \delta_1\}$ 
4: Sort the pairs  $(j, i)$  in  $S$  by the distance  $i - j$  from low to high
5: for  $(j, i)$  in  $S$  do
6:   if  $(j - i) == 1$  then
7:      $\hat{b}_{ji} = 0$ 
8:   end if
9:   if for every  $l$  with  $j < l < i$ :  $(j, l)$  or  $(l, i) \in S$  then
10:     $\hat{b}_{ji} = 0$ 
11:   else
12:     $S = S \setminus \{(j, i)\}$ 
13:   end if
14: end for
15: return  $\hat{\mathbf{B}}$ 

```

Lemma 4.4. *Algorithm 2 with threshold $\delta_1 > 0$ outputs an idempotent matrix, i.e., $\hat{\mathbf{B}} \odot \hat{\mathbf{B}} = \hat{\mathbf{B}}$ and there is no other idempotent matrix \mathbf{B}' such that $b'_{ji} = \hat{b}_{ji}$ whenever $\hat{b}_{ji} > \delta_1$ that contains more zero entries than $\hat{\mathbf{B}}$.*

Remark 4.5. *If we choose $\delta_1 \leq \min\{\hat{b}_{ji} : j < i\}$ no entry is set to 0, and if $\delta_1 > \max\{\hat{b}_{ji} : j < i\}$ all entries are set to 0 except for the diagonal. So in the first case, we obtain the complete DAG and in the second case the DAG consists of isolated nodes only.*

In order to estimate the minimum ML DAG \mathcal{D}^B it is not sufficient to decide if a path from j to i exists, i.e., if $b_{ji} > 0$. We need in particular to decide if the edge $j \rightarrow i$ belongs to \mathcal{D}^B . By continuity of the noise variables we may observe for the estimated path weights

$$\hat{b}_{ji} > \hat{b}_{jl}\hat{b}_{li}$$

even if $b_{ji} = b_{jl}b_{li}$. However, by Proposition 4.3, in this situation the difference $(\hat{b}_{ji} - \hat{b}_{jl}\hat{b}_{li}) \rightarrow 0$ a.s. as $n \rightarrow \infty$. Therefore, we introduce another threshold $\delta_2 > 0$ enforcing an edge in \mathcal{D}^B if this difference is greater than δ_2 . In Theorem 3.12 we have seen that the distribution of the ratio $\mathbb{P}(U_i/U_j \leq b_{ji}x)$ is asymptotically determined by $\mathbb{P}(\prod_{k \in S_p} \varepsilon_k - 1 \leq x)$ for $x \downarrow 0$. Hence, the rate of convergence of $(\hat{b}_{ji} - \hat{b}_{jl}\hat{b}_{li})$ depends crucially on the path length $m = |S_p|$. Ideally, we therefore choose $\delta_2 = \delta_2(n, m)$ depending not only on the sample size n , but also on the path length m .

More precisely, since $F_{\sum_{k \in S_p} \varepsilon_k}^{\leftarrow}(1/n) \sim F_{\prod_{k \in S_p} \varepsilon_{k-1}}^{\leftarrow}(1/n)$ (see Theorem 5.2 and its proof below), and assuming that \mathbf{C} is generic, we find that Algorithm 3 asymptotically identifies \mathcal{D}^B , if

$$F_{\sum_{k \in S_p} \varepsilon_k}^{\leftarrow}(1/n) = o(\delta_2(n, m)) \quad \text{for } n \rightarrow \infty.$$

In real life we do not know the number of critical edges in either of the three settings. We distinguish between setting (1) and settings (2)-(3) and propose Algorithm 3 with $\delta_2(m) := \delta_2(n, m)$, i.e., for a fixed sample size n we focus on the path length m . For setting (1) we do know the underlying unweighted DAG \mathcal{D} . Therefore, we do not need to decide whether some small value \hat{b}_{ji} corresponds to a path from j to i . However, we do not know the minimum ML DAG \mathcal{D}^B such that we would apply Algorithm 3 to estimate \mathcal{D}^B . For settings (2) and (3) we would apply first Algorithm 2 and afterwards Algorithm 3.

To further illustrate this, observe the diagram below. In setting 3, we start with $\tilde{\mathbf{B}}$, while in setting 2 with $\hat{\mathbf{B}}$ and for setting 1, we start with $\hat{\mathbf{B}}$.

$$\tilde{\mathbf{B}} \xrightarrow{\text{Alg. 1}} \hat{\pi} \xrightarrow{(4.10)} \hat{\mathbf{B}}; (\hat{\pi}, \hat{\mathbf{B}}) \xrightarrow{\text{Alg. 2}} \tilde{\mathbf{B}}; (\hat{\pi}, \tilde{\mathbf{B}}) \xrightarrow{\text{Alg. 3}} D^{\tilde{\mathbf{B}}}$$

In the next section we derive the asymptotic distribution of the estimators.

5. Asymptotic distribution of the minimum ratio estimators

With the goal of proving asymptotic distributional properties of the minimum ratio estimators for the different settings (1)-(3), we require regular variation of the noise variable ε in its left endpoint. Under this condition we first prove that also the minimum ratio estimators $\bigwedge_{k=1}^n (U_i^k/U_j^k)$ are regularly varying. Moreover, we show that their joint limit distribution is the product of Weibull distributions. In this section we assume \mathbf{C} is generic in the sense of Definition 3.5.

Algorithm 3 Approximating max-weighted paths

Input: Threshold sequences $\delta_2(1), \dots, \delta_2(d)$ and settings

(1): a known underlying DAG $\mathcal{D} := (V, E)$ and an estimate \hat{B} as in (4.2), or

(2-3): a (known or estimated) topological order $\pi : 1, \dots, d$ and a thresholded matrix \hat{B} obtained from Algorithm 2.

Output: An estimated minimum DAG $\mathcal{D}^{\hat{B}} = (V, E^{\hat{B}})$

$E^{\hat{B}} := \emptyset$ and $\mathcal{D}^{\hat{B}} := (V, E^{\hat{B}})$

(1): $S := \{(j, i) \in V \times V : j \in \text{pa}(i)\}$ and infer a topological order $\pi : 1, \dots, d$ from \mathcal{D}

(2)-(3): $S := \{(j, i) \in V \times V : j < i\}$

Sort pairs (j, i) in S by their distance $(i - j)$ according to the topological order from low to high

for (j, i) in S **do**

if \exists path p from j to i in $\mathcal{D}^{\hat{B}}$ **then**

 Set m as the maximum path length in $\mathcal{D}^{\hat{B}}$

 Set $l := \arg \max_{l \in V \setminus \{j, i\}} (\hat{b}_{jl} \hat{b}_{li})$

if $(\hat{b}_{ji} - \hat{b}_{jl} \hat{b}_{li}) > \delta_2(m)$ **then**

$E^{\hat{B}} := E^{\hat{B}} \cup \{(j, i)\}$

end if

else

if $\hat{b}_{ji} > 0$ **then**

$E^{\hat{B}} := E^{\hat{B}} \cup \{(j, i)\}$

end if

end if

end for

return $\mathcal{D}^{\hat{B}} = (V, E^{\hat{B}})$

The results can be extended to a non-generic model by similar methods as used in Corollary 3.14.

In what follows we assume that the random variables $\tilde{\varepsilon}_i := \ln(\varepsilon_i) > 0$ for $i = 1, \dots, d$ are iid regularly varying at zero with exponent $\alpha > 0$ and recall from Corollary 2.6(b) that this is equivalent to $(\varepsilon - 1) \in RV_\alpha^0$ or $\varepsilon \in RV_\alpha^1$.

We first prove that $\ln(U_i/U_j) - \ln(b_{ji})$ is regularly varying at zero which will be a consequence of Theorem 3.12. In this auxiliary result as well as in the theorems below we need that \mathbf{C} is generic. Further, for a path p we denote by $\zeta(p) = |S_p|$ its path length.

Lemma 5.1. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1) and assume that the path $p := [j \rightarrow \dots \rightarrow i]$ from j to i is generic. If $\ln(\varepsilon) \in RV_\alpha^0$, then $\ln(U_i/U_j) - \ln(b_{ji}) \in RV_{\zeta(p)\alpha}^0$.*

The following is the main result of this section and describes the asymptotic distribution of the minimum ratio estimator \hat{B} from (4.10). In particular, it shows that its entries are asymptotically independent.

Theorem 5.2. *Let $\mathbf{U} \in \mathbb{R}_+^d$ be a recursive ML vector with propagating noise as defined in (3.1). Assume that \mathbf{C} is generic and that $\tilde{\varepsilon} = \ln(\varepsilon) \in RV_\alpha^0$. For every path p_{ji} from j to i and node set $S_{p_{ji}}$ choose $a_n^{(ji)} \sim F_{\sum_{k \in S_{p_{ji}}} \tilde{\varepsilon}_k}^{\leftarrow} (1/n)$ as $n \rightarrow \infty$. If $\mathbf{U}^1, \dots, \mathbf{U}^n$ is an iid sample from \mathbf{U} , then*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{a_n^{(ji)} b_{ji}} \left(\bigwedge_{k=1}^n \frac{U_i^k}{U_j^k} - b_{ji} \right) \leq x_{ji} \forall (j, i) \in V \times V \text{ with } b_{ji} > 0 \right) \\ &= \prod_{\substack{(j,i) \in V \times V: \\ b_{ji} > 0}} \Psi_{\zeta_{(p_{ji})\alpha, (c^{(ji)})^{1/(\zeta_{(p_{ji})\alpha}})}(x_{ji}), \quad x_{ji} > 0, \end{aligned}$$

where $c^{(ji)} \in (0, 1)$ is defined as in Theorem 3.12.

If we know the minimum ML DAG $\mathcal{D}^B = (V, E(\mathcal{D}^B))$, it is preferable to estimate b_{ji} as in (4.2). Then Theorem 5.2 reduces as follows.

Corollary 5.3. *Let the assumptions of Theorem 5.2 hold and assume that the minimum ML DAG $\mathcal{D}^B(V, E(\mathcal{D}^B))$ is known. Then*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{a_n^{ji} b_{ji}} \left(\bigwedge_{k=1}^n \frac{U_i^k}{U_j^k} - b_{ji} \right) \leq x_{ji} \forall (j, i) \in E(\mathcal{D}^B) \right) \\ &= \prod_{(j,i) \in E(\mathcal{D}^B)} \Psi_{\alpha, (c^{(ji)})^{1/\alpha}}(x_{ji}), \quad x_{ji} > 0. \end{aligned}$$

6. Data analysis and simulation study

We want to apply the methods that we have developed over the past sections and consider a data example. For a quality assessment we also perform a simulation study.

6.1. Data example

We consider dietary supplement data of $n = 8327$ independent interviews taken from the NHANES report for the year 2015-2016, which is available at https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR1TOT_I.XPT. They have been part of a questionnaire as to “What We Eat in America”, which recorded the food and beverage consumed by all participants during the 24 hours period prior to the interview. The data contains 168 food components with the object of estimating the total intake of calories, macro and micro nutrients from foods and beverages. More details can be found on the website.

In [19], the data set has been considered in terms of an adapted k -means clustering algorithm for extremal observations. Moreover, assuming a recursive ML model and standardising the marginal data to regular variation at ∞ with $\alpha = 2$, [23] investigated the causal relationship between four nutrients using a different estimation method based on scalings.

In our data example we consider the same four nutrients, namely vitamin A (DR1TVARA), α -carotene (DR1TACAR), β -carotene (DR1TBCAR) and lutein+zeaxanthin (DR1TLZ) as in [23]. We abbreviate them by VA, AC, BC and LZ. In order to make results comparable to those of [23], we also use the empirical integral transform to standardize the data to Fréchet(2) margins (see e.g. [3], p. 381) by setting for $i = 1, 2, 3, 4$,

$$U_{li} := \left(-\log \left(\frac{1}{n+1} \sum_{j=1}^n \mathbf{1}_{\{\bar{U}_{ji} \leq \bar{U}_{li}\}} \right) \right)^{-1/2}, \quad l = 1, \dots, n = 8327,$$

where multiple ranks are uniformly randomly ordered.

We first consider the full matrix of minimum ratios $\check{\mathbf{B}} = (\check{b}_{ij})_{d \times d}$ with $\check{b}_{ij} = \bigwedge_{t=1}^n (X_j^t / X_i^t)$ given by

$$\begin{matrix} & \text{VA} & \text{AC} & \text{BC} & \text{LZ} & \\ \begin{pmatrix} 1 & 0.014 & 0.011 & 0.007 \\ 0.146 & 1 & 0.177 & 0.019 \\ 0.321 & 0.010 & 1 & 0.025 \\ 0.132 & 0.007 & 0.168 & 1 \end{pmatrix} & \text{VA} \\ & \text{AC} \\ & \text{BC} \\ & \text{LZ} \end{matrix}$$

We next apply Algorithm 1 to obtain an estimated topological order $\hat{\pi} := (AC, LZ, BC, VA)$. First we want to assess the quality of the estimated topological order $\hat{\pi}$, which also supports or contradicts the model assumption of a Bayesian network. Motivated by the coefficient R^2 of determination in regression we define the following.

Definition 6.1. For a given topological order π and an estimator \check{B} of the ML coefficient matrix we define the ML coefficient of determination

$$R_{\max}(\pi) = \frac{\sum_{\substack{(j,i) \in V \times V: \\ \pi(j) < \pi(i)}} \check{b}_{ji}}{\sum_{\substack{(j,i) \in V \times V: \\ j \neq i}} \check{b}_{ji}}.$$

The coefficient $R_{\max}(\pi)$ can take any value in the interval $[0, 1]$. Large $R_{\max}(\pi)$ supports the hypothesis that the underlying graph is a DAG and the estimated topological order lies in the equivalence class of topological orders defined in (4.6).

In our data example, we have $R_{\max}(\hat{\pi}) = 0.929$, strongly supporting the hypothesis of a recursive ML model. Now using the estimator (4.10), and applying Algorithms 2 and 3 with $\delta_1 = 0.02$ and $\delta_2(k) = 0.02$ for $k \in \{1, 2, 3\}$, we get the estimated minimum ML DAG $\mathcal{D}^{\hat{B}}$ and ML coefficient matrix \hat{B} , where we sorted the matrix according to $\hat{\pi}$. These are shown in Figure 3.

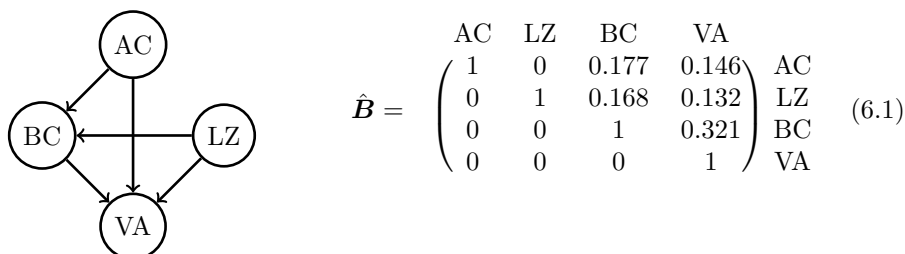


FIG 3. Estimated minimum ML DAG $\mathcal{D}^{\hat{B}}$ with estimated ML coefficient matrix \hat{B} .

Observe that, since we estimate the edge from AC to LZ to be absent, there are two possible topological orders.

From the estimates we observe that both, α -carotene and β -carotene lead to high amounts of vitamin A. This is in line with our expectation since β -carotene is a precursor to vitamin A and can be converted by β -carotene 15,15'-monooxygenase by many animals including humans. Similarly, also α -carotene can be converted to vitamin A. However, it is only half as active as β -carotene which explains that the edge weight from α -carotene to vitamin A is approximately half compared to the edge weight from β -carotene to vitamin A (0.146 compared to 0.321). Moreover, we can see that high amounts of lutein+zeaxanthin also lead to high amounts of β -carotene and high amounts of α -carotene also lead to high amounts of β -carotene. However, we did not find a significant connection between α -carotene and lutein+zeaxanthin. Observe that [23] inferred the same topological order, yet with one additional edge from α -carotene to lutein+zeaxanthin. However, it is also the edge with the smallest estimated edge weight. Similarly as in [23], we plot bivariate extremes in Figure 4 to underline our finding. The first 5 plots in Figure 4 look rather similar. For every large value of the substance on the vertical axis, we can see a large value of the substance on the horizontal axis. Moreover, these observations are shaped closely to a line. In contrast, a large value of the substance on the horizontal axis might as well coincide with a small value of the substance on the vertical axis. Therefore, e.g. a high amount of α -carotene leads to a high amount of vitamin A but a high amount of vitamin A does not necessarily lead to a high amount of α -carotene. This also supports that the dependence is not mutual and hence we can model it by a DAG. The same can be seen for any pair given in the plots 1-5. Moreover, since parts of the observations are shaped closely along a line, which we would expect for a recursive ML model, we can conclude that the recursive ML model fits the data very well.

The sixth plot is different from the other 5 plots, since for most large observations of α -carotene the level of lutein+zeaxanthin is not increased as most large observations in lutein+zeaxanthin do also not result in a high level of α -carotene. Therefore, the two substances do not seem to affect each other and we rightly concluded that there is no edge.

6.2. Simulation study

We want to illustrate the effect of observational noise in the ML model. We simulate recursive ML vectors with propagating noise, where the innovations Z_1, \dots, Z_4 are Fréchet(2) distributed and we use the estimated $\hat{\mathbf{B}}$ from (6.1) from the data analysis above for the ML coefficient matrix \mathbf{B} . Moreover, we simulate three different scenarios. In the first scenario, we assume the non-noisy model as given in (1.1), while for the second scenario we choose the propagating noise model with a medium sized noise and in the third setting we choose a noise variable which is stochastically larger. The scenarios are given as follows:

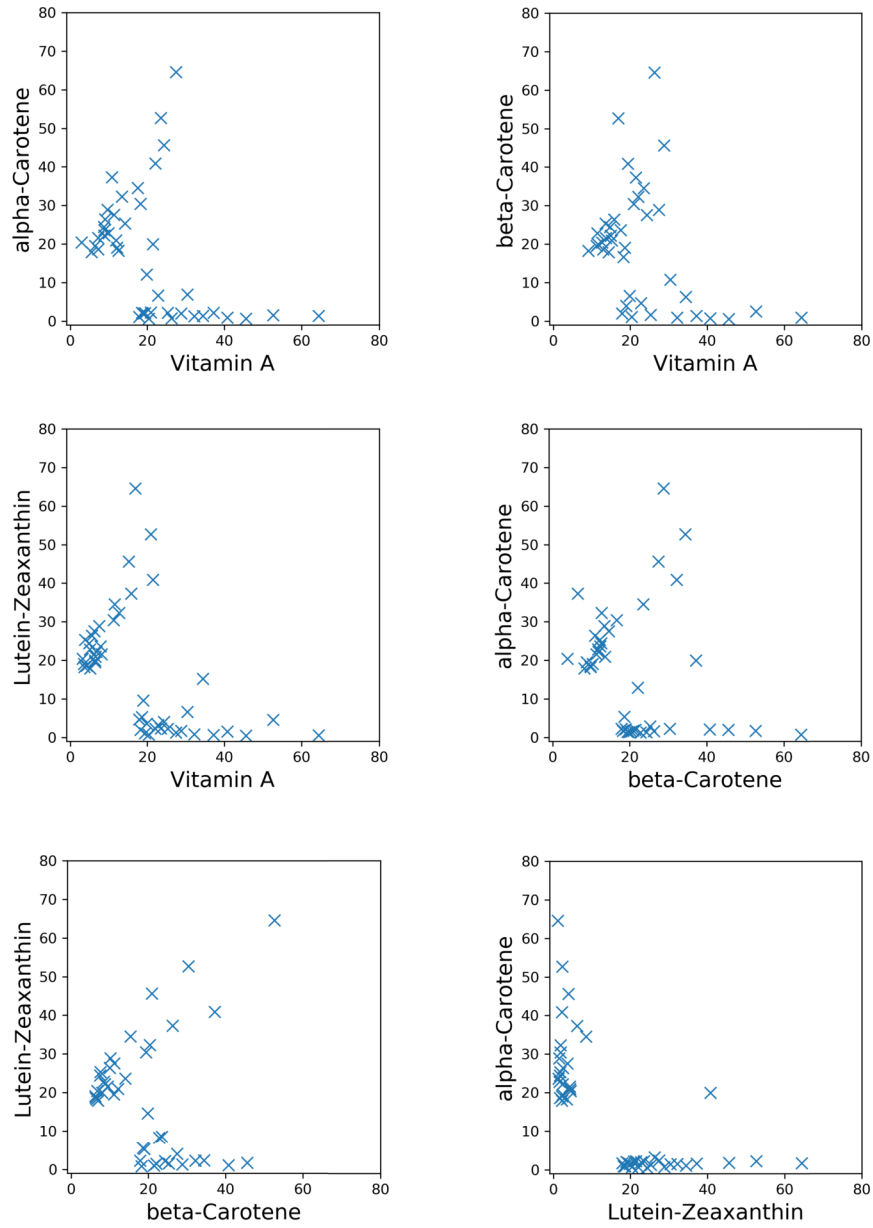


FIG 4. The empirical bivariate extremes (25 largest observations).

- (1) No noise
- (2) $\ln(\varepsilon_i) \sim \text{Gamma}(\lambda = 1, \alpha = 2)$ for $i \in \{1, 2, 3, 4\}$, which corresponds to $\mathbb{E}(\varepsilon_i) = 2$

TABLE 1

Empirical success probability for estimated topological order being in the equivalence class of topological orders for (1) No noise, (2): Gamma(1,2), (3); Gamma(2,2).

Sample Size	Correct Runs(1)	Correct Runs(2)	Correct Runs(3)
50	555	946	998
200	995	1000	1000
500	1000	1000	1000
1000	1000	1000	1000

(3) $\ln(\varepsilon_i) \sim \text{Gamma}(\lambda = 2, \alpha = 2)$ for $i \in \{1, 2, 3, 4\}$, which corresponds to $\mathbb{E}(\varepsilon_i) = 4$

We assume to have no information on the underlying DAG and we only consider the quality of the estimator \check{b}_{ji} given in (4.5). We choose the sample sizes $n \in \{50, 200, 500, 1000\}$ and 1000 simulation runs for each sample size. We first assess the success probabilities for Algorithm 1. Table 1 shows that the topological order can be correctly estimated even for small sample sizes. Moreover, the number of correct runs increases for larger noise variables. This is expected since the noise variables are one-sided. Therefore, for a path p from j to i the ratio $U_i/U_j \geq d_{ji}(p) \prod_{k \in S_p} \varepsilon_k$ increases, while the ratio $U_j/U_i \leq 1/(d_{ji}(p) \prod_{k \in S_p} \varepsilon_k)$ decreases. Therefore, it is easier to identify the paths in \mathcal{D} for larger noise.

Next, we want to assess the quality of the estimated ML coefficient matrix $\check{\mathbf{B}}$. To do so, for every pair (j, i) with $b_{ji} > 0$ and every simulation run $k \in \{1, \dots, 1000\}$, we denote the minimum ratio estimator given in (4.5) by \check{b}_{ji}^k .

We consider the empirical RMSE, standard deviation and bias for each $b_{ji} > 0$ in each model (1)-(3). In what follows we compare the three classical quantities

$$\text{bias}(\check{b}_{ji}) := \frac{1}{1000} \sum_{k=1}^{1000} \check{b}_{ji}^k - b_{ji}, \tag{6.2}$$

$$\text{SD}(\check{b}_{ji}) := \sqrt{\frac{1}{1000} \sum_{k=1}^{1000} (\check{b}_{ji}^k - \bar{\check{b}}_{ji})^2} \quad \text{with} \quad \bar{\check{b}}_{ji} = \frac{1}{1000} \sum_{k=1}^{1000} \check{b}_{ji}^k, \tag{6.3}$$

$$\text{RMSE}(\check{b}_{ji}) := \sqrt{\frac{1}{1000} \sum_{k=1}^{1000} (\check{b}_{ji}^k - b_{ji})^2}, \tag{6.4}$$

All three quantities are comparatively small even for small sample sizes and decrease whenever the sample size increases. Moreover, they are larger in the propagating noise model and larger noise terms also increase the three quantities. This is in line with what we can expect from the model as noise terms increase the ratios U_i/U_j and hence also increase the minimum ratio estimator. On the other hand, recall from above that with increasing noise the estimation of the DAG improves.

TABLE 2
Empirical Bias (6.2) for (1): No noise, (2): Gamma(1,2), (3): Gamma(2,2)

Sample Size	Edge	Edge Weight	Bias(1)	Bias(2)	Bias(3)
50	$AC \rightarrow BC$	0.177	0.012	0.020	0.046
50	$AC \rightarrow VA$	0.146	0.028	0.029	0.063
50	$LZ \rightarrow BC$	0.168	0.013	0.020	0.045
50	$LZ \rightarrow VA$	0.132	0.027	0.029	0.064
50	$BC \rightarrow VA$	0.321	0	0.014	0.053
200	$AC \rightarrow BC$	0.177	0	0.005	0.019
200	$AC \rightarrow VA$	0.146	0	0.007	0.026
200	$LZ \rightarrow BC$	0.168	0	0.005	0.020
200	$LZ \rightarrow VA$	0.132	0.001	0.007	0.026
200	$BC \rightarrow VA$	0.321	0	0.004	0.023
500	$AC \rightarrow BC$	0.177	0	0.002	0.012
500	$AC \rightarrow VA$	0.146	0	0.003	0.016
500	$LZ \rightarrow BC$	0.168	0	0.002	0.012
500	$LZ \rightarrow VA$	0.132	0	0.003	0.016
500	$BC \rightarrow VA$	0.321	0	0.001	0.015
1000	$AC \rightarrow BC$	0.177	0	0.001	0.008
1000	$AC \rightarrow VA$	0.146	0	0.001	0.011
1000	$LZ \rightarrow BC$	0.168	0	0.001	0.008
1000	$LZ \rightarrow VA$	0.132	0	0.001	0.010
1000	$BC \rightarrow VA$	0.321	0	0.001	0.010

TABLE 3
Empirical Standard Deviation (6.3) for (1): No noise, (2): Gamma(1,2), (3): Gamma(2,2)

Sample Size	Edge	Edge Weight	Std(1)	Std(2)	Std(3)
50	$AC \rightarrow BC$	0.177	0.031	0.023	0.031
50	$AC \rightarrow VA$	0.146	0.046	0.033	0.041
50	$LZ \rightarrow BC$	0.168	0.031	0.022	0.031
50	$LZ \rightarrow VA$	0.132	0.046	0.033	0.043
50	$BC \rightarrow VA$	0.321	0.006	0.015	0.031
200	$AC \rightarrow BC$	0.177	0.001	0.005	0.011
200	$AC \rightarrow VA$	0.146	0.002	0.006	0.015
200	$LZ \rightarrow BC$	0.168	0.002	0.005	0.012
200	$LZ \rightarrow VA$	0.132	0.005	0.008	0.017
200	$BC \rightarrow VA$	0.321	0	0.004	0.013
500	$AC \rightarrow BC$	0.177	0	0.002	0.007
500	$AC \rightarrow VA$	0.146	0	0.003	0.009
500	$LZ \rightarrow BC$	0.168	0	0.002	0.007
500	$LZ \rightarrow VA$	0.132	0.001	0.003	0.009
500	$BC \rightarrow VA$	0.321	0	0.001	0.008
1000	$AC \rightarrow BC$	0.177	0	0.001	0.004
1000	$AC \rightarrow VA$	0.146	0	0.001	0.006
1000	$LZ \rightarrow BC$	0.168	0	0.001	0.004
1000	$LZ \rightarrow VA$	0.132	0	0.001	0.006
1000	$BC \rightarrow VA$	0.321	0	0.001	0.005

Appendix A: Proofs of Section 3

PROOF OF THEOREM 3.2 Rewrite (3.1) in matrix form by means of the tropical matrix multiplication (2.4) as

$$U = (U \odot C \vee Z) \odot E_d.$$

TABLE 4
Empirical RMSE (6.4) for (1): No noise, (2): Gamma(1,2), (3): Gamma(2,2)

Sample Size	Edge	Edge Weight	RMSE(1)	RMSE(2)	RMSE(3)
50	AC → BC	0.177	0.033	0.030	0.056
50	AC → VA	0.146	0.054	0.044	0.075
50	LZ → BC	0.168	0.033	0.029	0.054
50	LZ → VA	0.132	0.053	0.044	0.077
50	BC → VA	0.321	0.006	0.021	0.061
200	AC → BC	0.177	0.001	0.007	0.023
200	AC → VA	0.146	0.002	0.009	0.030
200	LZ → BC	0.168	0.002	0.007	0.023
200	LZ → VA	0.132	0.006	0.011	0.031
200	BC → VA	0.321	0	0.005	0.027
500	AC → BC	0.177	0	0.003	0.014
500	AC → VA	0.146	0	0.004	0.018
500	LZ → BC	0.168	0	0.003	0.014
500	LZ → VA	0.132	0.001	0.004	0.018
500	BC → VA	0.321	0	0.002	0.017
1000	AC → BC	0.177	0	0.001	0.090
1000	AC → VA	0.146	0	0.002	0.120
1000	LZ → BC	0.168	0	0.001	0.090
1000	LZ → VA	0.132	0	0.002	0.012
1000	BC → VA	0.321	0	0.001	0.011

The associative law implies

$$U = (U \odot C \odot E_d) \vee (Z \odot E_d) \Leftrightarrow U = (U \odot \bar{C}) \vee \bar{Z}, \tag{A.1}$$

with $\bar{C} = C \odot E_d$, which is identical to (3.3), and $\bar{Z} = Z \odot E_d$. The right-most equation in (A.1) is of the same form as the non-noisy model in (2.5), so that analogously to its solution given in (2.6), we get the solution

$$B^* = (I_d \vee \bar{C})^{\odot(d-1)}, \quad U = \bar{Z} \odot B^* = Z \odot E_d \odot B^*,$$

where B^* is the Kleene star matrix of \bar{C} . Therefore, defining $\bar{B} = E_d \odot B^*$ yields the result. \square

PROOF OF COROLLARY 3.4 From (3.6) and the continuity of Z and ε we have

$$U_i = \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} Z_j = \bar{b}_{ki} Z_k \tag{A.2}$$

for some unique $k \in \text{An}(i)$. We want to show that this implies $U_i = \bar{b}_{ki} \tilde{U}_k$, i.e., $\tilde{U}_k = Z_k$. Applying first (3.7), then (3.6) and finally (3.4), we obtain

$$\tilde{U}_k = \frac{U_k}{\varepsilon_k} = \frac{\bigvee_{l \in \text{An}(k)} \bar{b}_{lk} Z_l}{\varepsilon_k} \geq \frac{\bar{b}_{kk} Z_k}{\varepsilon_k} = Z_k.$$

Now assume that $\tilde{U}_k > Z_k$. Then there exists an $l \in \text{an}(k) \subset \text{An}(i)$ with $\bar{b}_{lk} Z_l > \varepsilon_k Z_k$. Note also that the maximum random path weight from l to i

must be greater or equal than the maximum random path weight from l to i passing through node k . These two facts lead to

$$U_i = \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} Z_j \geq \bar{b}_{li} Z_l \geq \frac{\bar{b}_{lk} \bar{b}_{ki}}{\varepsilon_k} Z_l > \bar{b}_{ki} Z_k,$$

The above inequality, however, contradicts (A.2). Therefore, since $k \in \text{An}(i)$, we have

$$U_i \leq \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} \tilde{U}_j.$$

Now assume that $U_i < \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} \tilde{U}_j$. Then, with the same arguments as above, $U_i < \bigvee_{j \in \text{An}(i)} \bar{b}_{ji} Z_j$ which is a contradiction. \square

PROOF OF LEMMA 3.8 (a) We first assume that $j = i$. Since \mathcal{D} is a DAG, $\text{de}(i) \cap \text{pa}(i) = \emptyset$ and $\bar{b}_{ik} \bar{b}_{ki} = 0$ for all $k \neq i$. Therefore, the equality holds and the inequality is equivalent to $\bar{b}_{ii} \geq 0$ which obviously holds.

Next, assume $j \neq i$ and $j \notin \text{an}(i)$. Then by (3.4) $\bar{b}_{ji} = 0$ and there is no path from j to i . Therefore, $\text{de}(j) \cap \text{pa}(i) = \emptyset$. Hence, the right-hand side of the inequality equals zero. Moreover, the equality holds as well, otherwise $\bar{b}_{jk} > 0$ and $\bar{b}_{ki} > 0$ for some $k \in V$ and therefore, by (3.4) there would be a path from j to k and from k to i which contradicts $j \notin \text{an}(i)$.

For $j \in \text{pa}(i)$ with $\text{de}(j) \cap \text{pa}(i) = \emptyset$, the critical path must be the edge $j \rightarrow i$ since it is the only path from j to i . Furthermore, the equality $\bar{b}_{ji} = \frac{\bar{b}_{jk} \bar{b}_{ki}}{\bar{b}_{kk}}$ holds for $k = i$ and $k = j$ while for all $k \notin \{i, j\}$ it must hold that $\frac{\bar{b}_{jk} \bar{b}_{ki}}{\bar{b}_{kk}} = 0$. Therefore, the equality holds. Moreover, the right-hand side of the inequality again equals zero and we have strict inequality.

Now assume $j \in \text{an}(i)$ and $\text{de}(j) \cap \text{pa}(i) \neq \emptyset$. Then for every path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ with $n \geq 2$ from j to i and every $k_m \in \{k_1, \dots, k_{n-1}\}$, by (3.2),

$$\begin{aligned} \bar{d}_{ji}(p) &= \varepsilon_j \prod_{l=0}^{n-1} c_{k_l k_{l+1}} \varepsilon_{k_{l+1}} \\ &= \frac{\varepsilon_j \prod_{l=0}^{m-1} c_{k_l k_{l+1}} \varepsilon_{k_{l+1}} \cdot \varepsilon_{k_m} \prod_{l=m}^{n-1} c_{k_l k_{l+1}} \varepsilon_{k_{l+1}}}{\varepsilon_{k_m}} \\ &= \frac{\bar{d}_{jk_m}(p_1) \bar{d}_{k_m i}(p_2)}{\bar{b}_{k_m k_m}}, \end{aligned} \tag{A.3}$$

with $p_1 = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_m]$ and $p_2 = [k_m \rightarrow \dots \rightarrow k_n = i]$, where in the last step we have used that $\varepsilon_{k_m} = \bar{b}_{k_m k_m}$. Therefore, for the random critical path p with $\bar{b}_{ji} = \bar{d}_{ji}(p)$ it holds that every sub-path of this path is itself critical, otherwise we could find a path of larger random path weight by replacing the sub-path by a path of larger random weight. It follows that

$$\bar{b}_{ji} \geq \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{\bar{b}_{kk}}$$

with equality whenever the critical path p from j to i contains a node $k \in \text{de}(j) \cap \text{an}(i)$. Since for $k = i$ or $k = j$ we have $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$ and for $k \in V \setminus ((\text{an}(i) \cap \text{de}(j)) \cup \{j, i\})$ we have $\frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}} = 0$, the equality holds as well.

(b) First assume that there is a path $p := [j \rightarrow \dots \rightarrow k \rightarrow \dots \rightarrow i]$ with $\bar{d}_{ji}(p) = \bar{b}_{ji}$. Then by (A.3) we have $\bar{b}_{ji} = \frac{\bar{d}_{jk}(p_1)\bar{d}_{ki}(p_2)}{\bar{b}_{kk}}$. Now every sub-path of a random critical path must be itself critical, as explained in the proof of part a). Hence, $\bar{b}_{jk} = \bar{d}_{jk}(p_1)$ and $\bar{b}_{ki} = \bar{d}_{ki}(p_2)$ and for this reason $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$.

In contrast, let $\bar{d}_{ji}(p) < \bar{b}_{ji}$ for all $p \in P_{jki}$, where P_{jki} denotes all paths from j to i that pass through k . Now choose $p_1 = [j \rightarrow \dots \rightarrow k]$ and $p_2 = [k \rightarrow \dots \rightarrow i]$ such that $\bar{d}_{jk}(p_1) = \bar{b}_{jk}$ and $\bar{d}_{ki}(p_2) = \bar{b}_{ki}$. Then, for the path $p \in P_{jki}$ that results from concatenation of p_1 and p_2 we have by (3.4)

$$\bar{b}_{ji} > \bar{d}_{ji}(p) = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}},$$

which proves the reverse direction.

(c) For $j = i$ the inequality obviously holds, since $b_{ii} = 1$. If $j \notin \text{An}(i)$, then by definition $\bar{b}_{ji} = b_{ji} = 0$ and $\bar{b}_{jj} = \varepsilon_j \geq 1$. Therefore, the inequality is equivalent to $U_i/U_j \geq 0$, which is true. Now let $j \in \text{an}(i)$. Then by (3.2) and (3.4) the center ratio can be written as

$$\frac{\bar{b}_{ji}}{\bar{b}_{jj}} := \bigvee_{p \in P_{ji}} \frac{\bar{d}_{ji}(p)}{\bar{b}_{jj}} = \bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{l=0}^{n-1} \varepsilon_{k_{l+1}} \geq \bigvee_{p \in P_{ji}} d_{ji}(p) = b_{ji}, \quad (\text{A.4})$$

since $\bar{b}_{jj} = \varepsilon_j$ and $\varepsilon_i \geq 1$. Now we use (3.8) and obtain by (A.4)

$$\frac{U_i}{U_j} = \frac{\bigvee_{k \in \text{An}(i)} \bar{b}_{ki} \tilde{U}_k}{U_j} \geq \frac{\bar{b}_{ji} \tilde{U}_j}{U_j} = \frac{\bar{b}_{ji} U_j}{\varepsilon_j U_j} = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \geq b_{ji}.$$

(d) We first prove by contradiction that there is no lower bound for U_i/U_j of larger value than the one given in part (c). Assume $j \in \text{an}(i)$ and there is a lower bound $c > b_{ji}$. Since Z_1, \dots, Z_d are iid, every innovation Z_l can realize the maximum with positive probability, such that for every $l \in \text{An}(j)$,

$$\mathbb{P}\left(\left\{U_j = \bigvee_{k \in \text{An}(j)} \bar{b}_{kj} Z_k = \bar{b}_{lj} Z_l\right\} \cap \left\{U_i = \bigvee_{k \in \text{An}(i)} \bar{b}_{ki} Z_k = \bar{b}_{li} Z_l\right\}\right) > 0. \quad (\text{A.5})$$

Hence, without loss of generality we assume that this holds for $l = j$. Denote the random critical path $p := [j = k_0 \rightarrow \dots \rightarrow k_n = i]$ such that $\bar{d}_{ji}(p) = \bar{b}_{ji}$. Then, it follows on the event in (A.5) with $l = j$ from (3.2) that

$$\mathbb{P}\left(\frac{U_i}{U_j} < c\right) = \mathbb{P}\left(\frac{\bar{b}_{ji}}{\bar{b}_{jj}} < c\right) = \mathbb{P}\left(\frac{\varepsilon_j d_{ji}(p) \prod_{l=0}^{n-1} \varepsilon_{k_{l+1}}}{\varepsilon_j} < c\right)$$

$$= \mathbb{P} \left(d_{ji}(p) \prod_{l=0}^{n-1} \varepsilon_{k_{l+1}} < c \right) > 0,$$

since $d_{ji}(p) \leq b_{ji} < c$ and $\varepsilon \geq 1$. Hence, c is no lower bound and together with part c) this entails the support for $j \in \text{an}(i)$.

Now assume $j \notin \text{An}(i)$ such that $b_{ji} = 0$. Assume that U_i/U_j is lower bounded by some $c > 0$. Then by (3.6),

$$\frac{U_i}{U_j} \geq c \iff \bigvee_{k \in \text{An}(i)} \bar{b}_{ki} Z_k \geq c \bigvee_{k \in \text{An}(j)} \bar{b}_{kj} Z_k,$$

which is equivalent to

$$\bigvee_{k \in \text{An}(i)} \bar{b}_{ki} Z_k \geq c \left(\bigvee_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{kj} Z_k \vee \bigvee_{k \in \text{An}(j) \setminus \text{An}(i)} \bar{b}_{kj} Z_k \right).$$

Therefore, it holds in particular, that

$$\bigvee_{k \in \text{An}(i)} \bar{b}_{ki} Z_k \geq c \bar{b}_{lj} Z_l \tag{A.6}$$

for every $l \in \text{An}(j) \setminus \text{An}(i)$. This set is non-empty since $j \notin \text{An}(i)$, so it contains at least j . However, since the innovation and the noise variables are all independent and unbounded above, we have for every $l \in \text{An}(j) \setminus \text{An}(i)$

$$\mathbb{P} \left(Z_l \geq \frac{\bigvee_{k \in \text{An}(i)} \bar{b}_{ki} Z_k}{c \bar{b}_{lj}} \right) > 0,$$

contradicting (A.6) and, hence, the assumption of a lower positive bound c for U_i/U_j .

The upper interval limits of U_i/U_j for $j \in \text{an}(i)$ and $j \notin \text{An}(i)$ follow from changing the roles of i and j . For $j \neq i$, the ratio U_i/U_j always contains ε_i or ε_j and both random variables are atom-free and independent of all innovations Z_1, \dots, Z_d and ε_k for $k \neq i$ and $k \neq j$. Therefore, the ratio inherits the continuity of the noise variables and part d) follows.

(e) For $j = i$ we have $b_{ji} = 1 \neq 0 = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}$. If $j \notin \text{An}(i)$ we have $b_{ji} = \bar{b}_{ji} = 0$ by (3.4).

Next assume that $j \in \text{an}(i)$, and $b_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}} \neq 0$. Then there is a path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ from j to i with non-random path weight $d_{ji}(p) = b_{ji}$, which is not the edge $j \rightarrow i$.

For a contradiction, assume that $\bar{b}_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{\bar{b}_{kk}}$. This is equivalent to the edge $j \rightarrow i$ being the random critical path. However, every path $p \in P_{ji}$ has random path weight, which depends on both noise variables ε_i and ε_j , so in particular, the non-random critical path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$

from j to i with path weight $d_{ji}(p) = b_{ji}$ is one of these paths. Therefore, by (3.2) and since $b_{ji} > c_{ji}$, the random path weight of p is

$$\bar{d}_{ji}(p) = b_{ji} \varepsilon_j \prod_{l=0}^{n-1} \varepsilon_{k_{l+1}} \geq b_{ji} \varepsilon_j \varepsilon_i > c_{ji} \varepsilon_j \varepsilon_i = \bar{b}_{ji},$$

where we have used that $\varepsilon \geq 1$. This is a contradiction and hence $\bar{b}_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}$.

(f) The assumptions $b_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{b_{jk} b_{ki}}{b_{kk}}$ and $\text{de}(j) \cap \text{an}(i) \neq \emptyset$ are equivalent to the edge $p_{\max} = [j \rightarrow i]$ being the only non-random critical path.

Let $p' = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i] \neq p_{\max}$ be the path such that $\bigvee_{p \in P_{ji} \setminus \{p_{\max}\}} \bar{d}_{ji}(p) = \bar{d}_{ji}(p')$. Then

$$\bigvee_{p \in P_{ji} \setminus \{p_{\max}\}} \bar{d}_{ji}(p) = \bar{d}_{ji}(p') = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}},$$

otherwise we can construct a path of larger random path weight from j to i passing through k as explained in the proof of part a). First assume that $\bar{b}_{ji} = \bar{d}_{ji}(p_{\max})$. Then, $\bar{b}_{ji} > \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}$ and $\text{de}(j) \cap \text{an}(i) \neq \emptyset$ is by (3.2) and (3.4) equivalent to

$$\bar{b}_{ji} > \bar{d}_{ji}(p') = \varepsilon_i \varepsilon_j d_{ji}(p') \prod_{l=0}^{n-2} \varepsilon_{k_{l+1}} \iff \frac{b_{ji}}{d_{ji}(p')} > \prod_{l=0}^{n-2} \varepsilon_{k_{l+1}}. \tag{A.7}$$

Since $\varepsilon \geq 1$, also $b_{ji}/d_{ji}(p') > 1$. Hence, the event given by (A.7) has positive probability which is however, strictly smaller than one, since the noise variables do not have an upper bound. Therefore, since $\bar{b}_{ji} \geq \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}}$, by part a), the complementary event

$$\left\{ \bar{b}_{ji} = \bigvee_{k \in \text{de}(j) \cap \text{an}(i)} \frac{\bar{b}_{jk} \bar{b}_{ki}}{b_{kk}} \right\}$$

is also having positive probability. □

PROOF OF LEMMA 3.11 (a) Suppose there is an edge $k_l \rightarrow k_{l+1}$ in p such that $c_{k_l k_{l+1}} \notin \mathcal{D}^B$. Then, $\text{de}(j) \cap \text{an}(i) \neq \emptyset$ and by Lemma 3.8 e) $\mathbb{P}(\bar{b}_{k_l k_{l+1}} = c_{k_l k_{l+1}} \varepsilon_{k_l} \varepsilon_{k_{l+1}}) = 0$, so we can replace the edge $k_l \rightarrow k_{l+1}$ by some other path to get a new path from j to i of larger random path weight than p . Hence, p is not a possible critical path realization. The same argument can be used for the reverse.

(b) First consider $\neg(S_{p_1} \cap S_{p_2} = \emptyset$ or for every $r \in S_{p_1} \cap S_{p_2}$ the sub-path of p_1 from j to r is a sub-path of p_2 or the sub-path of p_2 from l to r is a sub-path of p_1). Then there exists some node $r \in S_{p_1} \cap S_{p_2}$ such that $p_1 = [j \rightarrow$

$\dots \rightarrow s \rightarrow r \rightarrow \dots \rightarrow i]$ and $p_2 = [l \rightarrow \dots \rightarrow t \rightarrow r \rightarrow \dots \rightarrow m]$ with $s \neq t$. Denote by $p_{11} := [j \rightarrow \dots \rightarrow s]$ the sub-path of p_1 from j to s . We want to show by contradiction that the event (3.9) has probability zero. Therefore, we consider the subset of Ω such that (3.9) holds and show that it is a null-set. Since on this subset, p_1 is the random critical path and passes through s , by Lemma 3.8 b) we have $\bar{b}_{ji} = \frac{\bar{b}_{js}\bar{b}_{si}}{\bar{b}_{ss}}$ and $U_s = U_j\bar{b}_{js}/\varepsilon_j = U_j d_{js}(p_{11}) \prod_{k \in S_{p_{11}}} \varepsilon_k$. With the same argument it also holds that $\bar{b}_{ji} = \frac{\bar{b}_{jr}\bar{b}_{ri}}{\bar{b}_{rr}}$ and $U_r = U_j\bar{b}_{jr}/\varepsilon_j = U_j d_{js}(p_{11}) \prod_{k \in S_{p_{11}}} \varepsilon_k c_{sr} \varepsilon_r$. Hence, it must hold that $U_r = U_s c_{sr} \varepsilon_r$. By the same arguments, we also must have $U_r = U_t c_{tr} \varepsilon_r$, which together leads to

$$U_s c_{sr} = U_t c_{tr}.$$

This is by (3.4) and (3.6) equivalent to

$$c_{sr} \prod_{l \in \text{An}(s)} \varepsilon_l \prod_{p \in P_{l_s}} d_{ls}(p) \prod_{k \in S_p} \varepsilon_k Z_l = c_{tr} \prod_{l \in \text{An}(t)} \varepsilon_l \prod_{p \in P_{lt}} d_{lt}(p) \prod_{k \in S_p} \varepsilon_k Z_l.$$

Now since \mathcal{D} is acyclic, there cannot be a path from s to t and from t to s ; so without loss of generality we can assume that there is no path from t to s . However, the right-hand side of the equation always contains ε_t which is not part of the left-hand side. Since Z_1, \dots, Z_d as well as $\varepsilon_1, \dots, \varepsilon_i$ are atom-free and independent random variables, this can only happen on a null-set.

Next consider the reverse, i.e., $S_{p_1} \cap S_{p_2} = \emptyset$ or for every $r \in S_{p_1} \cap S_{p_2}$ the sub-path of p_1 from j to r is a sub-path of p_2 or the sub-path of p_2 from l to r is a sub-path of p_1 .

If $S_{p_1} \cap S_{p_2} = \emptyset$, then the probability of (3.9) is obviously positive. Without loss of generality we now assume that for every $r \in S_{p_1} \cap S_{p_2}$ the sub-path of p_2 from l to r is a sub-path of p_1 . We now define r to be the last common node of the two paths p_1 and p_2 . Then, p_1 and p_2 induce the paths $p' = [j \rightarrow \dots \rightarrow l \rightarrow \dots \rightarrow r]$, $p'' = [r \rightarrow \dots \rightarrow i]$ and $p''' = [r \rightarrow \dots \rightarrow m]$. Then

$$\begin{aligned} & \left\{ U_i = U_j d_{ji}(p_1) \prod_{k \in S_{p_1}} \varepsilon_k, U_m = U_l d_{lm}(p_2) \prod_{k \in S_{p_2}} \varepsilon_k \right\} \\ &= \left\{ U_r = U_j d_{jr}(p') \prod_{k \in S_{p'}} \varepsilon_k, U_i = U_r d_{ri}(p'') \prod_{k \in S_{p''}} \varepsilon_k, U_m = U_r d_{rm}(p''') \prod_{k \in S_{p'''}} \varepsilon_k \right\}, \end{aligned}$$

which has positive probability, since $S_{p'} \cap S_{p''} \cap S_{p'''} = \emptyset$. □

PROOF OF THEOREM 3.12 By the law of total probability we have for $x \geq 1$,

$$\begin{aligned} I(x) &:= \mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x\right) = \mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}\right) + \mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x, U_i \neq \tilde{U}_j \bar{b}_{ji}\right) \\ &=: I_1(x) + I_2(x) \end{aligned}$$

We denote all paths from j to i by $P_{ji} = \{p_1, \dots, p_r, p_{\max}\}$. There are two situations, either $r = 0$ (where we interpret the above set of paths as $\{p_{\max}\}$), or

$r \geq 1$. We first give a proof for $r \geq 1$. We start with $I_1(x)$. Since p_{\max} is generic, every path $p \neq p_{\max}$ from j to i has non-random edge weight $d_{ji}(p) < b_{ji}$. Therefore, with (3.7) in the first line, (3.4) in the third and (3.2) in the last, we have for $x > 1$,

$$\begin{aligned} I_1(x) &= \mathbb{P}(U_i/(\tilde{U}_j \varepsilon_j) \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}) \\ &= \mathbb{P}(\bar{b}_{ji}/\varepsilon_j \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}) \\ &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} \bar{d}_{ji}(p)/\varepsilon_j \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}\right). \end{aligned} \quad (\text{A.8})$$

By definition of \bar{b}_{ji} in (3.4), there is a path $p \in \{p_{\max}, p_1, \dots, p_r\}$ such that $\bar{d}_{ji}(p) = \bar{b}_{ji}$ and by continuity of ε the probability that multiple paths satisfy the equation is equal to 0. Therefore, again applying the law of total probability, we find

$$\begin{aligned} I_1(x) &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bar{d}_{ji}(p_{\max}) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \quad (\text{A.9}) \\ &+ \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bigvee_{p \in \{p_1, \dots, p_r\}} \bar{d}_{ji}(p) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bar{d}_{ji}(p_{\max}) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &+ \sum_{s=1}^r \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bar{d}_{ji}(p_s) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &=: I_{11}(x) + I_{12}(x). \end{aligned}$$

We first find upper and lower bounds for $I_{11}(x)$. We denote by P_{kji} all paths from k to i which pass through j . Using the simple identity

$$\{z_1 \vee z_2 \leq a, z_1 \vee z_2 = z_1\} = \{z_1 \leq a, z_2 \leq z_1\}, \quad (\text{A.10})$$

(3.4) and (3.6) imply

$$\begin{aligned} &\left\{ \bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x \right\} \cap \left\{ \bar{d}_{ji}(p_{\max}) = \bar{b}_{ji} \right\} \cap \left\{ U_i = \tilde{U}_j \bar{b}_{ji} \right\} \quad (\text{A.11}) \\ &= \left\{ \bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq \bar{d}_{ji}(p_{\max}) \right\} \cap \left\{ \bar{d}_{ji}(p_{\max}) \leq b_{ji}x \right\} \cap \left\{ U_i = \tilde{U}_j \bar{b}_{ji} \right\} \\ &= \left\{ \bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k \right\} \cap \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \left\{ U_i = \tilde{U}_j \bar{b}_{ji} \right\} \\ &= \bigcap_{p \in P_{ji} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_{\max}}} \varepsilon_k \right\} \cap \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \end{aligned}$$

$$\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \leq b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\}. \quad (\text{A.12})$$

Cancelling all noise variables possible, and since $\varepsilon > 1$, we find a lower bound

$$\begin{aligned} I_{11}(x) &= \mathbb{P} \left(\bigcap_{p \in P_{ji} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_{\max}} \setminus S_p} \varepsilon_k \right\} \bigcap \right. \\ &\quad \left. \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \bigcap \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \right. \right. \right. \\ &\quad \left. \left. \leq b_{ji} \prod_{k \in S_{p_{\max}} \setminus (S_p \cup \{l\})} \varepsilon_k \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\ &\geq \mathbb{P} \left(\bigcap_{p \in P_{ji} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \bigcap \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \bigcap \right. \\ &\quad \left. \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\ &= \mathbb{P} \left(\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right. \\ &\quad \left. \bigcap_{p \in P_{ji} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \right) \mathbb{P} \left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right) \\ &=: c_1 \mathbb{P} \left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right), \end{aligned} \quad (\text{A.13})$$

for some constant $c_1 \in [0, 1]$ by independence of the noise variables.

We show that $c_1 > 0$. To do so, recall that $b_{ji}/d_{ji}(p) > 1$ for every $p \neq p_{\max}$. Therefore, since $\{p \in P_{ji} \setminus \{p_{\max}\}\} \neq \emptyset$ and $\varepsilon > 1$,

$$\mathbb{P} \left(\bigcap_{p \in P_{ji} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \right) > 0. \quad (\text{A.14})$$

Next, we want to show that also

$$\mathbb{P} \left(\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) > 0. \quad (\text{A.15})$$

For this, observe that the left-hand side of the inequality in (A.15) does not contain Z_j , since all paths from j to i pass through j . Since \bar{b}_{lj} and the left-hand side of the inequality in (A.15) is independent of Z_j for all $l \in \{1, \dots, d\}$ and Z_j has unbounded support, Z_j can become arbitrarily large with positive probability such that (A.15) holds.

The intersection of the two events has also positive probability since (A.14) is independent of Z_j . This implies that $c_1 > 0$ and a positive lower bound for $I_{11}(x)$.

To get an upper bound, observe that $\varepsilon \geq 1$ and, hence, for every set S_p we have

$$\left\{ \varepsilon_k : k \in S_{p_{\max}} \text{ and } \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \subseteq \left\{ \varepsilon_k : k \in S_{p_{\max}} \text{ and } \prod_{k \in S_{p_{\max}} \setminus S_p} \varepsilon_k \leq x \right\}.$$

Therefore, starting with (A.13) we find the upper bound

$$\begin{aligned} I_{11}(x) &\leq \mathbb{P} \left(\bigcap_{p \in P_{j_i} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} x \right\} \cap \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \right. \\ &\quad \left. \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lj_i}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} x \vee \bar{b}_{lj} Z_l \right\} \right\} \right) \\ &= \mathbb{P} \left(\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lj_i}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} x \vee \bar{b}_{lj} Z_l \right\} \right\} \right) \\ &\quad \cap \bigcap_{p \in P_{j_i} \setminus \{p_{\max}\}} \left\{ \prod_{k \in S_p \setminus S_{p_{\max}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} x \right\} \mathbb{P} \left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right) \\ &= c_2(x) \mathbb{P} \left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right). \end{aligned} \quad (\text{A.16})$$

Since the innovations and the noise variables are atom-free, it follows that $\lim_{x \downarrow 1} c_2(x) = c_1$ and, therefore,

$$I_{11}(x) \sim c_1 \mathbb{P} \left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right), \quad x \downarrow 1. \quad (\text{A.17})$$

We next show that $I_{12}(x) = o(I_{11}(x))$ as $x \downarrow 1$. We have for each summand $m \in \{1, \dots, r\}$, using the simple identity (A.10) to obtain the third line,

$$\begin{aligned} &\mathbb{P} \left(\bigvee_{p \in P_{j_i}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji} x, \bar{d}_{ji}(p_m) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji} \right) \\ &\leq \mathbb{P} \left(\bigvee_{p \in P_{j_i}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji} x, \bar{d}_{ji}(p_m) = \bar{b}_{ji} \right) \\ &= \mathbb{P} \left(\bigvee_{p \in P_{j_i}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq \bar{d}_{ji}(p_m), \bar{d}_{ji}(p_m) \leq b_{ji} x \right) \\ &= \mathbb{P} \left(\bigcap_{p \in P_{j_i} \setminus \{p_m\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq \frac{d_{ji}(p_m)}{d_{ji}(p)} \prod_{k \in S_{p_m}} \varepsilon_k \right\} \cap \left\{ \prod_{k \in S_{p_m}} \varepsilon_k \leq \frac{b_{ji} x}{d_{ji}(p_m)} \right\} \right) \\ &\leq \mathbb{P} \left(\left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq \frac{d_{ji}(p_m)}{b_{ji}} \prod_{k \in S_{p_m}} \varepsilon_k \right\} \cap \left\{ \prod_{k \in S_{p_m}} \varepsilon_k \leq \frac{b_{ji} x}{d_{ji}(p_m)} \right\} \right) \end{aligned} \quad (\text{A.18})$$

Now the first event rewrites as $\{\frac{b_{ji}}{d_{ji}(p_m)} \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq \prod_{k \in S_{p_m}} \varepsilon_k\} \subseteq \{\frac{b_{ji}}{d_{ji}(p_m)} \leq \prod_{k \in S_{p_m}} \varepsilon_k\}$, since $\varepsilon > 1$. Moreover,

$$\left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq \frac{d_{ji}(p_m)}{b_{ji}} \prod_{k \in S_{p_m}} \varepsilon_k \right\} \cap \left\{ \prod_{k \in S_{p_m}} \varepsilon_k \leq \frac{b_{ji}x}{d_{ji}(p_m)} \right\} \subseteq \left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\}.$$

Hence,

$$(A.18) \leq \mathbb{P}\left(\left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \left\{ \prod_{k \in S_{p_m}} \varepsilon_k \in \left[\frac{b_{ji}}{d_{ji}(p_m)}, \frac{b_{ji}x}{d_{ji}(p_m)} \right] \right\}\right),$$

Moreover, since $\varepsilon \geq 1$, we have for every subset $S \subseteq S_{p_{\max}}$ that $1 \leq \prod_{k \in S} \varepsilon_k \leq x$, whenever $1 \leq \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x$. Therefore, for another node set \tilde{S} with $S \cap \tilde{S} = \emptyset$ we have

$$\prod_{k \in S} \varepsilon_k \prod_{k \in \tilde{S}} \varepsilon_k \in [a, b] \Rightarrow \prod_{k \in \tilde{S}} \varepsilon_k \in [a/x, b]. \tag{A.19}$$

Finally, since $\bar{d}_{ji}(p_m) = \bar{b}_{ji}$ and $d_{ji}(p_m) < d_{ji}(p_{\max})$ we have $S_{p_m} \setminus S_{p_{\max}} \neq \emptyset$. In total, we obtain

$$\begin{aligned} (A.18) &\leq \mathbb{P}\left(\left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \left\{ \prod_{k \in S_{p_m}} \varepsilon_k \in \left[\frac{b_{ji}}{d_{ji}(p_m)}, \frac{b_{ji}x}{d_{ji}(p_m)} \right] \right\}\right) \\ &\leq \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right) \mathbb{P}\left(\prod_{k \in S_{p_m} \setminus S_{p_{\max}}} \varepsilon_k \in \left[\frac{b_{ji}}{x d_{ji}(p_m)}, \frac{b_{ji}x}{d_{ji}(p_m)} \right]\right) \\ &= \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right) o(1), \quad x \downarrow 1, \end{aligned}$$

as the interval in the second probability gets arbitrarily small and the distribution of ε is atom-free. Comparing this upper bound with (A.17) we can see that every summand of $I_{12}(x)$ is negligible with respect to $I_{11}(x)$ as $x \downarrow 1$. Since there are only finitely many nodes and hence finitely many paths from j to i , we have proved that $I_{12}(x) = o(I_{11}(x))$ as $x \downarrow 1$. Hence,

$$I_1(x) \sim c_1 \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right), \quad x \downarrow 1. \tag{A.20}$$

Next, we assume that $r = 0$, i.e., that there is only one path p_{\max} from j to i . Then from (A.9) we find that $I_1(x) = I_{11}(x)$ and simplifies (A.13) to

$$\begin{aligned} I_1(x) &= \mathbb{P}\left(\left\{ \prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x \right\} \cap \right. \\ &\quad \left. \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lj}} \left\{ d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \leq b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k \vee \bar{b}_{lj} Z_l \right\} \right\}\right) \end{aligned}$$

$$\begin{aligned}
&\geq \mathbb{P}\left(\left\{\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right\} \cap \right. \\
&\quad \left. \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{l_i} \setminus P_{l_{j_i}}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\
&= \mathbb{P}\left(\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{l_i} \setminus P_{l_{j_i}}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\
&\mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right) = c_1 \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right)
\end{aligned}$$

for $c_1 > 0$. On the other hand,

$$\begin{aligned}
I_1(x) &\leq \mathbb{P}\left(\bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{l_i} \setminus P_{l_{j_i}}} \left\{ d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \leq b_{ji} x \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\
&\mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right) = c_2(x) \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right),
\end{aligned}$$

and, since Z and ε are atom-free it again follows that $\lim_{x \downarrow 1} c_2(x) = c_1$ and therefore (A.20) holds also for $r = 0$.

We next show that $I_2(x) = o(I_1(x))$ as $x \downarrow 1$. Since $I_{12}(x) = o(I_{11}(x))$ as $x \downarrow 1$, we can and do assume that

$$\bar{b}_{ji} = b_{ji} \varepsilon_j \prod_{k \in S_{p_{\max}}} \varepsilon_k. \quad (\text{A.21})$$

Moreover, since for all paths $p \in P_{l_{j_i}}$ we have $l \in \text{An}(i)$ if and only if $l \in \text{An}(j)$,

$$\begin{aligned}
U_i &= \bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{j_i}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \vee \bigvee_{l \in \text{An}(j)} \bigvee_{p \in P_{l_{j_i}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \\
&= \bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{j_i}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \vee b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k U_j
\end{aligned}$$

by (A.21). If $\bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{j_i}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l > b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k U_j$, then it follows that

$$U_i = \bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{j_i}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l. \quad (\text{A.22})$$

Moreover, it holds by (3.8), (3.4) and (3.2)

$$U_i = \bigvee_{k \in \text{An}(i)} \bar{b}_{ki} \tilde{U}_k \geq \bar{b}_{ji} \tilde{U}_j \geq b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k U_j. \quad (\text{A.23})$$

Hence, $U_i/U_j \leq b_{ji}x$ implies that $\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x$. Therefore, using (3.7), (A.22) and (A.21) we get

$$I_2(x) = \mathbb{P}\left(\frac{U_i}{U_j} \leq b_{ji}x, U_i > \tilde{U}_j \bar{b}_{ji}\right) = \mathbb{P}\left(U_i \in (\tilde{U}_j \bar{b}_{ji}, \tilde{U}_j \varepsilon_j b_{ji}x]\right)$$

$$\begin{aligned} &\leq \mathbb{P}\left(\left\{\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right\} \cap \right. \\ &\quad \left. \left\{\bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{ji}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \in \left(b_{ji} \tilde{U}_j \varepsilon_j \prod_{k \in S_{p_{\max}}} \varepsilon_k, b_{ji} \tilde{U}_j \varepsilon_j x\right)\right\}\right) \\ &\leq \mathbb{P}\left(\left\{\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right\} \cap \right. \\ &\quad \left. \left\{\bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{ji}}} d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \in \left(b_{ji} \tilde{U}_j \varepsilon_j, b_{ji} \tilde{U}_j \varepsilon_j x\right)\right\}\right), \end{aligned}$$

since $\varepsilon \geq 1$. Using that $U_j = \tilde{U}_j \varepsilon_j$ and $j \notin S_{p_{\max}}$ and the same argument as in (A.19), we get

$$\begin{aligned} I_2(x) &\leq \mathbb{P}\left(\left\{\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right\} \cap \right. \\ &\quad \left. \left\{\bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{ji}}} d_{li}(p) \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \in \left(\frac{b_{ji} \tilde{U}_j \varepsilon_j}{x}, b_{ji} \tilde{U}_j \varepsilon_j x\right)\right\}\right) \\ &= \mathbb{P}\left(\left\{\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right\}\right) \\ &\quad \mathbb{P}\left(\left\{\bigvee_{l \in \text{An}(i)} \bigvee_{p \in P_{l_i} \setminus P_{l_{ji}}} \frac{d_{li}(p)}{\varepsilon_j} \prod_{k \in (S_p \cup \{l\}) \setminus S_{p_{\max}}} \varepsilon_k Z_l \in \left(\frac{b_{ji} \tilde{U}_j}{x}, b_{ji} \tilde{U}_j x\right)\right\}\right), \end{aligned}$$

since \mathcal{D} being acyclic implies that \tilde{U}_j and ε_j are independent of ε_k for every $k \in S_{p_{\max}}$. For $x \downarrow 1$ the interval in the second probability gets arbitrarily small. Since the distribution of the noise-variables is atom-free and the left-hand side contains ε_j that is not included in \tilde{U}_j , this probability tends to zero as $x \downarrow 1$. Comparing this upper bound with (A.20) we can see that $I_2(x) = o(I_1(x))$ as $x \downarrow 1$. Since $I_{12}(x) = o(I_{11}(x))$, we have

$$I(x) \sim I_1(x) \sim I_{11}(x) \sim c_1 \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x\right), \quad x \downarrow 1, \tag{A.24}$$

holds, where the last asymptotic equivalence follows from (A.20). Moreover, we have by (A.9), using (3.2),(3.4) and (3.7),

$$\begin{aligned} I(x) \sim I_{11}(x) &= \mathbb{P}\left(\bigvee_{p \in P_{j_i}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji} x, \bar{d}_{ji}(p_{\max}) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &= \mathbb{P}\left(\bar{d}_{ji}(p_{\max})/\varepsilon_j \leq b_{ji} x, U_i = \tilde{U}_j \bar{d}_{ji}(p_{\max})\right) \\ &= \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x, U_i = U_j b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k\right), \quad x \downarrow 1, \tag{A.25} \end{aligned}$$

which, together with (A.24), proves the result. □

PROOF OF COROLLARY 3.14. We first show the result for $P_{ji} \setminus \{p_1, \dots, p_n\} \neq \emptyset$, i.e., there exists a path p from j to i with $d_{ji} < b_{ji}$. We start as in the proof of Theorem 3.12 for $x \geq 1$

$$I(x) = I_1(x) + I_2(x)$$

and similarly to (A.8), we again apply the law of total probability to $I_1(x)$

$$\begin{aligned} I_1(x) &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bigvee_{p \in \{p_1, \dots, p_n\}} \bar{d}_{ji}(p) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &\quad + \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bigvee_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \bar{d}_{ji}(p) = \bar{b}_{ji}, U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &= \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \bigvee_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \bar{d}_{ji}(p) \leq \right. \\ &\quad \left. \bigvee_{p \in \{p_1, \dots, p_n\}} \bar{d}_{ji}(p), U_i = \tilde{U}_j \bar{b}_{ji}\right) + \mathbb{P}\left(\bigvee_{p \in P_{ji}} d_{ji}(p) \prod_{k \in S_p} \varepsilon_k \leq b_{ji}x, \right. \\ &\quad \left. \bigvee_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \bar{d}_{ji}(p) > \bigvee_{p \in \{p_1, \dots, p_n\}} \bar{d}_{ji}(p), U_i = \tilde{U}_j \bar{b}_{ji}\right) \\ &=: \tilde{I}_{11}(x) + \tilde{I}_{12}(x). \end{aligned}$$

With the same arguments as in the proof of Theorem 3.12 we find upper and lower bounds for $\tilde{I}_{11}(x)$. Analogously to (A.12) and (A.13) we find

$$\begin{aligned} \tilde{I}_{11}(x) &= \mathbb{P}\left(\bigcap_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \bigvee_{\bar{p} \in \{p_1, \dots, p_n\}} \prod_{k \in S_{\bar{p}}} \varepsilon_k \right\} \cap \right. \\ &\quad \bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\} \cap \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lj}} \left\{ d_{li}(p) \prod_{k \in S_p \cup \{l\}} \varepsilon_k Z_l \leq \right. \right. \\ &\quad \left. \left. b_{ji} \bigvee_{\bar{p} \in \{p_1, \dots, p_n\}} \prod_{k \in S_{\bar{p}}} \varepsilon_k \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\}) \\ &\geq \mathbb{P}\left(\bigcap_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p \setminus (\cup_{i=1}^n S_{p_i})} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \cap \bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\} \right. \\ &\quad \left. \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lj}} \left\{ d_{li}(p) \prod_{k \in S_p \cup \{l\} \setminus (\cup_{i=1}^n S_{p_i})} \varepsilon_k Z_l \leq b_{ji} \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \right) \\ &= c_1 \mathbb{P}\left(\bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\}\right) \end{aligned}$$

and analogously to (A.16) we have

$$\tilde{I}_{11}(x) \leq \mathbb{P}\left(\bigcap_{p \in P_{ji} \setminus \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p \setminus (\cup_{i=1}^n S_{p_i})} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} x \right\} \cap \right)$$

$$\begin{aligned} & \bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\} \\ & \bigcap_{l \in \text{An}(i)} \left\{ \bigcap_{p \in P_{li} \setminus P_{lji}} \left\{ d_{li}(p) \prod_{k \in S_p \cup \{l\} \setminus (\cup_{i=1}^n S_{p_i})} \varepsilon_k Z_l \leq b_{ji}x \bigvee_{l \in \text{An}(j)} \bar{b}_{lj} Z_l \right\} \right\} \\ & = c_2(x) \mathbb{P} \left(\bigcap_{p \in \{p_1, \dots, p_n\}} \left\{ \prod_{k \in S_p} \varepsilon_k \leq x \right\} \right) \end{aligned}$$

With the same arguments as in the previous proof, we can show that $c_1 \in (0, 1)$ and $c_2(x) \rightarrow c_1$ for $x \downarrow 1$ and $\tilde{I}_{12}(x) = o(\tilde{I}_{11}(x))$ and $I_2(x) = o(I_1(x))$. Hence, the result follows. If $P_{ji} \setminus \{p_1, \dots, p_n\} = \emptyset$ the result follows analogously. \square

PROOF OF COROLLARY 3.15 From Theorem 3.12 we have as $x \downarrow 1$,

$$\begin{aligned} & \mathbb{P} \left(\bigwedge_{k=0}^n \frac{U_i^k}{U_j^k} \leq b_{ji}x \right) = 1 - \left(1 - \mathbb{P} \left(\frac{U_i}{U_j} \leq b_{ji}x \right) \right)^n \\ & = 1 - \left(1 - c(1 + o(1)) \mathbb{P} \left(\prod_{k \in S_p} \varepsilon_k \leq x \right) \right)^n \\ & = 1 - \sum_{k=0}^n \binom{n}{k} \left(-c(1 + o(1)) \mathbb{P} \left(\prod_{k \in S_p} \varepsilon_k \leq x \right) \right)^k \sim cn \mathbb{P} \left(\prod_{i=1}^n \varepsilon_{k_i} \leq x \right), \end{aligned}$$

where we have used the binomial theorem and the fact that the summands for $k \geq 2$ are negligible when n is fixed. \square

PROOF OF THEOREM 3.16. We give a proof for $P_{ji} \setminus \{p_1\} \neq \emptyset$ and $P_{lm} \setminus \{p_2\} \neq \emptyset$, i.e., p_1 and p_2 are not the only paths from j to i and from l to m , respectively. All other cases follow analogously. By the law of total probability, we have

$$\begin{aligned} & \mathbb{P} \left(\frac{U_i}{U_j} \leq b_{ji}x_1, \frac{U_m}{U_l} \leq b_{lm}x_2 \right) \\ & = \mathbb{P} \left(\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1, \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2, U_i = U_j b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, U_m = U_l b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k \right) \\ & + \mathbb{P} \left(\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1, \frac{U_m}{U_l} \leq b_{lm}x_2, U_i = U_j b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, U_m \neq U_l b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k \right) \\ & + \mathbb{P} \left(\frac{U_i}{U_j} \leq b_{ji}x_1, \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2, U_i \neq U_j b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, U_m = U_l b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k \right) \\ & + \mathbb{P} \left(\frac{U_i}{U_j} \leq b_{ji}x_1, \frac{U_m}{U_l} \leq b_{lm}x_2, U_i \neq U_j b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k, U_m \neq U_l b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k \right) \\ & =: I_1(x_1, x_2) + I_2(x_1, x_2) + I_3(x_1, x_2) + I_4(x_1, x_2) \end{aligned}$$

We first consider $I_1(x_1, x_2)$. Observe that for $I_{11}(x)$ defined in (A.9) we have by (A.25)

$$I_{11}(x) = \mathbb{P}\left(\prod_{k \in S_{p_{\max}}} \varepsilon_k \leq x, U_i = U_j b_{ji} \prod_{k \in S_{p_{\max}}} \varepsilon_k\right)$$

and hence, $I_1(x_1, x_2)$ is the bivariate extension to $I_{11}(x)$. For this reason, we can follow the proof of Theorem 3.12 at (A.11), we again find upper and lower bounds based on the decomposition

$$\begin{aligned} & \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_1}} \varepsilon_k \right\} \cap \left\{ \bigvee_{p \in P_{lm} \setminus \{p_2\}} \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{lm}}{d_{lm}(p)} \prod_{k \in S_{p_2}} \varepsilon_k \right\} \cap \\ & \left\{ \prod_{k \in S_{p_1}} \varepsilon_k \leq x_1 \right\} \cap \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_p \cup \{n\}} \varepsilon_k Z_n \leq \right. \right. \\ & b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k \bigvee_{n \in \text{An}(j)} \bar{b}_{nj} Z_n \left. \right\} \right\} \cap \left\{ \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2 \right\} \cap \\ & \bigcap_{n \in \text{An}(m)} \left\{ \bigcap_{p \in P_{nm} \setminus P_{nlm}} \left\{ d_{nm}(p) \prod_{k \in S_p \cup \{n\}} \varepsilon_k Z_n \leq b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k \bigvee_{n \in \text{An}(l)} \bar{b}_{nl} Z_n \right\} \right\}. \end{aligned}$$

Now for three paths p , p_1 and p_2 and a node i we denote

$$S_{p+i|p_1+p_2} := (S_p \cup \{i\}) \setminus (S_{p_1} \cup S_{p_2}) \quad \text{and} \quad S_{p \setminus p_1+p_2} := S_p \setminus (S_{p_1} \cup S_{p_2}).$$

On the set $\{\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1\} \cap \{\prod_{k \in S_{p_2}} \varepsilon_k \leq x_2\}$ we have for $x_1, x_2 > 1$, since $\varepsilon > 1$,

$$\begin{aligned} & \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_1}} \varepsilon_k \right\} \\ & = \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p \setminus S_{p_1}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_1} \setminus S_p} \varepsilon_k \right\} \\ & \supseteq \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p \setminus S_{p_1}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \supseteq \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p \setminus p_1+p_2} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)x_2} \right\} \end{aligned}$$

as well as

$$\begin{aligned} & \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_p \cup \{n\}} \varepsilon_k Z_n \leq b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k \bigvee_{n \in \text{An}(j)} \bar{b}_{nj} Z_n \right\} \right\} \\ & = \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in (S_p \cup \{n\}) \setminus S_{p_1}} \varepsilon_k Z_n \leq b_{ji} \prod_{k \in S_{p_1} \setminus S_p} \varepsilon_k \right. \right. \\ & \left. \bigvee_{n \in \text{An}(j)} \bar{b}_{nj} Z_n \right\} \right\} \supseteq \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in (S_p \cup \{n\}) \setminus S_{p_1}} \varepsilon_k Z_n \leq \right. \right. \\ & \left. \left. b_{ji} \bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n \setminus p_1+p_2}} \varepsilon_k Z_n \right\} \right\} \end{aligned}$$

$$\begin{aligned} &\supseteq \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_{p+n \setminus p_1+p_2}} \varepsilon_k Z_n \leq \right. \right. \\ &\quad \left. \left. \frac{b_{ji}}{x_2} \bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n \setminus p_1+p_2}} \varepsilon_k Z_n \right\} \right\}. \end{aligned}$$

Therefore,

$$\begin{aligned} I_1(x_1, x_2) &\geq \mathbb{P} \left(\left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_{p \setminus p_1+p_2}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)x_2} \right\} \cap \left\{ \bigvee_{p \in P_{lm} \setminus \{p_1\}} \prod_{k \in S_{p \setminus p_1+p_2}} \varepsilon_k \leq \frac{b_{lm}}{d_{lm}(p)x_1} \right\} \right. \\ &\quad \left. \cap \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_{p+n \setminus p_1+p_2}} \varepsilon_k Z_n \leq \frac{b_{ji}}{x_2} \bigvee_{n \in \text{An}(j)} \right. \right. \right. \\ &\quad \left. \left. \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n \setminus p_1+p_2}} \varepsilon_k Z_n \right\} \right\} \cap \bigcap_{n \in \text{An}(m)} \left\{ \bigcap_{p \in P_{nm} \setminus P_{nlm}} \left\{ d_{nm}(p) \prod_{k \in S_{p+n \setminus p_1+p_2}} \varepsilon_k Z_n \leq \frac{b_{lm}}{x_1} \right. \right. \\ &\quad \left. \left. \bigvee_{n \in \text{An}(l)} \bigvee_{\tilde{p} \in P_{nl}} d_{nl}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n \setminus p_1+p_2}} \varepsilon_k Z_n \right\} \right\} \\ &\quad \cap \left(\left\{ \prod_{k \in S_{p_1}} \varepsilon_k \leq x_1 \right\} \cap \left\{ \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2 \right\} \right) \\ &=: c_3(x_1, x_2) \mathbb{P} \left(\left\{ \prod_{k \in S_{p_1}} \varepsilon_k \leq x_1 \right\} \cap \left\{ \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2 \right\} \right). \end{aligned}$$

For an upper bound, observe that on $\{\prod_{k \in S_{p_1}} \varepsilon_k \leq x_1\} \cap \{\prod_{k \in S_{p_2}} \varepsilon_k \leq x_2\}$ we have

$$\begin{aligned} &\left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \prod_{k \in S_{p_1}} \varepsilon_k \right\} = \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p \setminus S_{p_1}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right. \\ &\quad \left. \prod_{k \in S_{p_1} \setminus S_p} \varepsilon_k \right\} \subseteq \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_p \setminus S_{p_1}} \varepsilon_k \leq \frac{b_{ji}x_1}{d_{ji}(p)} \right\} \\ &\subseteq \left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_{p \setminus p_1+p_2}} \varepsilon_k \leq \frac{b_{ji}x_1}{d_{ji}(p)} \right\} \end{aligned}$$

as well as

$$\begin{aligned} &\bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_p \cup \{n\}} \varepsilon_k Z_n \leq b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k \bigvee_{n \in \text{An}(j)} \bar{b}_{nj} Z_n \right\} \right\} \\ &\subseteq \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in (S_p \cup \{n\}) \setminus S_{p_1}} \varepsilon_k Z_n \leq \right. \right. \\ &\quad \left. \left. b_{ji} x_1 x_2 \bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n \setminus p_1+p_2}} \varepsilon_k Z_n \right\} \right\} \\ &\subseteq \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_{p+n \setminus p_1+p_2}} \varepsilon_k Z_n \leq \right. \right. \end{aligned}$$

$$b_{ji}x_1x_2 \bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \Big\}.$$

For this reason,

$$I_1(x_1, x_2) \leq c_4(x_1, x_2) \mathbb{P}\left(\left\{ \prod_{k \in S_{p_1}} \varepsilon_k \leq x_1 \right\} \cap \left\{ \prod_{k \in S_{p_2}} \varepsilon_k \leq x_2 \right\}\right),$$

with

$$\begin{aligned} c_4(x_1, x_2) &:= \mathbb{P}\left(\left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_{p \setminus \{p_1+p_2\}}} \varepsilon_k \leq \frac{b_{ji}x_1}{d_{ji}(p)} \right\} \cap \left\{ \bigvee_{p \in P_{lm} \setminus \{p_1\}} \prod_{k \in S_{p \setminus \{p_1+p_2\}}} \varepsilon_k \leq \frac{b_{lm}x_2}{d_{lm}(p)} \right\} \right. \\ &\cap \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_{p+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \leq b_{ji}x_1x_2 \right. \right. \\ &\bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \Big\} \Big\} \cap \bigcap_{n \in \text{An}(m)} \left\{ \bigcap_{p \in P_{nm} \setminus P_{nlm}} \left\{ d_{nm}(p) \right. \right. \\ &\left. \left. \prod_{k \in S_{p+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \leq b_{lm}x_1x_2 \bigvee_{n \in \text{An}(l)} \bigvee_{\tilde{p} \in P_{nl}} d_{nl}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \right\} \right\}. \end{aligned}$$

Since all random variables are continuous, $c_4(x_1, x_2)$ tends to $c_3(x_1, x_2)$ for $x_1, x_2 \downarrow 1$, and

$$c_3(x_1, x_2) \leq c \leq c_4(x_1, x_2)$$

with

$$\begin{aligned} c &:= \mathbb{P}\left(\left\{ \bigvee_{p \in P_{ji} \setminus \{p_1\}} \prod_{k \in S_{p \setminus \{p_1+p_2\}}} \varepsilon_k \leq \frac{b_{ji}}{d_{ji}(p)} \right\} \cap \left\{ \bigvee_{p \in P_{lm} \setminus \{p_1\}} \prod_{k \in S_{p \setminus \{p_1+p_2\}}} \varepsilon_k \leq \frac{b_{lm}}{d_{lm}(p)} \right\} \right. \\ &\cap \bigcap_{n \in \text{An}(i)} \left\{ \bigcap_{p \in P_{ni} \setminus P_{nji}} \left\{ d_{ni}(p) \prod_{k \in S_{p+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \leq b_{ji} \bigvee_{n \in \text{An}(j)} \bigvee_{\tilde{p} \in P_{nj}} \right. \right. \\ &d_{nj}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \Big\} \Big\} \cap \bigcap_{n \in \text{An}(m)} \left\{ \bigcap_{p \in P_{nm} \setminus P_{nlm}} \left\{ d_{nm}(p) \prod_{k \in S_{p+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \right. \right. \\ &\left. \left. \leq b_{lm} \bigvee_{n \in \text{An}(l)} \bigvee_{\tilde{p} \in P_{nl}} d_{nl}(\tilde{p}) \prod_{k \in S_{\tilde{p}+n} \setminus \{p_1+p_2\}} \varepsilon_k Z_n \right\} \right\}. \end{aligned}$$

Since $i \neq m$ we can use the same arguments as for c_1 in the proof of Theorem 3.12 to show that $c > 0$. Therefore, we only need to show that $I_i(x_1, x_2) = o(I_1(x_1, x_2))$ for $i \in \{2, 3, 4\}$. It is obvious that $I_2(x_1, x_2) = o(I_1(x_1, x_2))$ implies the other two cases. Using the same arguments from the proof of Theorem 3.12 regarding $I_2(x) = o(I_1(x))$ and $I_{12}(x) = o(I_{11}(x))$, the result follows. \square

Appendix B: Proofs of Section 4

PROOF OF PROPOSITION 4.3 We first consider the convergence of the simple minimum ratio $\bigwedge_{k=1}^n (U_i^k / U_j^k)$. For $j = i$ the result is obvious. Moreover, for

$j \in \text{an}(i)$ we have with Lemma 3.8 d), for $x > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\bigwedge_{k=1}^n \frac{U_i^k}{U_j^k} - b_{ji}\right| > x\right) = \lim_{n \rightarrow \infty} \left(\mathbb{P}\left(\frac{U_i}{U_j} - b_{ji} > x\right)\right)^n = 0,$$

showing that the minimum ratio converges for $n \rightarrow \infty$ in probability to b_{ji} . Since $\bigwedge_{k=1}^n U_i^k/U_j^k$ is non-increasing, it converges almost surely. For $j \notin \text{an}(i)$ we have $b_{ji} = 0$ and the same result holds. Therefore, the estimators (4.1), (4.4) and (4.5) converge almost surely. Considering the inequality (4.3) for the estimator (4.2), this estimator as well converges almost surely. Finally using (4.7) and (4.8), the same also holds for the estimator (4.10). \square

PROOF OF LEMMA 4.4 Assume first that the output $\hat{\mathbf{B}}$ is not idempotent. Then there exists an entry \hat{b}_{ji} in $\hat{\mathbf{B}}$ such that $\hat{b}_{ji} < \hat{b}_{jk}\hat{b}_{ki}$. Therefore, since the input matrix $\check{\mathbf{B}}$ is idempotent, we have $\hat{b}_{ji} = 0$ while $\hat{b}_{jk} > 0$ and $\hat{b}_{ki} > 0$. This is a contradiction to the if-condition on line 9 in the algorithm.

Now assume that there is an idempotent matrix \mathbf{B}' that preserves all values that are larger than δ_1 but contains more zero entries. Then there is an entry b'_{ji} such that $b'_{ji} = 0$ while $\hat{b}_{ji} > 0$. Since $\hat{b}_{ji} > 0$ there must be some $k \in \{j+1, \dots, i-1\}$ such that $\hat{b}_{jk} \notin S$ and $\hat{b}_{ki} \notin S$, otherwise we would have set \hat{b}_{ji} equal to zero. Because we sort pairs (j, i) by distance and (j, k) and (k, i) both have smaller distance it must also hold that both, \hat{b}_{jk} and \hat{b}_{ki} are strictly greater than zero. In comparison, since \mathbf{B}' is idempotent, either b'_{jk} or b'_{ki} is equal to zero.

Therefore, we have $b'_{jk} = 0$ while $\hat{b}_{jk} > 0$ or $b'_{ki} = 0$ while $\hat{b}_{ki} > 0$. In both cases, the distance compared to the pair (j, i) is decreased. Repeating this argument we can assume that $(j-i) = 1$. This, however, leads to a contradiction since \hat{b}_{ji} is set to zero for all pairs (j, i) of distance one if $\check{b}_{ji} < \delta_1$. \square

Appendix C: Proofs of Section 5

PROOF OF LEMMA 5.1 By Theorem 3.12 we get for $x \downarrow 1$,

$$\begin{aligned} \lim_{t \downarrow 0} \frac{\mathbb{P}(\ln(U_i/U_j) - \ln(b_{ji}) \leq tx)}{\mathbb{P}(\ln(U_i/U_j) - \ln(b_{ji}) \leq t)} &= \lim_{t \downarrow 0} \frac{\mathbb{P}(U_i/U_j \leq b_{ji} \exp(tx))}{\mathbb{P}(U_i/U_j \leq b_{ji} \exp(t))} \\ &= \lim_{t \downarrow 0} \frac{c \mathbb{P}\left(\prod_{k \in S_p} \varepsilon_k \leq \exp(tx)\right)}{c \mathbb{P}\left(\prod_{k \in S_p} \varepsilon_k \leq \exp(t)\right)} = \lim_{t \downarrow 0} \frac{\mathbb{P}\left(\sum_{k \in S_p} \ln(\varepsilon_k) \leq tx\right)}{\mathbb{P}\left(\sum_{k \in S_p} \ln(\varepsilon_k) \leq t\right)} = x^{\zeta(p)\alpha} \end{aligned}$$

for $\zeta(p) = |S_p|$ by Corollary 2.6 a) and the fact that $\ln(\varepsilon_k) \in RV_\alpha^0$. \square

For the proof of Theorem 5.2 we need the following distribution family.

Definition C.1. A positive random variable Y is Fréchet distributed with shape $\alpha > 0$ and scale $s > 0$ and we write $Y \sim \text{Fréchet}(\alpha, s)$ if the distribution

function of Y is given by

$$\Phi_{\alpha,s}(x) = \exp\left(-\left(\frac{x}{s}\right)^{-\alpha}\right), \quad x > 0.$$

The proof of Theorem 5.2 is divided into a proof of the one-dimensional marginal limit distributions, followed by the proof of the multidimensional result. We start with the one-dimensional limits.

Proposition C.2. *Let \mathbf{U} be a recursive ML vector with propagating noise on a DAG \mathcal{D} as defined in (3.1) and assume that the path $p := [j \rightarrow \dots \rightarrow i]$ from j to i is generic. Assume further that $\tilde{\varepsilon} = \ln(\varepsilon) \in RV_{\alpha}^0$. For the node set S_p choose $a_n \sim F_{\sum_{k \in S_p} \tilde{\varepsilon}_k}^{\leftarrow}(1/n)$ as $n \rightarrow \infty$. Let U^1, \dots, U^n be an iid sample from \mathbf{U} . Then*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n b_{ji}} \left(\bigwedge_{k=1}^n U_i^k / U_j^k - b_{ji}\right) \leq x\right) = \Psi_{(\zeta(p)\alpha, c^{1/(\zeta(p)\alpha)})(x)}, \quad x > 0,$$

for the same constant c as in Theorem 3.12, and $\Psi_{\alpha,s}$ denotes the Weibull distribution from Definition 2.2 with $x_L = 0$.

Proof. Define $X := \ln(U_i/U_j) - \ln(b_{ji})$ with distribution function F_X . Then by Lemma 5.1, $X \in RV_{\zeta(p)\alpha}^0$, which implies that $1/X \in RV_{\zeta(p)\alpha}^{\infty}$. Using e.g. Theorem 3.3.7 of [11], for

$$a_{1/X}(n) \sim F_{1/X}^{\leftarrow}(1 - 1/n) \sim 1/F_X^{\leftarrow}(1/n) \rightarrow \infty, \quad n \rightarrow \infty, \tag{C.1}$$

we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigvee_{k=1}^n \left(\frac{1}{X}\right)^k \leq a_{1/X}(n)x\right) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigwedge_{k=1}^n X^k \geq \frac{1}{a_{1/X}(n)x}\right) \\ &= \Phi_{\zeta(p)\alpha,1}(x), \quad x > 0, \end{aligned}$$

which implies by the continuity of X ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigwedge_{k=1}^n X^k \leq \frac{x}{a_{1/X}(n)}\right) = 1 - \Phi_{\zeta(p)\alpha,1}(1/x), \quad x > 0. \tag{C.2}$$

Choose now

$$\tilde{a}_n \sim F_X^{\leftarrow}(1/n) \sim 1/a_{1/X}(n) \downarrow 0, \quad n \rightarrow \infty. \tag{C.3}$$

Hence, we have with (C.2) and (C.3) by Lemma 5.1,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\tilde{a}_n} \left(\bigwedge_{k=1}^n \ln(U_i^k / U_j^k) - \ln(b_{ji})\right) \leq x\right) = 1 - \Phi_{\zeta(p)\alpha,1}(1/x), \quad x > 0. \tag{C.4}$$

Recall from Theorem 3.12 and the regular variation of $\tilde{\varepsilon}$, that for the same c as defined in Theorem 3.12 we have

$$\begin{aligned} F_X(x) &\sim c \mathbb{P}\left(\sum_{k \in S_p} \tilde{\varepsilon}_k \leq x\right) \sim \mathbb{P}\left(\sum_{k \in S_p} \tilde{\varepsilon}_k \leq xc^{1/(\zeta(p)\alpha)}\right) \\ &= F_{\sum_{k \in S_p} \tilde{\varepsilon}_k}(xc^{1/(\zeta(p)\alpha)}), \quad x \downarrow 0, \end{aligned}$$

which implies that $1/n \sim F_X(\tilde{a}_n) \sim F_{\sum_{k \in S_p} \tilde{\varepsilon}_k}(\tilde{a}_n c^{1/(\zeta(p)\alpha)})$. For the generalized inverses this implies that

$$\tilde{a}_n \sim F_X^{\leftarrow}(1/n) \sim c^{-1/(\zeta(p)\alpha)} F_{\sum_{k \in S_p} \tilde{\varepsilon}_k}^{\leftarrow}(1/n) \sim c^{-1/(\zeta(p)\alpha)} a_n, \quad n \rightarrow \infty.$$

From this we find

$$\mathbb{P}\left(\frac{1}{\tilde{a}_n} \left(\bigwedge_{k=1}^n \ln(U_i^k/U_j^k) - \ln(b_{ji})\right) \leq x\right) = \mathbb{P}\left(\bigwedge_{k=1}^n U_i^k/U_j^k \leq \exp(\tilde{a}_n x) b_{ji}\right), \quad x > 0.$$

A Taylor expansion around 0 yields $\exp(\tilde{a}_n x) = 1 + \tilde{a}_n x(1 + o(1))$ as $n \rightarrow \infty$, because $\tilde{a}_n \downarrow 0$. Since for $x > 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\tilde{a}_n b_{ji}} \left(\bigwedge_{k=1}^n U_i^k/U_j^k - b_{ji}\right) \leq x\right) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n b_{ji}} \left(\bigwedge_{k=1}^n U_i^k/U_j^k - b_{ji}\right) \leq c^{1/(\zeta(p)\alpha)} x\right), \end{aligned}$$

we obtain with (C.4)

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n b_{ji}} \left(\bigwedge_{k=1}^n U_i^k/U_j^k - b_{ji}\right) \leq x\right) &= 1 - \Phi_{\zeta(p)\alpha, c^{-1/(\zeta(p)\alpha)}}(1/x) \\ &= \Psi_{\zeta(p)\alpha, c^{1/(\zeta(p)\alpha)}}(x), \end{aligned}$$

which proves the assertion. □

Now we can prove Theorem 5.2.

PROOF OF THEOREM 5.2 As we shall find asymptotic independence of estimates between different node pairs, it suffices to prove the bivariate result.

We first simplify notation as follows. Assume pairs of nodes $(j, i) \neq (l, m)$ and denote the generic paths $p_1 = p_{ji}$ with node set S_{p_1} and $p_2 = p_{lm}$ with node set S_{p_2} , respectively. Further denote $a_n^1 = a_n^{(ji)}$ and $a_n^2 = a_n^{(lm)}$.

By Proposition C.2 we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} \left(\bigwedge_{k=1}^n U_i^k/U_j^k - b_{ji}\right) \geq x_1\right) &= \Phi_{\zeta(p_1)\alpha, c_1^{-1/(\zeta(p_1)\alpha)}}(1/x_1) \\ &= \exp\left(-\frac{x_1^{\zeta(p_1)\alpha}}{c_1}\right), \end{aligned}$$

for $x_1 > 0$. Since $\lim_{n \rightarrow \infty} (1 - \frac{a}{n})^n = \exp(-a)$ for $a \in \mathbb{R}$, we get by independence of the ratios U_i^k/U_j^k for $k = 1, \dots, n$ as $n \rightarrow \infty$,

$$\mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} (U_i/U_j - b_{ji}) \leq x_1\right) = \frac{x_1^{\zeta(p_1)\alpha}}{c_1 n} (1 + o(1)), \quad x_1 > 0. \tag{C.5}$$

With the same argument, we have

$$\mathbb{P}\left(\frac{1}{a_n^2 b_{lm}} (U_m/U_l - b_{lm}) \leq x_2\right) = \frac{x_2^{\zeta(p_2)\alpha}}{c_2 n} (1 + o(1)), \quad x_2 > 0. \tag{C.6}$$

Therefore, we have on the one hand

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} \bigwedge_{k=1}^n (U_i^k/U_j^k - b_{ji}) \geq x_1\right) \mathbb{P}\left(\frac{1}{a_n^2 b_{lm}} \bigwedge_{k=1}^n (U_m^k/U_l^k - b_{lm}) \geq x_2\right) \\ &= \exp\left(-\frac{x_1^{\zeta(p_1)\alpha}}{c_1}\right) \exp\left(-\frac{x_2^{\zeta(p_2)\alpha}}{c_2}\right) = \exp\left(-\frac{x_1^{\zeta(p_1)\alpha}}{c_1} - \frac{x_2^{\zeta(p_2)\alpha}}{c_2}\right), \quad x_1, x_2 > 0, \end{aligned} \tag{C.7}$$

whereas, on the other hand, we have by independence of the bivariate ratios $(U_i^k/U_j^k, U_m^k/U_l^k)$ for $k = 1, \dots, n$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} \bigwedge_{k=1}^n (U_i^k/U_j^k - b_{ji}) \geq x_1, \frac{1}{a_n^2 b_{lm}} \bigwedge_{k=1}^n (U_m^k/U_l^k - b_{lm}) \geq x_2\right) \\ &= \lim_{n \rightarrow \infty} \left\{ \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} (U_i^k/U_j^k - b_{ji}) \geq x_1, \frac{1}{a_n^2 b_{lm}} (U_m^k/U_l^k - b_{lm}) \geq x_2\right) \right\}^n \\ &= \lim_{n \rightarrow \infty} \left\{ 1 - \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} (U_i/U_j - b_{ji}) \leq x_1\right) - \mathbb{P}\left(\frac{1}{a_n^2 b_{lm}} (U_m/U_l - b_{lm}) \leq x_2\right) \right. \\ & \quad \left. + \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} (U_i/U_j - b_{ji}) \leq x_1, \frac{1}{a_n^2 b_{lm}} (U_m/U_l - b_{lm}) \leq x_2\right) \right\}^n. \end{aligned} \tag{C.8}$$

By (A.23) we have $U_i \geq b_{ji} \prod_{k \in S_{p_1}} \varepsilon_k U_j$ and $U_m \geq b_{lm} \prod_{k \in S_{p_2}} \varepsilon_k U_l$, which implies

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} (U_i/U_j - b_{ji}) \leq x_1, \frac{1}{a_n^2 b_{lm}} (U_m/U_l - b_{lm}) \leq x_2\right) \\ & \leq \mathbb{P}\left(\frac{1}{a_n^1} \left(\prod_{k \in S_{p_1}} \varepsilon_k - 1\right) \leq x_1, \frac{1}{a_n^2} \left(\prod_{k \in S_{p_2}} \varepsilon_k - 1\right) \leq x_2\right). \end{aligned} \tag{C.9}$$

Since $(j, i) \neq (l, m)$, either $S_{p_1} \setminus S_{p_2} \neq \emptyset$ or $S_{p_2} \setminus S_{p_1} \neq \emptyset$ and without loss of generality, we assume $S_{p_2} \setminus S_{p_1} \neq \emptyset$. Since $\varepsilon \geq 1$ and all ε_k are independent, we get

$$(C.9) \leq \mathbb{P}\left(\frac{1}{a_n^1} \left(\prod_{k \in S_{p_1}} \varepsilon_k - 1\right) \leq x_1, \frac{1}{a_n^2} \left(\prod_{k \in S_{p_2} \setminus S_{p_1}} \varepsilon_k - 1\right) \leq x_2\right)$$

$$= \mathbb{P}\left(\frac{1}{a_n^1} \left(\prod_{k \in S_{p_1}} \varepsilon_k - 1 \right) \leq x_1\right) \mathbb{P}\left(\frac{1}{a_n^2} \left(\prod_{k \in S_{p_2} \setminus S_{p_1}} \varepsilon_k - 1 \right) \leq x_2\right). \quad (\text{C.10})$$

By Corollary 2.6 b) we know that $(\prod_{k \in S_{p_1}} \varepsilon_k - 1) \in RV_{\zeta(p_1)\alpha}^0$. Moreover, observe that by a Taylor expansion we have $a_n^1 \sim F_{\sum_{k \in S_p} \bar{\varepsilon}_k}^{\leftarrow}(1/n) \sim F_{\prod_{k \in S_p} \varepsilon_k - 1}^{\leftarrow}(1/n)$ as $n \rightarrow \infty$. Therefore, by Theorem 3.3.7 of [11] as in the proof of Proposition C.2, similarly to (C.4), it holds that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n^1} \left(\bigwedge_{t=1}^n \prod_{k \in S_{p_1}} \varepsilon_k^t - 1 \right) \geq x_1\right) = \Phi_{\zeta(p_1)\alpha, 1}(1/x_1) \\ & = \exp\left(-x_1^{\zeta(p_1)\alpha}\right), \quad x_1 > 0. \end{aligned}$$

We proceed as in (C.5) and (C.6) to obtain as $n \rightarrow \infty$

$$\mathbb{P}\left(\frac{1}{a_n^1} \left(\prod_{k \in S_{p_1}} \varepsilon_k - 1 \right) \leq x_1\right) = \left(\frac{x_1^{\zeta(p_1)\alpha}}{n}\right)(1 + o(1)), \quad x_1 > 0. \quad (\text{C.11})$$

Moreover, since $S_{p_2} \setminus S_{p_1} \neq \emptyset$, ε is atom-free and $a_n^2 \rightarrow 0$ as $n \rightarrow \infty$, we have

$$\mathbb{P}\left(\frac{1}{a_n^2} \left(\prod_{k \in S_{p_2} \setminus S_{p_1}} \varepsilon_k - 1 \right) \leq x_2\right) = o(1), \quad x_2 > 0. \quad (\text{C.12})$$

Therefore, we have by (C.10), (C.11) and (C.12)

$$(\text{C.9}) = \left(\frac{x_1^{\zeta(p_1)\alpha}}{n}\right)(1 + o(1))o(1), \quad x_1, x_2 > 0.$$

Comparing this with (C.5) and (C.6), we find that the last term in (C.8) is negligible. Hence, we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{a_n^1 b_{ji}} \bigwedge_{k=1}^n \left(U_i^k / U_j^k - b_{ji}\right) \geq x_1, \frac{1}{a_n^2 b_{lm}} \bigwedge_{k=1}^n \left(U_m^k / U_l^k - b_{lm}\right) \geq x_2\right) \\ & = \lim_{n \rightarrow \infty} \left(1 - \frac{x_1^{\zeta(p_1)\alpha}}{c_1 n} (1 + o(1)) - \frac{x_2^{\zeta(p_2)\alpha}}{c_2 n} (1 + o(1))\right)^n \\ & = \exp\left(-\frac{x_1^{\zeta(p_1)\alpha}}{c_1} - \frac{x_2^{\zeta(p_2)\alpha}}{c_2}\right). \end{aligned}$$

Comparing this to (C.7) yields the result. \square

References

- [1] C. Améndola, C. Klüppelberg, S. Lauritzen, and G. Tran. Conditional independence in max-linear Bayesian networks. *Ann. Appl. Prob.*, 2020. To appear. arXiv:2002.09233. MR3971300

- [2] F. Baccelli, G. Cohen, G.J. Olsder, and J-P. Quadrat. *Synchronization and Linearity: An Algebra for Discrete Event Systems*. Wiley, Chichester, 1993. [MR1204266](#)
- [3] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, 2004. [MR2108013](#)
- [4] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987. [MR1015093](#)
- [5] K.A. Bollen. *Structural Equations with Latent Variables*. Wiley, New York, 1989. [MR0996025](#)
- [6] P. Butkovič. *Max-linear Systems: Theory and Algorithms*. Springer, London, 2010. [MR2681232](#)
- [7] W. Chen, M. Drton, and Y.S. Wang. On causal discovery with equal variance assumption. *Biometrika*, 106(4):973–980, 2019. [MR4031210](#)
- [8] R.A. Davis and W.P. McCormick. Estimation for first-order autoregressive processes with positive or bounded innovations. *Stoch. Proc. Appl.*, 31:237–250, 1989. [MR0998115](#)
- [9] R.A. Davis and S.I. Resnick. Basic properties and prediction of max-arma processes. *Advances in Applied Probability*, 21(4):781–803, 1989. [MR1039628](#)
- [10] J. Einmahl, A. Kiriliouk, and J. Segers. A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, 21(2):205–233, 2018. [MR3800297](#)
- [11] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg, 1997. [MR1458613](#)
- [12] S. Engelke and A. Hitz. Graphical models for extremes. *JRSS B*, 82:871–932, 2020. [MR4136498](#)
- [13] N. Gissibl. *Graphical Modeling of Extremes: Max-linear Models on Directed Acyclic Graphs*. PhD thesis, Technical University of Munich, 2018.
- [14] N. Gissibl and C. Klüppelberg. Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720, 2018. [MR3779699](#)
- [15] N. Gissibl, C. Klüppelberg, and S. L. Lauritzen. Identifiability and estimation of recursive max-linear models. *Scandinavian Journal of Statistics*, 48(1):188–211, 2021. [MR4233170](#)
- [16] P. Hall, L. Peng, and Q. Yao. Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference*, 103:51–63, 04 2002. [MR1896983](#)
- [17] P. Hall and I. Van Keilegom. Nonparametric “regression” when errors are positioned at end-points. *Bernoulli*, 15(3):614–633, 2009. [MR2555192](#)
- [18] G.F. Italiano. Amortized efficiency of a path retrieval data structure. *Theoretical Computer Science*, 48(3):273–281, 1986. [MR0895800](#)
- [19] A. Janßen and P. Wan. k -means clustering of extremes. *Electronic Journal of Statistics*, 14(1):1211 – 1233, 2020. [MR4071364](#)
- [20] M. Jirak, A. Meister, and M. Reiss. Adaptive function estimation in nonparametric regression with one-sided errors. *Annals of Statistics*, 42(5):1970–2002, 2014. [MR3262474](#)
- [21] M. Joswig. *Essentials of Tropical Combinatorics*. Springer, Heidelberg. To

- appear, 2021.
- [22] J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, 27(4):887–906, 1956. [MR0086464](#)
 - [23] C. Klüppelberg and M. Krali. Estimating an extreme Bayesian network via scalings. *Journal of Multivariate Analysis*, 181:104672, 2021. doi.org/10.1016/j.jmva.2020.104672. [MR4157685](#)
 - [24] C. Klüppelberg and S. Lauritzen. Bayesian networks for max-linear models. In F. Biagini, G. Kauermann, and T. Meyer-Brandis, editors, *Network Science – An Aerial View from Different Perspectives*. Springer, 2020. [MR3971300](#)
 - [25] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA, 2009. [MR2778120](#)
 - [26] J.B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956. [MR0078686](#)
 - [27] S.L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, United Kingdom, 1996. [MR1419991](#)
 - [28] S.L. Lauritzen. Causal inference from graphical models. In O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, editors, *Complex Stochastic Systems*, pages 63–107. Chapman and Hall/CRC Press, London/Boca Raton, 2001. [MR1893411](#)
 - [29] S.L. Lauritzen, A.P. Dawid, B.N. Larsen, and H-G Leimer. Independence properties of directed Markov fields. *Networks*, 20(5):491–505, 1990. [MR1064735](#)
 - [30] D. Maclagan and B. Sturmfels. *Introduction to Tropical Geometry*. Graduate Studies in Mathematics, Vol. 161. American Mathematical Society, Providence, Rhode Island, 2015. [MR3287221](#)
 - [31] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2nd edition, 2009. [MR2548166](#)
 - [32] J. Peters, J.M. Mooij, D. Janzing, and B. Schölkopf. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15:2009–2053, 2014. [MR3231600](#)
 - [33] R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985. [MR0790201](#)
 - [34] R.L. Smith. Nonregular regression. *Biometrika*, 81(1):173–183, 1994. [MR1279665](#)
 - [35] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000. [MR1815675](#)
 - [36] N. Tran and J. Yu. Product-mix auctions and tropical geometry. *Mathematics of Operations Research*, 44(4):1396–1411, 2019. [MR4032448](#)