Technische Universität München
TUM School of Computation, Information and Technology

TUM

Exchangeable Minimum-Infinitely Divisible Sequences: Characterization, Simulation And A Detour To Model Selection

Florian Brück

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology

der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Mathias Drton

Prüfer*innen der Dissertation:

1.   Prof. Dr. Matthias Scherer
2.   Prof. Dr. Bruno Rémillard
3.   Prof. Dr. Marco Oesting

Die Dissertation wurde am 11.01.2023 bei der Technischen Universität München eingereicht

und durch die TUM School of Computation, Information and Technology am 28.04.2023

angenommen.

Technische Universität München

School of Computation, Information and Technology

Exchangeable Minimum-Infinitely Divisible Sequences: Characterization, Simulation And A Detour To Model Selection

Dissertation

by

Florian Brück

Abstract

This thesis contributes to the theoretical foundation of modeling high-dimensional random phenomena, with an emphasis on (exchangeable) multivariate extreme events.

First, it provides a one-to-one correspondence of sequences of exchangeable random variables whose multivariate marginal distributions are minimum-infinitely divisible and the class of non-negative and non-decreasing càdlàg infinitely divisible stochastic processes. We thereby unify several preceding research articles concerned with one-to-one correspondences of sequences of exchangeable random variables and non-negative and non-decreasing càdlàg processes under one common theoretical umbrella. Additionally, we provide an extension of de Finetti's seminal Theorem from finite (probability) measures to so-called exponent measures. The theoretical results are then used to describe construction schemes of sequences of exchangeable random variables whose multivariate marginal distributions are minimum-infinitely divisible.

Second, an exact simulation algorithm for continuous maximum-infinitely divisible stochastic processes is developed. The simulation algorithm is based on the exponent measure of the continuous maximum-infinitely divisible stochastic process and relies on the ability to simulate the exponent measure on certain sets of finite measure, while it avoids the derivation of the regular conditional probability distributions of the exponent measure. Moreover, the simulation algorithm is general enough to cope with non-continuous marginal distributions. Special emphasis is put on the simulation of sequences of exchangeable random variables whose multivariate marginal distributions are maximum-infinitely divisible.

Third, the thesis provides a general statistical model comparison framework with a focus on non-smooth criterion functions. The proposed model comparison test accounts for the influence of parameter estimation and asymptotically follows a normal distribution. Two estimators for the asymptotic variance of the test statistic are derived. A particular emphasis is put on the asymptotic normality of a suitably normalized Clarke test for non-nested model comparison, whose originally claimed asymptotic distribution in (Kevin A. Clarke. "Nonparametric model discrimination in international relations". In: *Journal of Conflict Resolution* 47.1 (2003), pp. 72–93. DOI: 10.1177/0022002702239512, Kevin A. Clarke. "A simple distribution-free test for nonnested model selection". In: *Political Analysis* 15.3 (2007), pp. 347–363. DOI: doi:10.1093/pan/mpm004) is thereby shown to be incorrect. Potential applications of such model comparison tests are sketched and a simulation study empirically illustrates the theoretical results.

Zusammenfassung

Diese Arbeit leistet einen Beitrag zu den theoretischen Grundlagen der Modellierung hochdimensionaler Zufallsphänomene, wobei der Schwerpunkt auf (austauschbaren) multivariaten Extremereignissen liegt.

Zuerst wird eine Eins-zu-Eins-Beziehung von Sequenzen austauschbarer Zufallsvariablen, deren multivariate Randverteilungen minimum-unendlich teilbar sind, und der Klasse der nicht-negativen und nicht-fallenden unendlich teilbaren stochastischen Prozesse mit càdlàg Pfaden hergeleitet. Dadurch betten wir mehrere vorausgehende Forschungsartikel, die sich mit Eins-zu-Eins-Beziehungen von Sequenzen austauschbarer Zufallsvariablen und nicht-negativer und nicht-fallender càdlàg stochastischer Prozesse befassen, in einen gemeinsamen theoretischen Rahmen ein. Zusätzlich erweitern wir de Finetti's Theorem von (Wahrscheinlichkeits-)Maßen auf sogenannte Exponentenmaße. Die theoretischen Ergebnisse werden dann verwendet, um Konstruktionsschemata von Sequenzen austauschbarer Zufallsvariablen zu beschreiben, deren multivariate Randverteilungen minimum-unendlich teilbar sind.

Des Weiteren wird ein exakter Simulationsalgorithmus für stetige maximum-unendlich teilbare stochastische Prozesse entwickelt. Der Simulationsalgorithmus basiert auf dem Exponentenmaß des stetigen maximum-unendlich teilbaren stochastischen Prozesses und basiert auf der Annahme, dass das Exponentenmaß auf bestimmten Mengen mit endlicher Masse simuliert werden kann, wobei vermieden wird, die regulären bedingten Wahrscheinlichkeitsverteilungen des Exponentenmaßes bestimmen zu müssen. Außerdem kann der Simulationsalgorithmus auch auf nicht-stetige Randverteilungen angewendet werden. Besonderes Augenmerk wird auf die Simulation von Sequenzen austauschbarer Zufallsvariablen gelegt, deren multivariate Randverteilungen maximum-unendlich teilbar sind.

Zuletzt wird ein allgemeiner Rahmen für statistische Modellvergleiche mit Schwerpunkt auf nicht-differenzierbaren Kriteriumsfunktionen entwickelt. Der vorgeschlagene Modellvergleichstest berücksichtigt den Einfluss von Parameterschätzungen und folgt asymptotisch einer Normalverteilung. Es werden zwei Schätzer für die asymptotische Varianz der Teststatistik hergeleitet. Ein besonderer Schwerpunkt liegt auf der asymptotischen Normalität eines geeignet normalisierten Clarke-Tests für nicht-verschachtelte Modellvergleiche, dessen ursprünglich postulierte asymptotische Verteilung in (Kevin A. Clarke. "Nonparametric model discrimination in international relations". In: *Journal of Conflict Resolution* 47.1 (2003), pp. 72–93. DOI: `10.1177/0022002702239512`, Kevin A. Clarke. "A simple distribution-free test for nonnested model selection". In: *Political Analysis* 15.3 (2007), pp. 347–363. DOI: `doi:10.1093/pan/mpm004`) korrigiert wird. Mögliche Anwendungen solcher Modellvergleichstests werden skizziert und eine Simulationsstudie illustriert die theoretischen Resultate.

Acknowledgements

First and foremost, I want to thank my supervisor Matthias Scherer for his continuous support during the last years. He generously supported all of my ideas and plans and gave me the freedom to explore every topic that seemed of interest to me. Moreover, I want to explicitly thank him for convincing me that mathematics is not only about theoretical results, but also about its communication and accessibility to others.

I want to thank Jan-Frederik Mai for the excellent collaboration in our joint research project. He was always there to answer questions and discuss ideas, even beyond the scope of the project. My gratitude also goes to Aleksey Min for always being available to help with any matter, be it scientific or not. Moreover, I want to thank Aleksey Min for the many interesting discussions on our joint research projects. Further gratitude goes to Jean-David Fermanian for the very productive collaboration with Aleksey Min and me in two research projects over the course of the last years. I am especially grateful to him for inviting me to CREST and providing me with kind hospitality during my stay.

Special thanks goes to WWK Versicherungen for their generous financial support of my PhD project. Moreover, I want to thank all of my colleagues at WWK for kindly taking me in as a member of the team from the first day and for enabling me to simultaneously pursue my academic interests while regularly working in a company.

Many thanks also go to all of my colleagues from M13 who I had great fun with in the last years. Especially, the countless coffee breaks with non-scientific discussions greatly contributed to keeping up my motivation. A special shout-out goes to the handball team of TSV Milbertshofen which largely tolerated my behavior when my brain was in sleep-mode and was responsible for many fun weekends.

Last but not least, I want to thank my family for the continuous support, not only during my PhD, but ever since I went to Munich. Even without having a clue of what I am doing, they encouraged me to pursue my interests and do what I enjoy. Without their support nothing of this would have been possible!

List of contributed publications

The following publications are included in this thesis as **core publications**:

- Florian Brück, Jan-Frederik Mai, and Matthias Scherer. "Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes". In: *Extremes* (2022). DOI: https://doi.org/10.1007/s10687-022-00450-w

- Florian Brück. "Exact simulation of continuous max-id processes with applications to exchangeable max-id sequences". In: *Journal of Multivariate Analysis* 193 (2023). DOI: https://doi.org/10.1016/j.jmva.2022.105117.

The following **additional publication** is included in this thesis:

- Florian Brück, Jean-David Fermanian, and Aleksey Min. "A corrected Clarke test for model selection and beyond". In: *Journal of Econometrics* (2022). DOI: https://doi.org/10.1016/j.jeconom.2021.12.013.

Contents

1 Introduction

Probabilistic modeling of complex multivariate phenomena is ubiquitous in applied sciences. In many fields, data for statistical investigations is available and the corresponding univariate phenomena are well understood and may be conveniently modeled. The embedding of these univariate phenomena into a multivariate model is often challenging, since it additionally requires to determine the usually unknown dependence structure between these phenomena. The easiest multivariate model that a researcher could think of is the assumption of independence between the observed univariate phenomena. However, this assumption essentially implies that there are no interactions between the phenomena, which can clearly be considered as an unreasonable assumption in many real world problems.

The simplest model assumption which can incorporate true interactions between observed phenomena is the assumption of (infinite)[1] exchangeability. In essence, exchangeability means that there is a single common factor, which determines the behavior of each observed phenomenon, but the individual realizations due to this predetermined behavior occur independently from each other. Therefore, one could argue that exchangeability is the slightest form of dependence between univariate phenomena. However, it should be noted that dependence relations in which perfect knowledge about one phenomenon determines the exact behavior of all other phenomena are also included in the concept of exchangeability. Thus, the extension of independence to exchangeability incorporates quite some modeling flexibility, while the inherent factor structure retains interpretability of the joint behavior of the univariate phenomena. When the assumption of exchangeability cannot be considered as adequate to model a multivariate phenomenon, the modeling complexity increases significantly. Usually, there is no general recipe to deal with these situations and researchers use approaches tailored to the specific modeling problem to reasonably represent the multivariate phenomenon.

In most modeling problems, several candidate models are explored. This poses the challenge of choosing the most appropriate model out of a collection of candidate models. In such situations, it is convenient to rely on statistical model comparison techniques that allow to decide if one model significantly outperforms another model. Such comparisons are usually based on statistical hypothesis tests, which are able to decide whether or not one model significantly outperforms another model in terms of a certain criterion function, which should reflect the relative goodness of fit of the compared models with respect to the modeling target.

In this thesis, we contribute to the theoretical foundation of the aforementioned aspects of modeling multivariate phenomena. Specifically, we contribute to the realm of modeling multivariate

---

[1]There are different concepts of exchangeability in the literature. First, the concept of finite exchangeability includes exchangeable random vectors for which there might not exists a higher dimensional exchangeable random vector in which it might be embedded as a subvector. On the other hand, the concept of infinite exchangeability assumes that an exchangeable random vector can be embedded into a sequence of exchangeable random variables. In this thesis we focus on infinite exchangeability, which is why we refer to infinite exchangeability when solely writing exchangeable, knowing that this simplification leads to a slight imprecision in our statements.

extreme events, meaning that we are focusing on modeling problems where dependencies of joint occurrences of "large" or "small" events are the modeling target. First, the core article [10] of this thesis derives the exchangeable dependence structures of those events, while, in a second step, the core article [8] of this thesis provides a tool to simulate and investigate non-exchangeable dependencies of those events. Finally, the additional article [9] of this thesis provides a general statistical model comparison framework, which allows to compare the goodness of fit of two competing models according to a possibly non-smooth criterion function.

2 Exchangeable sequences of random variables

The first rigorous mathematical formulation of the concept of (infinite) exchangeability has been developed by de Finetti in his seminal work [26]. Since then, there have been many generalizations of de Finetti's original formulation of exchangeability, e.g. [30] extended the concept of exchangeability to spaces of abstract objects, and many specific examples have been investigated in detail. Mathematically, exchangeability is a property of the law of a sequence of random objects. A sequence of random objects is called exchangeable (or exchangeable sequence) if its law is invariant under arbitrary deterministic permutations of finitely many of its components. We present de Finetti's Theorem in a form which is most appropriate for our purposes.

Theorem 1 (De Finetti's Theorem [26, 30]). Let $\boldsymbol{X} := (X_i)_{i \in \mathbb{N}} \in [-\infty, \infty]^{\mathbb{N}}$ denote a sequence of extended real-valued random variables. Then, $\boldsymbol{X}$ is an exchangeable sequence if and only if there exists some random distribution function $F := (F_t)_{t \in \mathbb{R}}$ such that, conditionally on F, $\boldsymbol{X}$ is an i.i.d. sequence. Moreover, the law of F is in one-to-one correspondence to the law of $\boldsymbol{X}$ and $\boldsymbol{X}$ has the stochastic representation

$$\boldsymbol{X} \sim \left(\inf \left\{ t \in \mathbb{R} \mid -\log \left(1 - F(t) \right) \geq E_i \right\} \right)_{i \in \mathbb{N}}, \tag{1}$$

where $(E_i)_{i \in \mathbb{N}}$ denotes a sequence of i.i.d. exponential random variables with mean 1.

De Finetti's Theorem tells us that every exchangeable sequence of extended real-valued random variables can be represented as the first passage time of a non-negative and non-decreasing càdlàg process

$$H := (H_t)_{t \in \mathbb{R}} := \left(-\log \left(1 - F(t) \right) \right)_{t \in \mathbb{R}}$$

over i.i.d. exponential barriers. This correspondence nicely illustrates that dependence properties of $\boldsymbol{X}$ may be associated to path properties of H. For example, one can easily deduce that $\boldsymbol{X}$ has ties, i.e. there exists some distinct $i, j \in \mathbb{N}$ such that $X_i = X_j$, if and only if H has jumps.

For many modeling purposes it is unrealistic to model a multivariate phenomenon directly via the representation (1). For instance, the marginal distribution function of X_i, given by $G(t) := P(X_i \leq t) = \mathbb{E}[F(t)]$, is identical for all margins of an exchangeable sequence, which is a constraint that is often violated in practice. Nevertheless, if one ignores the issue of identical

marginal distributions for a moment, the representation (1) is useful to study the dependence structure of $\boldsymbol{X}$, and it may be desirable to transfer this dependence structure to a model of the interactions of univariate phenomena.

It turns out that the just described way of thinking is indeed also mathematically justifiable, due to the celebrated result of Sklar in [63]. In essence, [63] tells us that the dependence structure of $\boldsymbol{X}$ may be isolated from its marginal distribution by analyzing the object

$$C_d(u_1, \ldots, u_d) = P\big(G(X_1) \leq u_1, \ldots, G(X_d) \leq u_d\big); \ (u_1, \ldots, u_d) \in [0, 1]^d, \tag{2}$$

which is called the copula of the d-dimensional margins of $\boldsymbol{X}$. C_d is called exchangeable, because it satisfies $C_d(u_1, \ldots, u_d) = C_d(u_{\pi(1)}, \ldots, u_{\pi(d)})$ for all permutations π of $\{1, \ldots, d\}$ as it represents the dependence structure of the exchangeable random vector $(X_1, \ldots, X_d)$. A copula C_d can then be used to construct a random vector $(Z_1, \ldots, Z_d)$ equipped with arbitrary marginal distributions $G_1, \ldots, G_d$ by defining its distribution function via

$$P\big(Z_1 \leq z_1, \ldots, Z_d \leq z_d\big) := C_d\big(G_1(z_1), \ldots, G_d(z_d)\big).$$

The corresponding random vector $(Z_1, \ldots, Z_d)$ shares the dependence structure (copula) with $(X_1, \ldots, X_d)$, but follows arbitrary marginal distributions $G_1, \ldots, G_d$.

For practical applications, it is often reasonable and convenient to derive the dependence structure of a random vector from (1) in terms of the copula C_d and to separately model its marginal distributions $G_1, \ldots, G_d$. When there is no contradicting evidence, an exchangeable dependence structure could naturally be considered, since it provides a simple, yet flexible, way to model the dependence of a random vector. However, there are also some pitfalls that are accompanied by an exchangeable dependence structure that should be kept in mind, e.g. induced hierarchies of marginal phenomena, as pointed out in [45]. Since the transition from the exchangeable sequence $\boldsymbol{X}$ to its copulas $(C_d)_{d \in \mathbb{N}}$ is rather simple in most of the considered examples, we will solely focus on studying exchangeable sequences $\boldsymbol{X}$ in the remainder of this thesis.

2.1 Associating classes of exchangeable sequences to classes of non-negative and non-decreasing càdlàg processes

De Finetti's Theorem shows that the variety of laws of exchangeable sequences is as large as the variety of laws of non-negative and non-decreasing càdlàg processes. Since the implications of this correspondence are hard to grasp in their full mathematically generality, a natural question that arises from this correspondence is whether certain classes or properties of exchangeable sequences may be connected to certain classes or properties of non-negative and non-decreasing càdlàg processes. This question has been intensively investigated in the literature. Here, we briefly recall some of the correspondences that have been established previously.

Exchangeable sequences associated to non-negative and non-decreasing functions with random parameter

We start with the simplest families of exchangeable sequences, apart from i.i.d. sequences, which arise in classical Bayesian statistics. Consider i.i.d. observations $\boldsymbol{X} := (X_1, X_2, \ldots)$ of a real-valued random variable X and assume that their unknown marginal distribution function stems from a finite-dimensional parametric family of distribution functions $(F_\theta(\cdot))_{\theta \in \Theta}$. To conduct inference about the law of X one must gather information about the unknown parameter θ. Classical Bayesian statistics attempts to accomplish this task by assuming that θ is a realization of a random variable with so-called prior distribution Q. Therefore, $\boldsymbol{X}$ becomes an exchangeable sequence of random variables associated to the non-negative and non-decreasing càdlàg process $\left(H_t^{(\theta)}\right)_{t \in \mathbb{R}} = (-\log(1 - F_\theta(t)))_{t \in \mathbb{R}}$. Inference about θ is then based on the conditional distribution of θ given the first n observations $(X_1, \ldots, X_n)$ from the exchangeable sequence $\boldsymbol{X}$, called the posterior distribution of θ.

Several characterizations of exchangeable sequences of random variables associated to certain non-negative and non-decreasing càdlàg processes of the form $\left(H_t^{(\theta)}\right)_{t \in \mathbb{R}}$ have been obtained in the literature. Here, we focus on some well-known examples for the special case of real-valued θ. For example, consider an i.i.d. sequence $(Y_1, Y_2, \ldots)$ of real-valued random variables with marginal distribution function $F(\cdot)$ and consider an independent non-negative random variable θ. Then

$$\boldsymbol{X} := (\theta Y_1, \theta Y_2, \ldots)$$

defines an exchangeable sequence of random variables with associated non-negative and non-decreasing càdlàg process $\left(H_t^{(\theta)}\right)_{t \in \mathbb{R}} := (-\log(1 - F(t/\theta)))_{t \in \mathbb{R}}$. Schönberg's Theorem [61] implies that when $(Y_1, Y_2, \ldots)$ denotes a sequence of i.i.d. standard normal distributed random variables, the law of $\boldsymbol{X}$ is spherical, since its d-dimensional marginal distributions are invariant under orthogonal transformations. Moreover, Kimberling's Theorem [36] implies that when $(Y_1, Y_2, \ldots)$ denotes a sequence of i.i.d. unit exponential distributed random variables, the law of $\boldsymbol{X}$ is 1-norm symmetric, since its d-dimensional marginal survival functions only depend on the 1-norm of its argument. In this case, the copula associated to $-\boldsymbol{X}$ belongs to the family of Archimedean copulas [48], a family of dependence structures which is well investigated in the literature due to its simple one-factor dependence structure and convenient analytical form. As a third example, consider $(Y_1, Y_2, \ldots)$ i.i.d. uniformly distributed on $[0, 1]$ and assume that θ has a (sufficiently regular) density. Then, according to [29], the d-dimensional marginal distributions of $\boldsymbol{X}$ have a density which solely depends on the ∞-norm of its argument. Finally, by providing a suitable density for Y_1, a unifying characterization of non-negative exchangeable sequences $(\theta Y_1, \theta Y_2, \ldots)$ with p-norm symmetric densities is provided in [55].

Exchangeable sequences associated to Lévy subordinators

The starting point of a series of research papers was the article [44], which associated the class of non-negative and non-decreasing càdlàg processes with independent and stationary increments that start at 0 in time 0, also known as Lévy subordinators, to the class of exchangeable sequences whose d-dimensional margins can be represented as

$$(X_1, \ldots, X_d) \sim \big(\min\{E_I \mid I \subset \{1, \ldots, d\}, i \in I\} \big)_{1 \leq i \leq d}, \tag{3}$$

where $(E_I)_{I \subset \{1,\ldots,d\}}$ denote independent exponential random variables whose mean only depends on the cardinality $|I|$. Random vectors with the stochastic representation (3) were introduced in [47] and their laws are known as so-called exchangeable Marshall–Olkin distributions. The Marshall–Olkin distribution may be viewed as a multivariate extension of the exponential distribution, since it has marginal exponential distributions and satisfies a specific multivariate lack-of-memory property. At the same time, Lévy subordinators are closely connected to the theory of infinitely divisible distributions, i.e. to the law of random variables H with the property that for every $n \in \mathbb{N}$ there exist independent and identically distributed random variables $\big(H^{(i,1/n)}\big)_{1 \leq i \leq n}$ such that $H \sim \sum_{i=1}^n H^{(i,1/n)}$. It is well-known that every non-negative infinitely divisible distribution may be obtained as the distribution of a Lévy subordinator at unit time and, vice versa, that every marginal distribution of a Lévy subordinator follows a non-negative infinitely divisible distribution. Moreover, due to the Lévy–Khintchine Theorem [40, 35], a non-negative infinitely divisible distribution is in one-to-one correspondence with a measure υ on $(0, \infty]$ satisfying a certain integrability condition and a scalar $b \geq 0$, which are called Lévy measure and drift of the infinitely divisible distribution. Thus, exchangeable sequences whose multivariate marginal distributions are of Marshall–Olkin kind are not only in one-to-one correspondence with Lévy subordinators, but also in one-to-one correspondence with tuples (υ, b) of Lévy measure and drift of non-negative infinitely divisible distributions.

Exchangeable sequences associated to non-negative and non-decreasing strong-idt processes

Extending the correspondence established in [44], [43] proved that when H in (1) is chosen to be strongly infinitely divisible with respect to time (strong-idt), i.e for i.i.d. copies $\big(H^{(i)}\big)_{i \in \mathbb{N}}$ of H and every $n \in \mathbb{N}$ the process H satisfies $(H_t)_{t \geq 0} \sim \Big(\sum_{i=1}^n H_{t/n}^{(i)} \Big)_{t \geq 0}$, one obtains that the d-dimensional marginal distributions of the associated exchangeable sequence follow so-called exchangeable min-stable multivariate exponential distributions. Min-stable multivariate exponential distributions naturally arise as the limit laws of scaled minima of i.i.d. random vectors. They are defined as those laws that arise when there exist possibly dependent exponential random variables $(X_i)_{1 \leq i \leq d}$, such that there are some i.i.d. random vectors $\big((Z_{i,1}, \ldots, Z_{i,d})\big)_{i \in \mathbb{N}}$ and sequences $(a_{n,j})_{1 \leq j \leq d}$ and $(b_{n,j})_{1 \leq j \leq d} \geq 0$ which satisfy

$$\left(\min_{1 \leq i \leq n} (Z_{i,1} - a_{n,1})/b_{n,1}, \ldots, \min_{1 \leq i \leq n} (Z_{i,d} - a_{n,d})/b_{n,d} \right) \xrightarrow{law} (X_1, \ldots, X_d). \tag{4}$$

5

It turns out that that the construction in (4) is already general enough to obtain all copulas associated to the limiting laws of scaled minima of i.i.d. random vectors. Since the possible limits of minima of univariate i.i.d. random variables are well-studied and rather easily described, the (survival) copulas associated to the limiting laws of scaled minima of i.i.d. random vectors are usually the central object of study in multivariate extreme value theory. These copulas are therefore called extreme value copulas and can be analytically characterized by the property that

$$C_d(u_1^t, \ldots, u_d^t) = C_d(u_1, \ldots, u_d)^t \text{ for all } t > 0.$$

[43] proved that the correspondence of non-negative and non-decreasing strong-idt processes to exchangeable sequences whose d-dimensional marginal distributions follow an exchangeable min-stable multivariate exponential distribution is in fact general enough to incorporate all exchangeable sequences with exponential margins whose multivariate marginal distributions are associated to an exchangeable extreme value copula.

Exchangeable sequences associated to additive subordinators

Another extension of the correspondence established in [44] was obtained by [41, 64], who proved that when H in (1) belongs to the class of stochastically continuous non-negative and non-decreasing càdlàg processes with independent increments that start at 0 in time 0, called additive subordinators, the associated class of exchangeable sequences has d-dimensional margins with stochastic representation

$$(X_1, \ldots, X_d) \sim \big(\min\{\tilde{E}_I \mid I \subset \{1, \ldots, d\}, i \in I\} \big)_{1 \leq i \leq d}, \tag{5}$$

where $\big(\tilde{E}_I \big)_{I \subset \{1, \ldots, d\}}$ denote independent non-negative random variables with continuous distribution function, which satisfy $\tilde{E}_{I_1} \sim \tilde{E}_{I_2}$ if $|I_1| = |I_2|$. Random vectors with representation (5) are called exchangeable exogenous shock models and may be seen as a natural extension of the Marshall–Olkin distribution due to their similarity with the stochastic representation (3), replacing the exponential law by arbitrary non-negative continuous distributions. Exchangeable exogenous shock models naturally arise in non-parametric Bayesian statistics, when one puts a prior distribution on the unknown distribution function of an extended real-valued random variable by setting $(F_t)_{t \geq 0} := (1 - \exp(-H_t))_{t \geq 0}$ for some additive subordinator H. Thus, under the prior distribution, the i.i.d. sequence associated to $(F_t)_{t \geq 0}$ becomes an exchangeable sequence with multivariate margins of the form (5). The appealing feature of this specific prior choice is that the posterior distribution of H is again a càdlàg process with independent increments, i.e. the prior distribution of F (resp. H) is (almost) conjugate, see [24, 18, 25].

Another theoretical application of the correspondence of exchangeable exogenous shock models and additive subordinators has been provided by [42], where the authors investigated the subclass of exchangeable exogenous shock models which correspond to self-similar additive subordinators, i.e. additive subordinators for which there exists some index of self-similarity $\gamma > 0$ such that

$(H_{at})_{t \in \mathbb{R}} \sim (a^\gamma H_t)_{t \in \mathbb{R}}$ for all $a > 0$. The law of H_1 is known to be self-decomposable, meaning that for every $c \in (0, 1)$ there exists some random variable $Z^{(c)}$, independent of H_1, such that $H_1 \sim c H_1 + Z^{(c)}$. On the other hand, for every self-decomposable distribution π and every index of self-similarity γ, there exists a (unique in law) self-similar additive subordinator H with index of self-similarity γ such that $H_1 \sim \pi$. Thus, there is a one-to-one correspondence of non-negative self-decomposable distributions to self-similar additive subordinators. The authors of [42] characterized the exchangeable sequences associated with self-similar additive subordinators $(H_t)_{t \in \mathbb{R}}$ via (1) in terms of analytical properties of the self-decomposable law of H_1. Moreover, it can be shown that the index of self-similarity of a self-similar additive process only influences the marginal distributions of the associated exchangeable sequence, but not its copulas. Thus, [42] provides a one-to-one correspondence of non-negative self-decomposable laws and the copulas of certain exchangeable sequences whose multivariate margins have stochastic representation (5).

Exchangeable sequences associated to random walks with non-negative infinitely divisible jumps

The geometric distribution[2] is the only univariate discrete distribution on $\mathbb{N}$ with a lack-of-memory property. Analogously to the univariate exponential distribution, there are several multivariate extensions of the univariate geometric distribution. The article [46] investigates, among other things, exchangeable sequences associated to the so-called narrow-sense multivariate geometric distributions, which are a specific multivariate extension of the geometric distribution. More precisely, in analogy to (3), the multivariate margins of narrow-sense multivariate geometric distributions have the stochastic representation

$$(X_1, \ldots, X_d) \sim \big(\min\{G_I \mid I \subset \{1, \ldots, d\}, i \in I\} \big)_{1 \leq i \leq d}, \tag{6}$$

where $(G_I)_{I \subset \{1, \ldots, d\}}$ denote independent geometric random variables whose success probabilities only depend on the cardinality $|I|$. [46] prove that an exchangeable sequence has multivariate margins with stochastic representation (6) if and only if the associated non-negative and non-decreasing càdlàg process is a random walk $(H_t)_{t \in \mathbb{R}} := \left(\sum_{i=1}^{\lfloor t \rfloor} A_i \right)_{t \in \mathbb{R}}$ with i.i.d. infinitely divisible jumps $(A_i)_{i \in \mathbb{N}}$.

Exchangeable sequences associated to non-negative and non-decreasing infinitely divisible processes as a unifying framework

The previous sections roughly summarized the series of research papers [46, 43, 41, 64], initiated by [44], on the correspondences of certain classes of exchangeable sequences to certain classes of non-negative and non-decreasing càdlàg processes. If one compares the classes of non-negative and non-decreasing càdlàg processes appearing in [43, 41, 64], one observes that none of these

---

[2]We define the univariate geometric distributions as the distribution of the first success in a series of independent Bernoulli trials.

classes is a subclass of the other and that their intersection is the class of Lévy subordinators, whose correspondence with exchangeable sequences was investigated in [44]. Furthermore, the class of random walks with infinitely divisible jumps appearing in [46] has no intersection with any of the classes of non-negative and non-decreasing càdlàg processes appearing in [44, 43, 41, 64]. Thus, since the article [44], the research on the correspondences of non-negative non-decreasing càdlàg processes and exchangeable sequences has developed into several directions. The main contribution of the core article [10] of this thesis is to show that there exists a natural mathematical framework which embeds the articles [44, 46, 43, 41, 64], thereby unifying the literature under one common umbrella.

The key observation is that all non-negative and non-decreasing càdlàg processes H appearing in [44, 46, 43, 41, 64] satisfy the property that for every $n \in \mathbb{N}$ there exist i.i.d. càdlàg processes $\left(H^{(j,1/n)}\right)_{1 \leq j \leq n}$ such that for all $(t_i)_{1 \leq i \leq d} \in \mathbb{R}^d$

$$(H_{t_1}, \ldots, H_{t_d}) \sim \left(\sum_{j=1}^{n} H_{t_1}^{(j,1/n)}, \ldots, \sum_{j=1}^{n} H_{t_d}^{(j,1/n)} \right). \tag{7}$$

Such processes are called infinitely divisible and they have been investigated, among others, in [39, 49, 34, 5, 59]. The most recent work [59] elegantly unifies the literature by providing a modern mathematical description of the analytical characterization of infinitely divisible processes in terms of a so-called Lévy measure ν on the space of càdlàg functions, a deterministic càdlàg function b called drift and a non-negative definite (covariance) function Σ on $\mathbb{R} \times \mathbb{R}$. Since we are only interested in non-negative infinitely divisible processes, one always has $\Sigma = 0$ and the law of a non-negative and non-decreasing infinitely divisible càdlàg process can be analytically described in terms of the family of its d-variate Laplace transforms given by

$$\mathbb{E}\left[\exp\left(-\sum_{i=1}^{d} a_i H_{t_i} \right) \right]$$
$$= \exp\left(-\sum_{i=1}^{d} a_i b(t_i) - \int_{D^{\infty}_{\nearrow, \geq 0}(\mathbb{R})} 1 - \exp\left(-\sum_{i=1}^{d} a_i x(t_i) \right) \nu(dx) \right), \tag{8}$$

where $(a_i)_{1 \leq i \leq d} \in [0, \infty]^d$, b is a non-negative and non-decreasing, ν satisfies $\int_{D^{\infty}_{\nearrow, \geq 0}(\mathbb{R})} |x(t)| \nu(dx) < \infty$ for every $t \in \mathbb{R}$ and $D^{\infty}_{\nearrow, \geq 0}(\mathbb{R})$ denotes the space of non-negative and non-decreasing extended real-valued càdlàg functions. Conversely, for every Lévy measure ν satisfying $\int_{D^{\infty}_{\nearrow, \geq 0}(\mathbb{R})} |x(t)| \nu(dx) < \infty$ for every $t \in \mathbb{R}$ and every non-negative and non-decreasing drift b there exists a non-negative and non-decreasing infinitely divisible càdlàg process H which has corresponding d-variate Laplace transforms of the form (8).

The class of exchangeable sequences associated to non-negative and non-decreasing infinitely divisible càdlàg processes via (1) has multivariate survival functions $\bar{G}_d(t_1, \ldots, t_d) := P(X_1 > t_1, \ldots, X_d > t_d)$ of the form

$$\bar{G}_d(t_1, \ldots, t_d) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{d} H_{t_i} \right) \right]$$
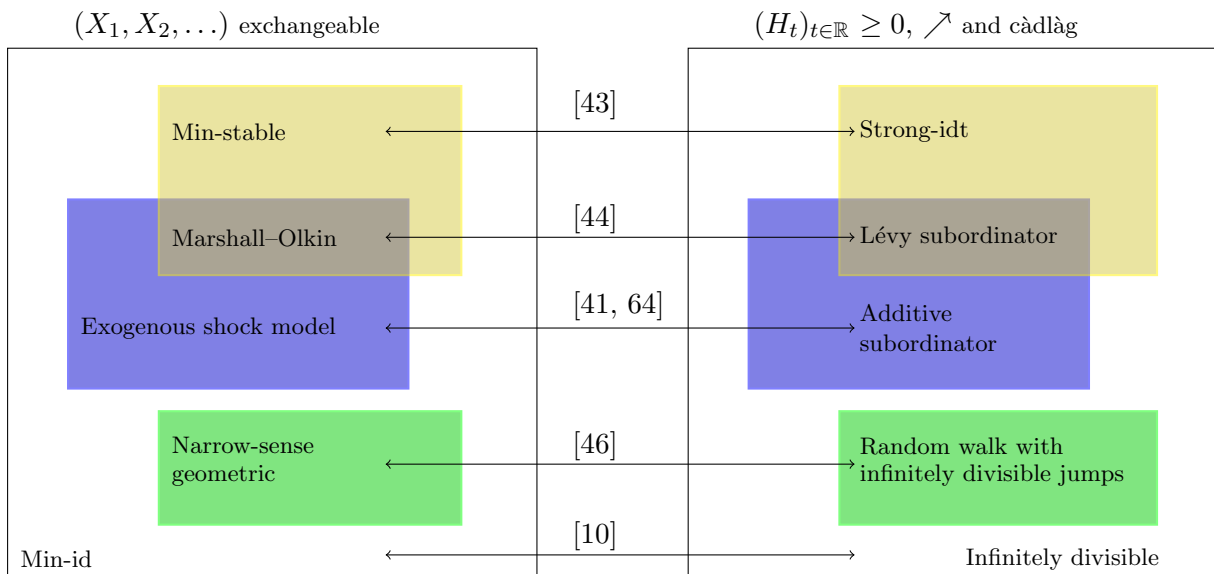
8

Figure 1: Embedding of previously established correspondences of non-negative and non-decreasing càdlàg processes and exchangeable sequences into the framework of [10]. The Figure is adapted from [10, Figure 2].

$$= \exp\left(-\sum_{i=1}^{d} b(t_i) - \int_{D_{\nearrow,\geq 0}^{\infty}(\mathbb{R})} 1 - \exp\left(-\sum_{i=1}^{d} x(t_i)\right)\nu(dx)\right).$$

It is easy to see that, for every $s > 0$, the family of multivariate survival functions $(\bar{G}_d^s)_{d\in\mathbb{N}}$ coincides with the multivariate survival functions of an exchangeable sequence associated to a non-negative and non-decreasing infinitely divisible càdlàg process with drift $sb(\cdot)$ and Lévy measure $s\nu$. Thus, by choosing $s = 1/n$ for some $n \in \mathbb{N}$, the exchangeable sequence associated to the non-negative and non-decreasing infinitely divisible càdlàg processes with drift $b(\cdot)$ and Lévy measure ν has the stochastic representation

$$\boldsymbol{X} \sim \min_{1\leq i\leq n} \boldsymbol{X}^{(i,1/n)} := \left(\min_{1\leq i\leq n} X_1^{(i,1/n)}, \min_{1\leq i\leq n} X_2^{(i,1/n)}, \dots\right), \tag{9}$$

where $\boldsymbol{X}^{(i,1/n)}$ denote i.i.d. copies of an exchangeable sequence associated to a non-negative and non-decreasing infinitely divisible càdlàg processes with drift $b(\cdot)/n$ and Lévy measure ν/n. Due to the stochastic representation (9), the d-dimensional margins of $\boldsymbol{X}$ are called minimum infinitely divisible (min-id) and $\boldsymbol{X}$ is called an exchangeable min-id sequence. Similarly, the d-dimensional marginal distributions of $\boldsymbol{X}$ are called exchangeable min-id distributions. Since all exchangeable sequences investigated in [44, 46, 43, 41, 64] are associated to a non-negative and non-decreasing infinitely divisible càdlàg process, they may be embedded into the unifying probabilistic framework of exchangeable min-id sequences. Figure 1 visualizes the correspondences of exchangeable sequences and non-negative and non-decreasing càdlàg processes that were embedded into the framework of [10].

Min-id distributions naturally appear when investigating minima of independent random vectors. First, the bivariate case was investigated in [4]. Later, under some reasonably mild assumptions,

[32, 3] have shown that when there exist some independent, but not necessarily identically distributed, random vectors $((Z_{i,1}, \ldots, Z_{i,d}))_{i \in \mathbb{N}}$ and sequences $(a_{n,j})_{1 \le j \le d}$ and $(b_{n,j})_{1 \le j \le d} \ge 0$ such that

$$\left(\min_{1 \le i \le n} (Z_{i,1} - a_{n,1})/b_{n,1}, \ldots, \min_{1 \le i \le n} (Z_{i,d} - a_{n,d})/b_{n,d} \right) \xrightarrow{law} (X_1, \ldots, X_d),$$

then $(X_1, \ldots, X_d)$ is min-id. On the other hand, it is easy to see that every min-id distribution can be obtained as the distribution of the weak limit of scaled minima of independent random vectors. Thus, min-id distributions essentially constitute the class of possible limit distributions of scaled minima of independent random vectors. In particular, exchangeable min-id sequences essentially constitute the class of exchangeable limits of scaled minima of independent sequences of random variables.

Building on the results of [4], [67] showed that the survival function of a min-id sequence $\boldsymbol{X}$, i.e. of a sequence of random variables with stochastic representation (9) in terms of minima of i.i.d. sequences of random variables, can be represented as

$$P(X_1 > t_1, X_2 > t_2, \ldots) = \exp\left(-\mu \left((\times_{i \in \mathbb{N}} (t_i, \infty])^{\complement} \right) \right), \tag{10}$$

where μ denotes a unique Radon measure on $[-\infty, \infty]^{\mathbb{N}} \setminus \times_{i \in \mathbb{N}} \{\infty\}$. The measure μ is called exponent measure of $\boldsymbol{X}$ and it is easy to see that when choosing μ as the intensity measure of a Poisson random measure $M := \sum_{i \in \mathbb{N}} \delta_{\boldsymbol{x}_i} := \sum_{i \in \mathbb{N}} \delta_{(x_{i,1}, x_{i,2}, \ldots)}$ on $[-\infty, \infty]^{\mathbb{N}} \setminus \times_{i \in \mathbb{N}} \{\infty\}$ one obtains an alternative stochastic representation of $\boldsymbol{X}$ as

$$\boldsymbol{X} \sim \min_{\boldsymbol{x}_i \in M} \boldsymbol{x}_i = \left(\min_{\boldsymbol{x}_i \in M} x_{i,1}, \min_{\boldsymbol{x}_i \in M} x_{i,2}, \ldots \right), \tag{11}$$

which is often convenient when analyzing statistical properties of $\boldsymbol{X}$. The core article [10] of this thesis provides detailed descriptions of the connections of the stochastic representations (1), (9), and (11) when $\boldsymbol{X}$ is an exchangeable min-id sequence. Moreover, it uniquely connects the Lévy measure and drift of the associated non-negative and non-decreasing infinitely divisible càdlàg process to the exponent measure of $\boldsymbol{X}$.

A particular result that should be emphasized is that [10] shows that the exponent measure of an exchangeable min-id sequence can be decomposed into

$$\mu(\cdot) = \mu_b(\cdot) + \int_{D^\infty_{\nearrow, \ge 0}(\mathbb{R})} \otimes_{i \in \mathbb{N}} P_{1 - \exp(-x(\cdot))}(\cdot) \nu(\mathrm{d}x). \tag{12}$$

where μ_b denotes the exponent measure associated to an i.i.d. sequence with marginal distribution function $1 - \exp(-b(\cdot))$ and $\otimes_{i \in \mathbb{N}} P_{1 - \exp(-x(\cdot))}$ denotes the law of an i.i.d. sequence with marginal distribution function $1 - \exp(-x(\cdot))$. Thus, μ may be decomposed into the exponent measure of an i.i.d. sequence and a mixture of i.i.d. distributions, which may be seen as an extension of de Finetti's Theorem from finite (probability) measures to (possibly non-finite) exponent measures. Figure 2 graphically summarizes the connections of the analytical and probabilistic characterizations of exchangeable min-id sequences provided by [10].
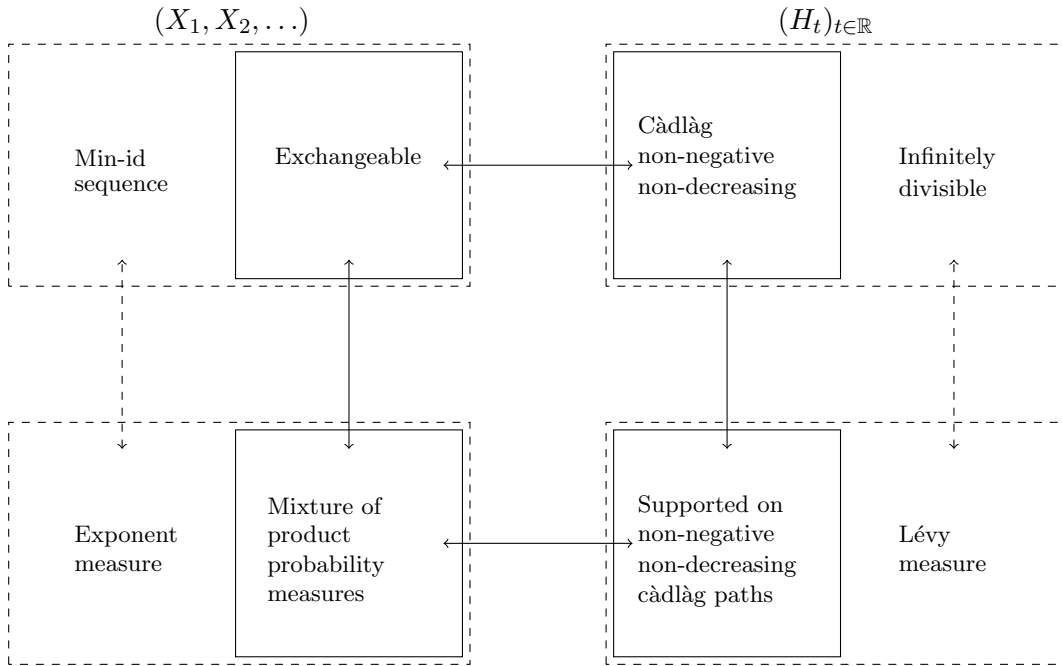
$(X_1, X_2, \ldots)$ $(H_t)_{t \in \mathbb{R}}$

| Min-id sequence | Exchangeable | Càdlàg non-negative non-decreasing | Infinitely divisible |

| Exponent measure | Mixture of product probability measures | Supported on non-negative non-decreasing càdlàg paths | Lévy measure |

Figure 2: Correspondences of min-id sequences (top, left), exponent measures (bottom, left), infinitely divisible processes (top, right), and Lévy measures (bottom, right). The Figure is adapted from [10, Figure 1].

For practical applications it is often necessary to be able to simulate the d-dimensional margins of the exchangeable min-id sequence $\boldsymbol{X}$. To simulate $(X_1, \ldots, X_d)$, one may choose a suitable stochastic representation from (1), (9), or (11), or one may try to directly simulate its d-dimensional margins via standard simulation techniques, e.g. based on the density of $(X_1, \ldots, X_d)$ w.r.t. the Lebesgue measure. Usually, it is not possible to exploit (9) or any standard simulation technique based on the density of $(X_1, \ldots, X_d)$, which might not even exist. Moreover, the simulation of the Poisson random measure from (11) also seems challenging, since there are possibly infinitely many atoms of the Poisson random measure that could determine the realization of $(X_1, \ldots, X_d)$. Therefore, at first sight, the stochastic representation (1) seems to be best suited to simulate the d-dimensional margins of $\boldsymbol{X}$, since it "simply" boils down to simulating the first passage times of the associated non-negative and non-decreasing infinitely divisible càdlàg process H over i.i.d. unit exponential barriers. Clearly, (1) provides an elegant construction method for $\boldsymbol{X}$, where properties of $\boldsymbol{X}$ may be directly inferred from the properties of H or (b, ν). However, even when the Lévy measure and drift of H are known and H is from one of the better known subclasses of infinitely divisible processes, e.g. from the class of Lévy, additive, or strong-idt processes, it is usually quite challenging to even approximately sample a path of H. Nevertheless, it would be desirable to be able to use (1) at least to construct $\boldsymbol{X}$, while having a tool to simulate $\boldsymbol{X}$, provided at least (b, ν), at hand. Exploiting the representation (11), the second core article [8] of this thesis provides a solution to the just described problem by providing an exact simulation algorithm for so-called continuous maximum-infinitely divisible (max-id) processes and viewing $-\boldsymbol{X}$ as a continuous max-id process with index set $\mathbb{N}$ and exponent measure determined by (b, ν).

11

3   Continuous max-id processes

A continuous max-id process $\tilde{\boldsymbol{X}} := \left(\tilde{X}_t\right)_{t \in T}$ is a continuous stochastic process with the property that for every $n \in \mathbb{N}$ there exist i.i.d. stochastic processes $\left(\left(\tilde{X}_t^{(i,1/n)}\right)_{t \in T}\right)_{1 \leq i \leq n}$ such that

$$\tilde{\boldsymbol{X}} \sim \left(\max_{1 \leq i \leq n} \tilde{X}_t^{(i,1/n)}\right)_{t \in T}.$$

[28] were the first to investigate continuous max-id processes, relying on results of [51]. They proved that for every suitably normalized continuous max-id process $\tilde{\boldsymbol{X}}$ with sufficiently regular index set T there exists a Poisson random measure $M := \sum_{i \in \mathbb{N}} \delta_{f_i}$ on the space of continuous functions on T, denoted as $C(T)$, such that $\tilde{\boldsymbol{X}}$ has the stochastic representation

$$\left(\tilde{X}_t\right)_{t \in T} \sim \left(\max_{f_i \in M} f_i(t)\right)_{t \in T}. \tag{13}$$

On the other hand, if the intensity measure $\mu(\cdot) = \mathbb{E}\left[M(\cdot)\right]$ of the Poisson random measure M satisfies certain regularity conditions, they proved that every stochastic process defined via the right hand side of (13) is continuous and max-id. Essentially, this means that a stochastic process is a continuous max-id process if and only if it admits the stochastic representation (13). Again, μ is called the exponent measure of $\tilde{\boldsymbol{X}}$, since the d-dimensional marginal distribution functions of $\tilde{\boldsymbol{X}}$ may be expressed as

$$P\left(\tilde{X}_{t_1} \leq x_1, \ldots, \tilde{X}_{t_d} \leq x_d\right) = \exp\left(-\mu\left(\{f \in C(T) \mid f(t_i) > x_i \text{ for some } 1 \leq i \leq d\}\right)\right).$$

In analogy to min-id distributions, continuous max-id processes arise as weak limits of maxima of (triangular arrays of) independent stochastic processes, which was first proven in [3]. Therefore, they may be seen as the natural extension of continuous max-stable processes, which arise as the possible continuous limits of scaled maxima of i.i.d. stochastic processes and thus are the central object of study in the extreme value theory of i.i.d. continuous stochastic processes. Max-stable processes are well-investigated and highly popular in the modeling of extreme events in space and time, see e.g. [58, 16] for a relatively recent overview about the topic. However, one major flaw of max-stable processes is their limited dependence structure. For example, assuming that all margins of $\tilde{\boldsymbol{X}}$ follow the same distribution, the so-called extremal coefficient $\theta_{t_1,\ldots,t_d}(z) = \log\left(P(\tilde{X}_{t_1} \leq z, \ldots, \tilde{X}_{t_d} \leq z)\right) / \log\left(P(\tilde{X}_{t_1} \leq z)\right)$ is independent of the "level" z, which contradicts the features of many real-world datasets, see, e.g., the datasets used in [31, 11, 2]. On the other hand, due to their richer dependence structure, there exist continuous max-id processes such that $\theta_{t_1,\ldots,t_d}(z)$ is level-dependent, which makes them a suitable candidate for modeling extreme events in space and time when the assumption of max-stability is questionable.

3.1   Simulation of continuous max-id processes

The additional modeling flexibility of max-id processes in comparison to max-stable processes has recently generated new theoretical and practical interest in the topic. For example, [33, 21,

20, 34] have investigated theoretical properties of max-id processes, whereas [53, 6, 31, 70] are concerned with applications of max-id processes in the modeling of spatial extremes. A practical obstacle when modeling a random phenomenon by a max-id process is that their d-dimensional marginal distributions are usually not expressible in a convenient analytical form. Therefore, one must often resort to simulation techniques to capture the properties of a max-id process. However, no general simulation scheme for max-id processes has been available in the literature.

Simulation schemes for the subfamily of max-stable processes are well-known in the literature and, more recently, two exact simulation schemes for continuous max-stable processes have been proposed by [19] and [52]. The algorithm of [52] is based on a certain representation of (parts of) the exponent measure μ of a non-negative max-stable process, which ensures that μ is essentially concentrated on uniformly bounded functions at predefined locations (or compact sets) of interest $(t_i)_{1 \leq i \leq d}$. More precisely, [52] provide a representation of the exponent measure μ which ensures that there exists some $k > 0$ such that the Poisson random measure M with intensity μ may be represented as $\sum_{j \in \mathbb{N}} \delta_{(\zeta_j f_j)}$, where $\zeta_1 \geq \zeta_2, \ldots$ and $f_j(t_i) \leq k$ for all $j \in \mathbb{N}$ and $1 \leq i \leq d$. Then,

$$\left(\tilde{X}_{t_i} \right)_{1 \leq i \leq d} \sim \left(\max_{(\zeta_j f_j) \in M} \zeta_j f_j(t_i) \right)_{1 \leq i \leq d}$$

and $\tilde{\boldsymbol{X}}$ may be simulated exactly at locations $(t_i)_{1 \leq i \leq d}$ by iteratively simulating pairs (ζ_j, f_j) and stopping as soon as $\zeta_n k \leq \max_{1 \leq j \leq n-1} f_j(t_i)$ for all $1 \leq i \leq d$. The derivation of this specific representation of μ relies on a change of measure technique, which cannot directly be extended to continuous max-id processes. Moreover, the technique inherently requires to deviate from the given description of μ and the measure change further depends on the locations $(t_1, \ldots, t_d)$, which are features that could further complicate the extension of the core ideas of [52] to continuous max-id processes.

A different approach is pursued by [19], who provide a simulation algorithm for continuous max-stable processes by only simulating those atoms of the Poisson random measure M which assume the argmax in (13) at least at one of finitely many predefined locations $(t_i)_{1 \leq i \leq d}$. Thus, the goal of their algorithm is to simulate the set of so-called extremal functions at locations $(t_1, \ldots, t_d)$ defined as

$$N^+_{(t_1, \ldots, t_d)} := \bigcup_{1 \leq i \leq d} \operatorname{argmax}_{f_j \in M} f_j(t_i).$$

$\tilde{\boldsymbol{X}}$ is then approximated as

$$\tilde{\boldsymbol{X}} \approx \left(\max_{f \in N^+_{(t_1, \ldots, t_d)}} f(t) \right)_{t \in T},$$

while we have the equality $\tilde{X}_{t_i} = \max_{f \in N^+_{(t_1, \ldots, t_d)}} f(t_i)$ for all $1 \leq i \leq d$. The theoretical foundation of the simulation algorithm of [19] is based on the results of [21], who provide the conditional distribution of $N^-_{(t_1, \ldots, t_n)} := \{f \in C(T) \mid f \text{ atom of } M\} \setminus N^+_{(t_1, \ldots, t_n)}$ given $N^+_{(t_1, \ldots, t_n)}$ for

every $n \in \mathbb{N}$. This decomposition of M into $N^-_{(t_1,\ldots,t_n)}$ and $N^+_{(t_1,\ldots,t_n)}$ allows them to iteratively simulate $\left(N^+_{(t_1,\ldots,t_n)}\right)_{1 \leq n \leq d}$. Moreover, [19] rely on a standardized way of simulating the Poisson random measure M of a continuous max-stable process such that, for fixed $\bar{t} \in T$, one has $f_1(\bar{t}) \geq f_2(\bar{t}) \geq \ldots$. This way of simulating M is based on the regular conditional probability distribution of μ, provided in [20][3], and a change of measure technique. However, the bottleneck of their algorithm is the derivation of the regular conditional probability distribution of μ, which is usually quite challenging.

Even though exact simulation schemes for continuous max-stable processes are available in the literature, none of them had been extended to a general simulation scheme for continuous max-id processes. The core article [8] of this thesis fills this gap in the literature and extends the ideas of [19] to continuous max-id processes. Essentially, the algorithm simulates $N^+_{(t_1,\ldots,t_d)}$ for a predetermined number of locations $(t_1,\ldots,t_d)$ and thus provides an exact simulation algorithm for the d-dimensional margins and a pathwise approximation of a continuous max-id processes. Additionally, it avoids the derivation of the regular conditional probability distribution of μ and allows for non-continuous marginal distributions of $\tilde{\boldsymbol{X}}$. It should be noted that, concurrently and independently of [8], [69] also provided a simulation algorithm for continuous max-id processes with compact index sets and continuous marginal distributions, which is also based on the ideas of [19], but still requires the derivation of the regular conditional probability distribution of μ.

The main ingredient of the simulation algorithm proposed in [8] is the ability to simulate from "finite slices" of the intensity measure μ of the Poisson random measure M of the form $A_{t,k} := \{f \in C(T) \mid f(t) \geq k\}$ for every $k > \operatorname{essinf} \tilde{X}_t$ and $t \in T$. To illustrate this requirement, notice that (11) implies that every exchangeable min-id sequence $\boldsymbol{X}$ can be transformed into a continuous max-id process $-\boldsymbol{X}$ with index set $T = \mathbb{N}$. Assuming that $\boldsymbol{X}$ is associated to a driftless infinitely divisible càdlàg process with Lévy measure ν, (12) implies that the exponent measure μ of $-\boldsymbol{X}$ is a mixture of the law of i.i.d. sequences. Then, for each $i \in \mathbb{N}$, μ restricted to $A_{i,k}$ may be represented as

$$\int_{D^\infty_{\nearrow,\geq 0}(\mathbb{R})} \otimes_{l \in \mathbb{N}} P_{1-\exp(-x(\cdot))}\big((-f(j))_{j \in \mathbb{N}} \in \cdot \mid f(i) \leq -k\big)(1-\exp(-x(-k)))\nu(\mathrm{d}x),$$

where $(1-\exp(-x(-k)))\nu(\mathrm{d}x)$ defines a finite measure. Therefore, to simulate from μ restricted to $A_{i,k}$ one may, in a first step, simulate a function $x_k \sim (1-\exp(-x(-k)))\nu(\mathrm{d}x)/K_k$, where $K_k := \mu(A_{i,k})$ denotes the normalizing constant such that $(1-\exp(-x(-k)))\nu(\mathrm{d}x)/K_k$ becomes a probability measure. Then, in a second step, and conditioned on x_k, simulate an i.i.d. sequence $(f(j))_{j \in \mathbb{N}}$ according to the marginal distribution $1-\exp(-x_k(\cdot))$. Finally, set $f(i) = Z$, where $Z \sim \mathbf{1}_{\{\cdot \leq -k\}}(1-\exp(-x_k(\cdot)))/(1-\exp(-x_k(-k)))$, i.e. Z follows the distribution $1-\exp(-x_k(\cdot))$ conditioned on being less than or equal to $-k$. Then, $(-f(j))_{j \in \mathbb{N}}$ has distribution $\mu(\cdot \cap A_{i,k})/K_k$ and $N \sim Poi(K_k)$ independent draws from $\mu(\cdot \cap A_{i,k})/K_k$ provide a sample of the atoms of a Poisson random measure with intensity $\mu(\cdot \cap A_{i,k})$.

---

[3]See [20, Appendix A2] for the formal definition of the regular conditional probability distribution of an exponent measure.

Clearly, the main difficulty of this procedure is simulating from $(1 - \exp(-x(-k))\nu(\mathrm{d}x)/K_k$. However, the complexity of simulating from $(1 - \exp(-x(-k))\nu(\mathrm{d}x)/K_k$ is usually much lower than the complexity of simulation from $-\boldsymbol{X}$ directly. For example, [8] investigates the particular case of exchangeable exogenous shock models $\boldsymbol{X}$, which correspond to self-similar additive subordinators via (1). The d-dimensional margins of most exchangeable exogenous shock models do not possess a Lebesgue density, which is why standard simulation techniques usually cannot be applied. The elaborations in [8] show that simulation from $(1 - \exp(-x(-k))\nu(\mathrm{d}x)/K_k$ for exchangeable exogenous shock models associated to self-similar additive subordinators essentially boils down to simulating a two-dimensional random vector with known Lebesgue density, which is usually much simpler than simulating $-\boldsymbol{X}$ directly.

4   Model selection with non-smooth criterion functions

For many practical applications, several candidate models for a modeled phenomenon are proposed. Therefore, it has to be decided which model, out of a collection of competing models, is best suited to describe the modeled phenomenon. Usually, the model selection procedure is conducted via pairwise[4] comparisons of the proposed models according to a certain criterion function. The criterion function is commonly determined as an expectation of a measurable function C w.r.t. to the distribution P of the underlying Data Generating Process X, e.g. see [68, 17, 38, 62, 15, 14], and thus may be represented as

$$c(M_1, M_2) := \mathbb{E}\left[\mathsf{C}(M_1(X), M_2(X), X)\right],$$

where $(M_i)_{i=1,2}$ represent two competing models for X. It is common to suitably normalize C such that $c(M_1, M_2) = 0$ may be interpreted as M_1 and M_2 are equally well suited to the modeling problem, $c(M_1, M_2) > 0$ may be interpreted as M_1 is better suited to the modeling problem than M_2 and $c(M_1, M_2) < 0$ may be interpreted as M_2 is better suited to the modeling problem than M_1. For example, $c(M_1, M_2)$ and C may be chosen such that $c(M_1, M_2)$ has the form $c(M_1, M_2) = d(M_2, P) - d(M_1, P)$, where $(d(M_i, P))_{i=1,2}$ denote distances between the candidate probability distributions modeled by $(M_i)_{i=1,2}$ and P. Since $c(M_1, M_2)$ cannot be observed, the researcher needs to rely on a sample $\boldsymbol{X} = (X_1, X_2, \ldots)$ from X to decide which of the models M_1 or M_2 is better suited to describe the modeled phenomenon. A common approach is to conduct a hypothesis test of the form

$$H_0: \ c(M_1, M_2) = 0,$$

which is based on the asymptotic distribution of the empirical counterpart

$$c_n(M_1, M_2) := \frac{1}{n}\sum_{i=1}^{n}\mathsf{C}(M_1(X_i), M_2(X_i), X_i)$$

---

[4]This procedure may not yield a "best" model among all the competing models, but usually the information about how each model compares to the other models is sufficient.

of $c(M_1, M_2)$.

A famous representative of the triplet (C, M_1, M_2) is given by $\left(\log\left(\frac{f_1}{f_2}\right), f_1, f_2\right)$, where f_1 and f_2 denote some proposal densities for the law of X and $c(f_1, f_2) = \mathbb{E}\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right]$ is the expected logarithmic likelihood ratio. The popularity of this specific choice of $c(f_1, f_2)$ stems from the fact that $c(f_1, f_2) = K_p(f_2) - K_p(f_1)$, where $K_p(\cdot)$ denotes the Kullback-Leibler divergence [37] defined as $K_p(h) := \mathbb{E}\left[\log\left(\frac{p(X)}{h(X)}\right)\right]$ and p denotes the true density of X w.r.t. some common dominating measure. Another well-known triplet (C, M_1, M_2) is given by $\left(\left|f_2(X^{(1)}) - X^{(2)}\right|^q - \left|f_1(X^{(1)}) - X^{(2)}\right|^q, f_1, f_2\right)$, where $X = \left(X^{(1)}, X^{(2)}\right)$, $q \geq 1$ and $f_1(X^{(1)})$ and $f_2(X^{(1)})$ denote prediction functions of $X^{(2)}$ given $X^{(1)}$. Then, $c(M_1, M_2) = \mathbb{E}\left[\left|f_2(X^{(1)}) - X^{(2)}\right|^q\right] - \mathbb{E}\left[\left|f_1(X^{(1)}) - X^{(2)}\right|^q\right]$ yields the difference in expected L_q-loss of the predictive models f_1 and f_2.

In most applications, $(M_i)_{i=1,2}$ are represented as parametric families of models, i.e. $M_i = \{f_i(\cdot, \theta_i) \mid \theta_i \in \Theta_i \subset \mathbb{R}^{d_i}\}$ for some fixed functions f_i, while there exists an unknown "optimal"[5] parameter $\theta_i^\star$ for each model M_i. Clearly, a researcher wants to compare M_1 and M_2 in terms of their "optimal" models, but $c(M_1, M_2) := c\left(f_1(\cdot, \theta_1^\star), f_2(\cdot, \theta_2^\star)\right)$ cannot be empirically estimated by $c_n(f_1(\cdot, \theta_1^\star), f_2(\cdot, \theta_2^\star))$ and one must resort to the asymptotic distribution of

$$\hat{c}_n := c_n\left(f_1(\cdot, \hat{\theta}_{1,n}), f_2(\cdot, \hat{\theta}_{2,n})\right)$$

to test H_0, where $(\hat{\theta}_{i,n})_{i=1,2}$ denote consistent estimators of $(\theta_i^\star)_{i=1,2}$.

It is well known that the asymptotic distribution of a statistic with estimated parameters may differ from the asymptotic distribution of the statistic with fixed parameters, see e.g. [65, 22] for early examples of this phenomenon. To illustrate this behavior, consider $c(M_1, M_2) = K_p(f_2(\cdot, \theta_2^\star)) - K_p(f_1(\cdot, \theta_1^\star))$ with its empirical counterpart

$$\hat{c}_n = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f_1(X_i, \hat{\theta}_{1,n})}{f_2(X_i, \hat{\theta}_{2,n})}\right)$$

which is the test statistic of the famous Vuong test [68]. Vuong shows that, under H_0, $\hat{c}_n$ converges either to a weighted sum of Chi-squared distributed random variables with rate n^{-1} or to a normal distribution with rate $n^{-1/2}$, depending on whether or not $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$. In particular, the asymptotic distribution of $n\hat{c}_n$ is, among other things, dependent on the dimension of $(\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$.

In general, the derivation of the asymptotic distribution of $\hat{c}_n$ is traditionally carried out under the assumption that C and $(f_i)_{i=1,2}$ are sufficiently smooth functions of their arguments, since then $\mathsf{C}(f_1(\cdot, \theta_1), f_2(\cdot, \theta_2), \cdot)$ can be approximated by a Taylor expansion of $(\theta_1, \theta_2) \mapsto \mathsf{C}(f_1(\cdot, \theta_1), f_2(\cdot, \theta_2), \cdot)$ around $(\theta_1^\star, \theta_2^\star)$ and one may resort to classical techniques from theoretical statistics to obtain the asymptotic distribution of $\hat{c}_n$. However, when C or $(f_i)_{i=1,2}$ are not differentiable, the asymptotic distribution of $\hat{c}_n$ is considerably more difficult to obtain.

---

[5]The term "optimal" simply refers to a target parameter which is estimated, but it does not need to be truly optimal in any sense.

For example, consider the triplet $(\mathbf{1}_{\{f_1 > f_2\}} - 1/2, f_1, f_2)$, where $f_1(\cdot, \theta_1)$ and $f_1(\cdot, \theta_2)$ denote the logarithm of some parametric candidate densities for the law of X. Clearly, $\mathbf{1}_{\{f_1 > f_2\}} - 1/2$ is not everywhere differentiable and has vanishing derivative whenever it exists. Assuming that $f_1(\cdot, \theta_1^\star) \neq f_2(\cdot, \theta_2^\star)$, the corresponding statistic

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}} - \frac{1}{2}$$

is (equivalent to) the test statistic of the Clarke test, which was introduced in [15, 14]. Clarke's test was motivated by Vuong's test, but he focused on the median of the logarithm of the likelihood ratio instead of its expectation. Besides the fact that, unlike an expectation, the median always exists, considering the median of the logarithm of the likelihood ratio may be motivated as follows. For linear regression models the log-likelihood of an observation $(X^{(1)}, X^{(2)})$ is equal to the L_2-error of the prediction $\tilde{f}_i(X^{(1)}, \theta_i^\star)$ of $X^{(2)}$. Therefore, Clarke's null hypothesis translates to

$$H_0: \quad P\left(\left|\tilde{f}_1(X^{(1)}, \theta_1^\star) - X^{(2)}\right|^2 > \left|\tilde{f}_2(X^{(1)}, \theta_2^\star) - X^{(2)}\right|^2\right) = 1/2$$
$$\Leftrightarrow P\left(\left|\tilde{f}_1(X^{(1)}, \theta_1^\star) - X^{(2)}\right|^q > \left|\tilde{f}_2(X^{(1)}, \theta_2^\star) - X^{(2)}\right|^q \text{ for all } q > 0\right) = 1/2,$$

which may be interpreted as the absence of a structural difference in the error terms of the competing models. In contrast to the more classical null hypothesis that the L_2-loss of the competing models is identical, the null hypothesis of the Clarke test rather means that none of the models more frequently makes predictions that are closer to the truth.

The Clarke test is a drastic example when it comes to the influence of parameter estimation on the asymptotic distribution of the test statistic $\hat{c}_n$. When there is no parameter estimation, it is easy to see that, under H_0,

$$n c_n(f_1(\cdot, \theta_1^\star), f_2(\cdot, \theta_2^\star)) = \sum_{i=1}^{n} \mathbf{1}_{\{f_1(X_i, \theta_1^\star) > f_2(X_i, \theta_2^\star)\}}$$

is Binomial$(n, 1/2)$ distributed for every $n \in \mathbb{N}$. However, when parameter estimation is present,

$$n \hat{c}_n = \sum_{i=1}^{n} \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}}$$

is not (even asymptotically) Binomial distributed under H_0, due to the dependence induced by $\hat{\theta}_{i,n}$ and contrary to the claims in Clarke's articles [15, 14].

The article [9] of this thesis took the example of the Clarke test as a motivation to investigate the asymptotic distribution of

$$\sqrt{n} \hat{c}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{C}\left(f_1(X_i, \hat{\theta}_{1,n}), f_2(X_i, \hat{\theta}_{2,n}), X_i\right)$$

with particular emphasis on non-smooth functions C and $(f_i)_{i=1,2}$. Moreover, the required assumptions on (C, f_1, f_2) are tailored to econometric applications, even though model comparison

based on non-smooth criterion functions may also be interesting for competing Machine Learning models, which are trained with convex non-differentiable loss functions.

Surprisingly, lots of research has been devoted to obtaining M- or Z-estimators of non-smooth criterion functions, see e.g. [1, 12, 54], but there is not much literature about the asymptotic distribution of non-smooth criterion functions with estimated parameters. An early attempt to tackle such problems was introduced by [57], based on ideas of [65], who established the asymptotic normality of $\sqrt{n}\hat{c}_n$ under the assumption that

$$\delta \mapsto \mathbb{E}\left[\sup_{(\theta_1,\theta_2):\ \|(\theta_1,\theta_2)-(\theta_1^\star,\theta_2^\star)\|<\delta} |\mathsf{C}(f_1(X,\theta_1),f_2(X,\theta_2),X) - \mathsf{C}(f_1(X,\theta_1^\star),f_2(X,\theta_2^\star),X)|\right]$$

is Lipschitz-continuous in a neighborhood of 0. The assumption is quite reasonable when C is "almost" smooth, e.g. for continuous non-differentiable functions, but, due to the involved supremum, it seems not very well suited when C has jumps or is an indicator function. Thus, to derive the asymptotic distribution of $\hat{c}_n$ when C is non-continuous with jumps, other techniques need to be applied.

Empirical process theory investigates the asymptotic distribution of the stochastic process $(\mathbb{G}_n h)_{h\in\mathcal{H}} := \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n h(X_i) - \mathbb{E}[h(X)]\right)_{h\in\mathcal{H}}$, where $\mathcal{H}$ denotes a set of functions. One key advantage of empirical process theory for our applications is that the convergence of $(\mathbb{G}_n h)_{h\in\mathcal{H}}$ to a Gaussian process is essentially equivalent to the so-called asymptotic equicontinuity of $(\mathbb{G}_n h)_{h\in\mathcal{H}}$ which is defined as

$$\forall \epsilon > 0:\ \lim_{m\to\infty}\lim_{n\to\infty} P\left(\sup_{h_1,h_2\in\mathcal{H}:\ \|h_1-h_2\|_{L_2(P)}<\frac{1}{m}} |\mathbb{G}_n h_1 - \mathbb{G}_n h_2| > \epsilon\right) = 0.$$

Thus, heuristically speaking, when two functions $h_1, h_2 \in \mathcal{H}$ are "close" in $L_2(P)$, then their limits $\lim_{n\to\infty} \mathbb{G}_n h_1$ and $\lim_{n\to\infty} \mathbb{G}_n h_2$ are "close" with high probability. In our desired application, this allows to define $\mathcal{H} := \{\mathsf{C}(f_1(\cdot,\theta_1), f_2(\cdot,\theta_2), \cdot) \mid (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\}$ and, under certain regularity conditions, to replace the weak limit of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathsf{C}(f_1(X_i,\hat{\theta}_{1,n}), f_2(X_i,\hat{\theta}_{2,n}), X_i)$ with the weak limit of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathsf{C}(f_1(X_i,\theta_1^\star), f_2(X_i,\theta_2^\star), X_i)$ plus an additional bias.

Another advantage of empirical process theory is that the asymptotic Gaussianity of $(\mathbb{G}_n h)_{h\in\mathcal{H}}$ is usually not based on any smoothness conditions on $\mathcal{H}$, but is rather based on the complexity of the class of functions $\mathcal{H}$ in certain L_2 spaces. More precisely, the complexity of $\mathcal{H}$ is measured in terms of its covering or bracketing number[6] in certain L_2-spaces, which essentially can be interpreted as the size of $\mathcal{H}$. For example, when X is real-valued and $\mathcal{H} = \{\mathbf{1}_{\{t\geq\cdot\}} \mid t \in \mathbb{R}\}$, then $(\mathbb{G}_n h)_{h\in\mathcal{H}}$ is a scaled and centered version of the usual empirical distribution function of $(X_1,\ldots,X_n)$ and it may be shown that $\mathcal{H}$ satisfies the complexity bounds that are required for the asymptotic Gaussianity of $(\mathbb{G}_n h)_{h\in\mathcal{H}}$, independently of the law of X. Furthermore, empirical process theory immediately provides the asymptotic law of bootstrapped versions of $(\mathbb{G}_n h)_{h\in\mathcal{H}}$, essentially without strong additional assumptions. Thus, apart from providing results about

---

[6]See [66, Section 2.6 and 2.7] for more details on covering and bracketing numbers.

the asymptotic behaviour of $(\mathbb{G}_n h)_{h \in \mathcal{H}}$, it provides a way of determining the critical values for testing null hypotheses which are based on the asymptotic law of $(\mathbb{G}_n h)_{h \in \mathcal{H}}$.

The article [9] derives the asymptotic distribution of $\sqrt{n} \hat{c}_n$ in the framework of empirical process theory, which is especially well-suited for triplets of parametric non-smooth functions (C, f_1, f_2). In the particular case of the Clarke test it provides conditions which ensure the asymptotic normality of the test statistic

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}} - \frac{1}{2} \right)$$

$$= \mathbb{G}_n \left(\mathbf{1}_{\{f_1(\cdot, \hat{\theta}_{1,n}) > f_2(\cdot, \hat{\theta}_{2,n})\}} \right) + \sqrt{n} \left(\mathbb{E}_X \left[\mathbf{1}_{\{f_1(X, \hat{\theta}_{1,n}) > f_2(X, \hat{\theta}_{2,n})\}} \right] - \frac{1}{2} \right)$$

under the null hypothesis of the Clarke test $H_0: \quad P\left(f_1(X, \theta_1^\star) > f_2(X, \theta_2^\star)\right) = \frac{1}{2}$, assuming that $f_1(\cdot, \theta_1^\star) \neq f_2(\cdot, \theta_2^\star)$. Therefore, it corrects the falsely claimed[7] Binomial distribution of $\sum_{i=1}^n \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}}$ in [15, 14]. Furthermore, it provides the consistency of bootstrapped versions of $\sqrt{n} \hat{c}_n$, which allow to determine the critical values for a test of the null hypothesis $H_0: c(M_1, M_2) = 0$.

5 Outlook

The aim of this section is to identify and roughly sketch potential further research topics and remaining open questions, which are closely related to the contributions of the articles [9, 10, 8] constituting the main content of this thesis.

First, the article [10] embeds the research articles [44, 46, 43, 41, 64] in a unifying framework and provides some explicit examples of exchangeable min-id sequences and non-negative and non-decreasing infinitely divisible processes outside the realm of the previously known correspondences. For example, exchangeable min-id sequences with Archimedean copula [50] and log-completely monotone generator as well as exchangeable min-id sequences with reciprocal Archimedean copula [27] are embedded into the framework of [10]. Moreover, exchangeable min-id sequences with finite exponent measure and exchangeable min-id sequences associated to Lévy processes subordinated by an integrated non-negative infinitely divisible càdlàg process are investigated. However, a thorough analysis of statistical and analytical properties of these sequences and their respective copulas is missing. For example, [23] provide examples of non-negative infinitely divisible processes, including their respective Lévy measures, which may serve as a basic tool to build non-negative and non-decreasing infinitely divisible càdlàg processes and their associated exchangeable min-id sequences. Another potentially promising family of exchangeable min-id sequences is the class of exchangeable min-id sequences associated to Poisson cluster processes, since they are usually neither additive nor strong-idt processes and admit a

---

[7][9] also proves that the asymptotic variance of $\sqrt{n} \left(n^{-1} \sum_{i=1}^n \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}} - 1/2 \right)$ may deviate from $1/4$, which shows that $\sum_{i=1}^n \mathbf{1}_{\{f_1(X_i, \hat{\theta}_{1,n}) > f_2(X_i, \hat{\theta}_{2,n})\}}$ is not even asymptotically Binomial distributed.

closed-form representation of the survival function of the associated exchangeable min-id sequence. The goal of such investigations should be to derive convenient analytical and stochastic representations of the considered sequences such that simulation of the d-dimensional margins of these sequences is feasible, e.g. via the algorithm proposed in [8]. Potential applications of these specific exchangeable min-id distributions may be found in the modeling of dependence structures of extreme events, where the assumption of min(or max)-stability may not be reasonable for the data at hand.

Second, the correspondence of exchangeable min-id sequences and non-negative and non-decreasing càdlàg processes has thus far only been used to construct sequences of exchangeable random variables with certain statistical properties determined by a given non-negative and non-decreasing infinitely divisible càdlàg process. However, one may also use this correspondence to construct non-negative and non-decreasing infinitely divisible càdlàg processes by specifying the law of an exchangeable min-id sequence. For practical applications it is usually necessary to not only specify the law of the non-negative and non-decreasing infinitely divisible càdlàg process, but also to be able to simulate this process. By observing that the empirical survival function $\left(\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{X_i>t\}}\right)_{t\in\mathbb{R}}$ of a realization of an exchangeable min-id sequence may be transformed into a realization of the associated infinitely divisible process $\left(-\log\left(\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{X_i>t\}}\right)\right)_{t\in\mathbb{R}}$, one immediately obtains an (approximate) simulation scheme of the associated infinitely divisible process if one is able to simulate the (finite-dimensional margins of the) associated exchangeable min-id sequence, e.g. via the simulation algorithm proposed in [8]. Moreover, since the empirical distribution function is a well-investigated object, this approximate simulation algorithm for infinitely divisible processes may also come with guarantees on the approximation quality in finite sample sizes. In most of the academic literature, simulation algorithms for infinitely divisible processes are based on certain series representations of the process, but there are generally no guarantees on the approximation quality and on the feasibility of the simulation of the relevant terms of the series representation. Thus, the simulation algorithm for non-negative and non-decreasing infinitely divisible stochastic processes via exchangeable min-id sequences would provide a completely new approach for the simulation of non-negative and non-decreasing infinitely divisible stochastic processes, which might be advantageous in certain applications.

Third, one could extend the framework of exchangeable sequences of random variables whose d-dimensional marginal distributions are min-id to exchangeable sequences of d'-dimensional random vectors whose multivariate marginal distributions on $\mathbb{R}^{d'\times d}$ are min-id for all $d \in \mathbb{N}$, when viewed as a random field. The logarithm of the corresponding random multivariate survival function, whose existence is ensured by [30], is then presumably infinitely divisible, which is consistent with every margin being associated to a non-negative and non-decreasing infinitely divisible processes. The analytical properties of the random multivariate survival function may be derived from the frameworks of [56, 59]. However, it is not obvious if the associated random survival function should be viewed as a multivariate infinitely divisible process, as an infinitely

divisible random measure or as an infinitely divisible process with multivariate index set. On the one hand, such an extension would be valuable from a theoretical point of view, since it nicely embeds the univariate theory into a multivariate context. On the other hand, such ideas may have applications in non-parametric Bayesian inference about the unknown distribution of a random vector, similar to additive priors on the unknown distribution function of a random variable as proposed in [24, 18, 25]. Roughly speaking, the d'-dimensional random survival function associated to the exchangeable sequence of random vectors would then act as a prior on the space of d'-dimensional survival functions.

Fourth, the literature on max-id processes, which are not max-stable processes, is rather scarce. With the simulation algorithm [8] at hand, one may construct new parametric families of continuous max-id processes which are suitable for simulation and satisfy desirable statistical and analytical properties. Moreover, fitting such models to empirical data is rather complicated, since the full likelihood (if it even exists) of the data is often not tractable. However, there are modern statistical approaches whose model fitting procedures solely require the ability to simulate the candidate models and do not require independent input data, such as Maximum-Mean-Discrepancy based estimation [13]. Applying such approaches to max-id or max-stable models comes with solid statistical consistency guarantees and may lower the computational burden in comparison to likelihood-based methods, as stochastic gradient descent approaches may be applied. Furthermore, if necessary, one may extend the simulation algorithm for continuous max-id processes presented in [8] to upper semi-continuous max-id processes. However, this would come at the cost of carefully rechecking many proofs of preliminary results by [21, 20].

Fifth, the derivations of the asymptotic distribution of the model selection test proposed in [9] are mainly based on the asymptotic normality of the empirical process. Such asymptotic normality results are not limited to independent observations of the data, but, e.g., also available for α- or β-mixing data. Therefore, the model selection test proposed in [9] may be extended to dependent data. Furthermore, the null hypothesis of the Clarke test $H_0 : P(f_1(X, \theta_1^\star) > f_2(X, \theta_2^\star)) = 1/2$ is obviously not satisfied when $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$, which is why [9] proposed to test the modified Clarke null hypothesis $\tilde{H}_0 : P(f_1(X, \theta_1^\star) > f_2(X, \theta_2^\star)) = P(f_1(X, \theta_1^\star) < f_2(X, \theta_2^\star))$ when it cannot be excluded that $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$. However, when $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$, the map $(\theta_1, \theta_2) \mapsto P(f_1(X, \theta_1) > f_2(X, \theta_2))$ may become highly irregular in a neighborhood of $(\theta_1^\star, \theta_2^\star)$ and the assumptions of the stated model selection test in [9] may fail to hold. Therefore, it would be desirable to find a regularization mechanism, possibly similar to the ideas of [62, 60], which ensures that a modified version of the test statistic proposed in [9] is asymptotically normal even when $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$.

References

[1] Donald W.K. Andrews. "Asymptotics for semiparametric econometric models via stochastic equicontinuity". In: *Econometrica* 62.1 (1994), pp. 43–72. DOI: https://doi.org/10.2307/2951475.

[2] Jean-Noël Bacro et al. "Hierarchical Space-Time Modeling of Asymptotically Independent Exceedances With an Application to Precipitation Data". In: *Journal of the American Statistical Association* 115.530 (2020), pp. 555–569. DOI: 10.1080/01621459.2019.1617152.

[3] August A. Balkema, Laurens de Haan, and Rajeeva L. Karandikar. "Asymptotic Distribution of the Maximum of n Independent Stochastic Processes". In: *Journal of Applied Probability* 30.1 (1993), pp. 66–81. DOI: https://doi.org/10.2307/3214622.

[4] August A. Balkema and Sindey I. Resnick. "Max-infinite divisibility". In: *Journal of Applied Probability* 14.2 (1977), pp. 309–319. DOI: 10.2307/3213001.

[5] Ole E. Barndorff–Nielsen, Makoto Maejima, and Ken I. Sato. "Infinite Divisibility for Stochastic Processes and Time Change". In: *Journal of Theoretical Probability* 19 (2006), pp. 411–446. DOI: https://doi.org/10.1007/s10959-006-0020-7.

[6] Gregory P. Bopp, Benjamin A. Shaby, and Raphaël Huser. "A Hierarchical Max-Infinitely Divisible Spatial Model for Extreme Precipitation". In: *Journal of the American Statistical Association* 116.533 (2021), pp. 93–106. DOI: 10.1080/01621459.2020.1750414.

[7] Florian Brück. "Clarke's Test For Non-Nested Model Comparison". Masterarbeit. Technische Universität München, 2019. URL: https://mediatum.ub.tum.de/1472565.

[8] Florian Brück. "Exact simulation of continuous max-id processes with applications to exchangeable max-id sequences". In: *Journal of Multivariate Analysis* 193 (2023). DOI: https://doi.org/10.1016/j.jmva.2022.105117.

[9] Florian Brück, Jean-David Fermanian, and Aleksey Min. "A corrected Clarke test for model selection and beyond". In: *Journal of Econometrics* (2022). DOI: https://doi.org/10.1016/j.jeconom.2021.12.013.

[10] Florian Brück, Jan-Frederik Mai, and Matthias Scherer. "Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes". In: *Extremes* (2022). DOI: https://doi.org/10.1007/s10687-022-00450-w.

[11] Daniela Castro-Camilo, Linda Mhalla, and Thomas Opitz. "Bayesian space-time gap filling for inference on extreme hot-spots: an application to Red Sea surface temperatures". In: *Extremes* 24.1 (2021), pp. 105–128. DOI: https://doi.org/10.1007/s10687-020-00394-z.

[12] Xiaohong Chen, Oliver Linton, and Ingrid Van Keilegom. "Estimation of semiparametric models when the criterion function is not smooth". In: *Econometrica* 71.5 (2003), pp. 1591–1608. URL: https://www.jstor.org/stable/1555514.

[13] Badr-Eddine Chérief-Abdellatif and Pierre Alquier. "Finite sample properties of parametric MMD estimation: Robustness to misspecification and dependence". In: *Bernoulli* 28.1 (2022), pp. 181–213. DOI: `10.3150/21-BEJ1338`.

[14] Kevin A. Clarke. "A simple distribution-free test for nonnested model selection". In: *Political Analysis* 15.3 (2007), pp. 347–363. DOI: `doi:10.1093/pan/mpm004`.

[15] Kevin A. Clarke. "Nonparametric model discrimination in international relations". In: *Journal of Conflict Resolution* 47.1 (2003), pp. 72–93. DOI: `10.1177/0022002702239512`.

[16] Dipak K. Dey and Jun Yan. *Extreme value modeling and risk analysis: methods and applications*. CRC Press, 2016. DOI: `https://doi.org/10.1201/b19721`.

[17] Francis X. Diebold and Robert S. Mariano. "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3 (1995), pp. 253–263. DOI: `https://doi.org/10.1198/073500102753410444`.

[18] Kjell Doksum. "Tailfree and Neutral Random Probabilities and Their Posterior Distributions". In: *The Annals of Probability* 2.2 (1974), pp. 183–201. DOI: `10.1214/aop/1176996703`.

[19] Clément Dombry, Sebastian Engelke, and Marco Oesting. "Exact simulation of max-stable processes". In: *Biometrika* 103.2 (2016), pp. 303–317. DOI: `10.1093/biomet/asw008`.

[20] Clément Dombry and Frédéric Eyi-Minko. "Regular conditional distributions of continuous max-infinitely divisible random fields". In: *Electronic Journal of Probability* 18 (2013), pp. 1–21. DOI: `10.1214/EJP.v18-1991`.

[21] Clément Dombry and Frédéric Eyi-Minko. "Strong mixing properties of max-infinitely divisible random fields". In: *Stochastic Processes and their Applications* 122.11 (2012), pp. 3790–3811. DOI: `https://doi.org/10.1016/j.spa.2012.06.013`.

[22] James Durbin. "Weak Convergence of the Sample Distribution Function when Parameters are Estimated". In: *The Annals of Statistics* 1.2 (1973), pp. 279–290. DOI: `10.1214/aos/1176342365`.

[23] Nathalie Eisenbaum and Jan Rosiński. *Lévy measures of infinitely divisible positive processes – examples and distributional identities*. 2021. DOI: `10.48550/ARXIV.2106.10175`.

[24] Thomas S. Ferguson. "A Bayesian Analysis of Some Nonparametric Problems". In: *The Annals of Statistics* 1.2 (1973), pp. 209–230. DOI: `10.1214/aos/1176342360`.

[25] Thomas S. Ferguson. "Prior Distributions on Spaces of Probability Measures". In: *The Annals of Statistics* 2.4 (1974), pp. 615–629. DOI: `10.1214/aos/1176342752`.

[26] Bruno de Finetti. "Funzione caratteristica di un fenomeno aleatorio". In: *Atti del Congresso Internazionale dei Matematici, Bologna 3-10 Settembre* Tomo VI (1928), pp. 179–190.

[27] Christian Genest, Johanna G. Nešlehová, and Louis-Paul Rivest. "The class of multivariate max-id copulas with ℓ_1-norm symmetric exponent measure". In: *Bernoulli* 24.4B (2018), pp. 3751–3790. DOI: `10.3150/17-BEJ977`.

[28] Evarist Giné, Marjorie G. Hahn, and Pirooz Vatan. "Max-infinitely divisible and max-stable sample continuous processes". In: *Probability theory and related fields* 87.2 (1990), pp. 139–165. DOI: `https://doi.org/10.1007/BF01198427`.

[29] Alexander V. Gnedin. "On a class of exchangeable sequences". In: *Statistics & Probability Letters* 25.4 (1995), pp. 351–355. DOI: `https://doi.org/10.1016/0167-7152(94)00240-3`.

[30] Edwin Hewitt and Leonard J. Savage. "Symmetric measures on Cartesian products". In: *Transactions of the American Mathematical Society* 80.2 (1955), pp. 470–501. DOI: `https://doi.org/10.1090/S0002-9947-1955-0076206-8`.

[31] Raphaël Huser, Thomas Opitz, and Emeric Thibaud. "Max-infinitely divisible models and inference for spatial extremes". In: *Scandinavian Journal of Statistics* 48.1 (2021), pp. 321–348. DOI: `https://doi.org/10.1111/sjos.12491`.

[32] Jürg Hüsler. "Limit properties for multivariate extreme values in sequences of independent, non-identically distributed random vectors". In: *Stochastic Processes and their Applications* 31.1 (1989), pp. 105–116. DOI: `https://doi.org/10.1016/0304-4149(89)90105-1`.

[33] Zakhar Kabluchko and Martin Schlather. "Ergodic properties of max-infinitely divisible processes". In: *Stochastic Processes and their Applications* 120.3 (2010), pp. 281–295. DOI: `https://doi.org/10.1016/j.spa.2009.12.002`.

[34] Zakhar Kabluchko and Stilian Stoev. "Stochastic integral representations and classification of sum- and max-infinitely divisible processes". In: *Bernoulli* 22.1 (2016), pp. 107–142. DOI: `10.3150/14-BEJ624`.

[35] Alexander J. Khintchine. "Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze". In: *Matematicheskii Sbornik* 44.1 (1937), pp. 79–119.

[36] Clark H. Kimberling. "A probabilistic interpretation of complete monotonicity". In: *Aequationes Mathematicae* 10 (1974), pp. 152–164. DOI: `https://doi.org/10.1007/BF01832852`.

[37] Solomon Kullback and Richard A. Leibler. "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. DOI: `10.1214/aoms/1177729694`.

[38] Pascal Lavergne and Quang H. Vuong. "Nonparametric Selection of Regressors: The Nonnested Case". In: *Econometrica* 64.1 (1996), pp. 207–219. DOI: `https://doi.org/10.2307/2171929`.

[39] Peter M. Lee. "Infinitely divisible stochastic processes". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 7.2 (1967), pp. 147–160. DOI: `https://doi.org/10.1007/BF00536328`.

[40] Paul Lévy. *Théorie de l'Addition des Variables Aléatoires*. Vol. 1. Monographies des Probabilités / publiés sous la direction de Emile Borel. Gauthier-Villars, 1937.

[41] Jan-Frederik Mai, Steffen Schenk, and Matthias Scherer. "Exchangeable exogenous shock models". In: *Bernoulli* 22 (2016), pp. 1278–1299. DOI: 10.3150/14-BEJ693.

[42] Jan-Frederik Mai, Steffen Schenk, and Matthias Scherer. "Two novel characterizations of self-decomposability on the half-line". In: *Journal of Theoretical Probability* 30.1 (2017), pp. 365–383. DOI: https://doi.org/10.1007/s10959-015-0644-6.

[43] Jan-Frederik Mai and Matthias Scherer. "Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time". In: *Extremes* 17.1 (2014), pp. 77–95. DOI: https://doi.org/10.1007/s10687-013-0175-4.

[44] Jan-Frederik Mai and Matthias Scherer. "Lévy-frailty copulas". In: *Journal of Multivariate Analysis* 100.7 (2009), pp. 1567–1585. DOI: https://doi.org/10.1016/j.jmva.2009.01.010.

[45] Jan-Frederik Mai and Matthias Scherer. "What makes dependence modeling challenging? Pitfalls and ways to circumvent them". In: *Statistics & Risk Modeling* 30.4 (2013), pp. 287–306. DOI: doi:10.1524/strm.2013.2001.

[46] Jan-Frederik Mai, Matthias Scherer, and Natalia Shenkman. "Multivariate geometric distributions, (logarithmically) monotone sequences, and infinitely divisible laws". In: *Journal of Multivariate Analysis* 115 (2013), pp. 457–480. DOI: https://doi.org/10.1016/j.jmva.2012.11.012.

[47] Albert W. Marshall and Ingram Olkin. "A Multivariate Exponential Distribution". In: *Journal of the American Statistical Association* 62.317 (1967), pp. 30–44. DOI: 10.1080/01621459.1967.10482885.

[48] Albert W. Marshall and Ingram Olkin. "Families of Multivariate Distributions". In: *Journal of the American Statistical Association* 83.403 (1988), pp. 834–841. DOI: https://doi.org/10.2307/2289314.

[49] Gisiro Maruyama. "Infinitely divisible processes". In: *Theory of Probability & Its Applications* 15.1 (1970), pp. 1–22. DOI: https://doi.org/10.1137/1115001.

[50] Alexander J. McNeil and Johanna Nešlehová. "Multivariate Archimedean Copulas, d-Monotone Functions and ℓ_2-Norm Symmetric Distributions". In: *The Annals of Statistics* 37.5B (2009), pp. 3059–3097. DOI: 10.1214/07-AOS556.

[51] Tommy Norberg. *On the existence and convergence of probability measures on continuous semi-lattices*. Tech. rep. University of North Carolina, 1986. URL: https://apps.dtic.mil/sti/citations/ADA179267.

[52] Marco Oesting, Martin Schlather, and Chen Zhou. "Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation". In: *Bernoulli* 24.2 (2018), pp. 1497–1530. DOI: 10.3150/16-BEJ905.

[53] Simone A. Padoan. "Extreme dependence models based on event magnitude". In: *Journal of Multivariate Analysis* 122 (2013), pp. 1–19. DOI: `https://doi.org/10.1016/j.jmva.2013.07.009`.

[54] Ariel Pakes and David Pollard. "Simulation and the asymptotics of optimization estimators". In: *Econometrica* 57.5 (1989), pp. 1027–1057. DOI: `https://doi.org/10.2307/1913622`.

[55] S. T. Rachev and L. Rüschendorf. "Approximate Independence of Distributions on Spheres and Their Stability Properties". In: *The Annals of Probability* 19.3 (1991), pp. 1311–1337. DOI: `10.1214/aop/1176990346`.

[56] Balram S. Rajput and Jan Rosiński. "Spectral representations of infinitely divisible processes". In: *Probability Theory and Related Fields* 82.3 (1989), pp. 451–487. DOI: `https://doi.org/10.1007/BF00339998`.

[57] Ronald H. Randles. "On the Asymptotic Normality of Statistics with Estimated Parameters". In: *The Annals of Statistics* 10.2 (1982), pp. 462–474. DOI: `10.1214/aos/1176345787`.

[58] Mathieu Ribatet. "Spatial extremes: Max-stable processes at work". In: *Journal de la Société Française de Statistique* 154.2 (2013), pp. 156–177. URL: `http://www.numdam.org/item/JSFS_2013__154_2_156_0/`.

[59] Jan Rosiński. "Representations and isomorphism identities for infinitely divisible processes". In: *The Annals of Probability* 46.6 (2018), pp. 3229–3274. DOI: `10.1214/17-AOP1246`.

[60] Susanne M. Schennach and Daniel Wilhelm. "A simple parametric model selection test". In: *Journal of the American Statistical Association* 112.520 (2017), pp. 1663–1674. DOI: `https://doi.org/10.1080/01621459.2016.1224716`.

[61] Isaac J. Schoenberg. "Metric spaces and positive definite functions". In: *Transactions of the American Mathematical Society* 44.3 (1938), pp. 522–536. DOI: `https://doi.org/10.2307/1989894`.

[62] Xiaoxia Shi. "A nondegenerate Vuong test". In: *Quantitative Economics* 6.1 (2015), pp. 85–121. DOI: `10.3982/QE382`.

[63] Abe Sklar. "Fonctions de répartition à n dimensions et leurs marges". In: *Publications de l'Institut de Statistique de l'Université de Paris* 8 (1959), pp. 229–231.

[64] Henrik Sloot. "The deFinetti representation of generalised Marshall–Olkin sequences". In: *Dependence Modeling* 8.1 (2020), pp. 107–118. DOI: `https://doi.org/10.1515/demo-2020-0006`.

[65] Balkrishna V. Sukhatme. "Testing the Hypothesis That Two Populations Differ Only in Location". In: *The Annals of Mathematical Statistics* 29.1 (1958), pp. 60–78. DOI: `10.1214/aoms/1177706706`.

[66] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, 1996. DOI: https://doi.org/10.1007/978-1-4757-2545-2.

[67] Pirooz Vatan. "Max-infinite divisibility and max-stability in infinite dimensions". In: *Probability in Banach Spaces V*. Springer Berlin Heidelberg, 1985, pp. 400–425. ISBN: 978-3-540-39645-1.

[68] Quang H. Vuong. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses". In: *Econometrica* 57.2 (1989), pp. 307–333. DOI: https://doi.org/10.2307/1912557.

[69] Peng Zhong, Raphaël Huser, and Thomas Opitz. "Exact Simulation of Max-Infinitely Divisible Processes". In: *Econometrics and Statistics* (2022). DOI: https://doi.org/10.1016/j.ecosta.2022.02.007.

[70] Peng Zhong, Raphaël Huser, and Thomas Opitz. "Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude". In: *The Annals of Applied Statistics* 16.1 (2022), pp. 272–299. DOI: 10.1214/21-AOAS1504.

A   Core Publications

A.1   Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes [10]

Summary

The main result of this paper is a one-to-one correspondence of exchangeable min-id sequences and non-negative and non-decreasing infinitely divisible càdlàg processes. More precisely, we show that every exchangeable min-id sequence may be constructed via

$$\boldsymbol{X} := \Big(\inf \Big\{ t \in \mathbb{R} \mid H(t) \geq E_i \Big\} \Big)_{i \in \mathbb{N}}, \tag{14}$$

where $(H_t)_{t \in \mathbb{R}}$ denotes a non-negative and non-decreasing infinitely divisible càdlàg process and $(E_i)_{i \in \mathbb{N}}$ denotes a sequence of independent and identically distributed exponential random variables with mean 1. Furthermore, we show that each particular choice of (the law of) the non-negative and non-decreasing infinitely divisible càdlàg processes $(H_t)_{t \in \mathbb{R}}$ in (14) yields a unique (in distribution) exchangeable min-id sequence $\boldsymbol{X}$. Thus, the law of an exchangeable min-id sequence is in one-to-one correspondence to the law of a non-negative and non-decreasing infinitely divisible càdlàg process.

Based on this correspondence, we deduce that the exponent measure of the exchangeable min-id sequence $\boldsymbol{X}$ is in one-to-one correspondence to the (unique) tuple (ν, b) of Lévy measure and drift of the associated non-negative and non-decreasing infinitely divisible càdlàg process H. We infer that the exponent measure of an exchangeable min-id sequence can be decomposed into the sum of the exponent measure of an i.i.d. sequence with marginal survival function $\big(\exp(-b(t)) \big)_{t \in \mathbb{R}}$ and a mixture of the law of an i.i.d. sequence of the form

$$\int_{D_{\nearrow, \geq 0}^{\infty}(\mathbb{R})} \otimes_{i \in \mathbb{N}} P_{1-\exp(-x(\cdot))} \nu(\mathrm{d}x), \tag{15}$$

where $\otimes_{i \in \mathbb{N}} P_{1-\exp(-x(\cdot))}$ denotes the law of an i.i.d. sequence with marginal distribution function $\big(1 - \exp(-x(t)) \big)_{t \in \mathbb{R}}$ and $D_{\nearrow, \geq 0}^{\infty}(\mathbb{R})$ denotes the space of non-negative and non-decreasing extended real-valued càdlàg functions. This result may be viewed as an extension of de Finetti's seminal Theorem [26] from finite (probability) measures to sigma-finite exponent measures.

Our results allow to unify several preceding research articles under one common umbrella. To be specific, starting with the article [44], which proved that the exchangeable sequences constructed in (14) when H is a non-negative and non-decreasing Lévy process have finite-dimensional marginal distribution of the Marshall–Olkin kind, the literature has developed into two "separate" directions. First, [43] proved that all exchangeable sequences with extreme value copulas and exponential margins can be obtained if H in (14) is a strong-idt process. Second, [41] and [64] characterized the class of exchangeable sequences that are obtained via (14) when H is an additive process as so-called exogenous shock models. Since neither the class of strong-idt processes

nor the class of additive processes is a superclass of the other and their intersection is precisely the class of Lévy processes, there was no common mathematical framework for the results of [43, 41, 64]. However, strong-idt processes and additive processes are specific subfamilies of the class of infinitely divisible processes. Therefore, we have unified the work of [43, 41, 64] in a common mathematical framework. Additionally, particular classes of exchangeable sequences which could not be associated to any of the previously investigated classes of exchangeable sequences may now be embedded into our framework. For example, when the stochastic process H in (14) is a random walk with infinitely divisible jumps, the corresponding multivariate marginal distributions of the associated exchangeable sequences were identified as members of the so-called narrow sense multivariate geometric distributions by [46]. Further examples include exchangeable sequences with reciprocal Archimedean copula [27] as well as exchangeable sequences with Archimedean copula and log-completely monotone generator [50].

To provide a mathematically sound framework, we also contribute to the study of non-negative and non-decreasing infinitely divisible càdlàg processes. We prove that the Lévy measures of such processes resemble their path properties as they are shown to be concentrated on non-negative and non-decreasing càdlàg functions as well. Moreover, we show that each infinitely divisible càdlàg process may be represented as the sum of arbitrary many i.i.d. infinitely divisible càdlàg processes, which was formulated as an open problem in [5]. Another side result is that we have proven that a min-id random vector is exchangeable if and only if its associated finite-dimensional exponent measure is exchangeable.

Individual contributions

Jan-Frederik Mai has conjectured that exchangeable min-id sequences are in one-to-one correspondence to non-negative and non-decreasing infinitely divisible càdlàg processes and provided an idea to prove this correspondence. The derivations and detailed formulations of the mathematical results and proofs in the paper were worked out by myself. The families of exchangeable sequences which were embedded into our framework were mostly identified by Jan-Frederik Mai and Matthias Scherer.

I wrote the first version of the entire manuscript. All authors carefully edited the work in several iterations, including the requested changes in the resubmission process. Jan-Frederik Mai and Matthias Scherer particularly modified the introduction and the examples section that embeds many existing families of exchangeable sequences into the new framework.

I am the main author of this article.

Permission to include the article

Springer   SpringerLink shop ⊞

Subjects   **Services**   **About Us**

Permissions

Get permission to reuse Springer Nature content

Springer Nature is partnered with the Copyright Clearance Center to meet our customers' licensing and permissions needs.

Copyright Clearance Center's RightsLink® service makes it faster and easier to secure permission for the reuse of Springer Nature content to be published, for example, in a journal/magazine, book/textbook, coursepack, thesis/dissertation, annual report, newspaper, training materials, presentation/slide kit, promotional material, etc.

Simply visit SpringerLink and locate the desired content;

- Go to the article or chapter page you wish to reuse content from. (Note: permissions are granted on the article or chapter level, not on the book or journal level). Scroll to the botton of the page, or locate via the side bar, the "Reprints and Permissions" link at the end of the chapter or article.
- Select the way you would like to reuse the content;
- Complete the form with details on your intended reuse. Please be as complete and specific as possible ao as not to delay your permission request;
- Create an account if you haven't already. A RightsLink account is different than a SpringerLink account, and is necessary to receive a licence regardless of the permission fee. You will receive your licence via the email attached to your RightsLink receipt;
- Accept the terms and conditions and you're done!

For questions about using the RightsLink service, please contact Customer Support at Copyright Clearance Center via phone +1-855-239-3415 or +1-978-646-2777 or email springernaturesupport@copyright.com.

How to obtain permission to reuse Springer Nature content not available online on SpringerLink

Requests for permission to reuse content (e.g. figure or table, abstract, text excerpts) from Springer Nature publications currently not available online must be submitted in writing. Please be as detailed and specific as possible about what, where, how much, and why you wish to reuse the content.

Your contacts to obtain permission for the reuse of material from:

- books: bookpermissions@springernature.com
- journals: journalpermissions@springernature.com

Author reuse

Please check the Copyright Transfer Statement (CTS) or Licence to Publish (LTP) that you have signed with Springer Nature to find further information about the reuse of your content.

Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published article in their thesis according to current citation standards. Material from: 'AUTHOR, TITLE, JOURNAL TITLE, published [YEAR], [publisher - as it appears on our copyright page]'

If you are any doubt about whether your intended re-use is covered, please contact journalpermissions@springernature.com for confirmation.

Self-Archiving

- Journal authors retain the right to self-archive the final accepted version of their manuscript. Please see our self-archiving policy for full details:
https://www.springer.com/gp/open-access/authors-rights/self-archiving-policy/2124

- Book authors please refer to the information on this link:
https://www.springer.com/gp/open-access/publication-policies/self-archiving-policy

---

Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes

Florian Brück[1] · **Jan-Frederik Mai**[2] · **Matthias Scherer**[1]

Abstract

We establish a one-to-one correspondence between (i) exchangeable sequences of random variables whose finite-dimensional distributions are minimum (or maximum) infinitely divisible and (ii) non-negative, non-decreasing, infinitely divisible stochastic processes. The exponent measure of an exchangeable minimum infinitely divisible sequence is shown to be the sum of a very simple "drift measure" and a mixture of product probability measures, which uniquely corresponds to the Lévy measure of a non-negative and non-decreasing infinitely divisible process. The latter is shown to be supported on non-negative and non-decreasing functions. In probabilistic terms, the aforementioned infinitely divisible process is equal to the conditional cumulative hazard process associated with the exchangeable sequence of random variables with minimum (or maximum) infinitely divisible marginals. Our results provide an analytic umbrella which embeds the de Finetti subfamilies of many interesting classes of multivariate distributions, such as exogenous shock models, exponential and geometric laws with lack-of-memory property, min-stable multivariate exponential and extreme-value distributions, as well as reciprocal Archimedean copulas with completely monotone generator and Archimedean copulas with log-completely monotone generator.

✉ Florian Brück
  florian.brueck@tum.de

  Jan-Frederik Mai
  Jan-Frederik.Mai@xaia.com

  Matthias Scherer
  scherer@tum.de

1 Lehrstuhl für Finanzmathematik, Technische Universität München, Parkring 11, Garching 857480, Germany

2 XAIA Investment GmbH, München, Germany

 Springer

1 Introduction

The present article bridges the gap between two well-established theories: exchangeable sequences of minimum/maximum infinitely divisible (min-id/max-id) random variables on the one hand and non-negative and non-decreasing (nnnd) infinitely divisible stochastic processes on the other hand. So far, these topics are studied in (mostly) separate communities and we seek to address both. On the one hand, our results can be understood as a particular application of the theory of infinitely divisible (id) processes. General id-processes are studied in Barndorff-Nielsen et al. (2006), Kabluchko and Stoev (2016), Rosiński (2018), related literature concerned with specific subfamilies comprises Skorohod (1991), Bertoin (1999), Sato (1999), Mansuy (2005), Hakassou and Ouknine (2012, 2013), Kopp and Molchanov (2018), Mai and Scherer (2019). The article Rosiński (2018) unifies the literature by establishing a general analytical apparatus to deal with id-processes by means of a Lévy measure on the path space $\mathbb{R}^{\mathbb{R}}$. We refine these results by restricting our attention to processes with nnnd càdlàg paths, but remain totally general aside from this assumption. On the other hand, finite-dimensional probability distributions that are max- (or min-)id naturally arise as limit laws of suitably scaled maxima of independent random vectors, see Hüsler (1989), a textbook account being Resnick (2007). Such distributions and prominent subfamilies, like max- (or min-) stable laws, are well-established in the applied probability and statistics literature, see e.g. Balkema and Resnick (1977), Marshall and Olkin (1990), Alzaid and Proschan (1994), Joe and Hu (1996), Mulinacci (2015), Genest et al. (2018), and have recently gained interest in the modeling of spatial extremes, see e.g. Padoan (2013), Huser et al. (2018), Bopp et al. (2020), Huser et al. (2021). In analytical terms, such probability distributions are canonically described by a so-called exponent measure and the work of Vatan (1985) generalizes this framework to infinite sequences of random variables.

In our article, we apply de Finetti's Theorem, see de Finetti (1931), (Aldous 1985, Chapter 1.3) to derive a correspondence between nnnd càdlàg id-processes and exchangeable sequences of random variables, all of whose finite-dimensional distributions are min- (or max-)id. In fact, given an nnnd càdlàg id-process $H = (H_t)_{t \in \mathbb{R}}$, we may define an infinite exchangeable sequence $X = (X_1, X_2, \ldots)$ of min-id random variables via the almost sure relation

$$\mathbb{P}\left(\bigcap_{i \in \mathbb{N}} \{X_i > t_i\} \,\middle|\, H \right) = \prod_{i \in \mathbb{N}} \exp\left(-H_{t_i}\right), \ (t_1, t_2, \ldots) \in [-\infty, \infty)^{\mathbb{N}}. \tag{1}$$

The sequence $X = (X_1, X_2, \ldots)$ is a conditionally (on H) i.i.d. sequence with conditional marginal survival function $t \mapsto \exp(-H_t)$,[1] where $H_{-\infty}$ is defined as the almost sure limit $\lim_{t \to -\infty} H_t$. While we may plug in arbitrary nnnd càdlàg id-processes H on the right-hand side of Eq. (1), we prove that one actually obtains each exchangeable min-id sequence X on the left-hand side of Eq. (1) via this construction, i.e. we prove a one-to-one correspondence between exchangeable min-id sequences and nnnd càdlàg id-processes.

---

[1] This follows from (1) by choosing $(t_1, t_2, \ldots)$ such that all but one t_i are equal to $-\infty$.

The assumption of (infinite) exchangeability often appears in practical applications when the overall complexity of the model needs to be limited. For example, if a high-dimensional multivariate phenomenon cannot be modeled by i.i.d. components, but further information about the dependence structure of the margins is missing, it is often reasonable to assume exchangeability of the components, as it provides an appealing scheme for analytical computation and simulation. The assumption of exchangeability is convenient in such situations, since it preserves flexibility between independence and full dependence. Moreover, even when there are legitimate reasons to believe that the modeled margins are not exchangeable, it is often reasonable to divide the modeled components into homogeneous subgroups, which are intrinsically exchangeable. Building a hierarchical model which has exchangeable subgroups allows to maintain a quite simple dependence structure inside these subgroups, while the non-exchangeable dependence structure between subgroups remains tractable.

A particular advantage of min- (or max-)id distributions in the modeling of extreme events is their flexible dependence structure. For example, the so-called extremal coefficient, which measures the distance of the model to the model of independent components at a fixed threshold, can actually be chosen to be threshold-dependent, a feature that is often discovered in real world data sets, see e.g. Huser et al. (2021). On the other hand, if one would resort to classical min-stable models, such features cannot be modeled, since the extremal coefficient of min-stable models with identical margins is known to be independent of the chosen threshold. Moreover, going from min-stability to min-infinite divisibility comes essentially without additional mathematical technicalities, while min-id models easily incorporate arbitrary marginal distributions in contrast to the restricted flexibility of marginal distributions in min-stable models.

Regarding the analytical treatments of H and X, we characterize the Lévy measures of nnnd càdlàg id-processes H as precisely those which are concentrated on nnnd paths and characterize the exponent measures of exchangeable min-id sequences X as precisely those that are the sum of some simple "drift measure" and a (possibly infinite) mixture of product probability measures. Figure 1 summarizes our findings in a nutshell.

For many subfamilies of id-processes, there exist well-established theories and applications on the stochastic process level. The relation between the nnnd instances of these subfamilies with the multivariate probability laws of X via de Finetti's Theorem has been explored in several previous articles, which are unified and extended by the present work. Firstly, if H is a non-decreasing Lévy process (aka Lévy subordinator), the finite-dimensional distributions of the corresponding sequence X are so-called Marshall-Olkin distributions, a result first found in Mai and Scherer (2011), re-discovered and further explored in Sun et al. (2017). Secondly and slightly more general, if H is an nnnd additive process (aka additive subordinator), the survival function defined by $t \mapsto \exp(-H_t)$ is called a neutral-to-the-right prior in non-parametric Bayesian statistics. The resulting non-parametric Bayesian estimation techniques are explored, e.g., in Kalbfleisch (1978), Hjort (1990), Epifani et al. (2003), Regazzini et al. (2003), James et al. (2009), James (2005), with the most prominent representative being the Dirichlet process developed by Ferguson (1973). The finite-dimensional distributions of the associated exchangeable sequence X are shown to correspond to exogenous shock models in Mai et al. (2016), Sloot (2020). A special case of particular interest is obtained in case H is a Sato subordinator, leading to a characterization
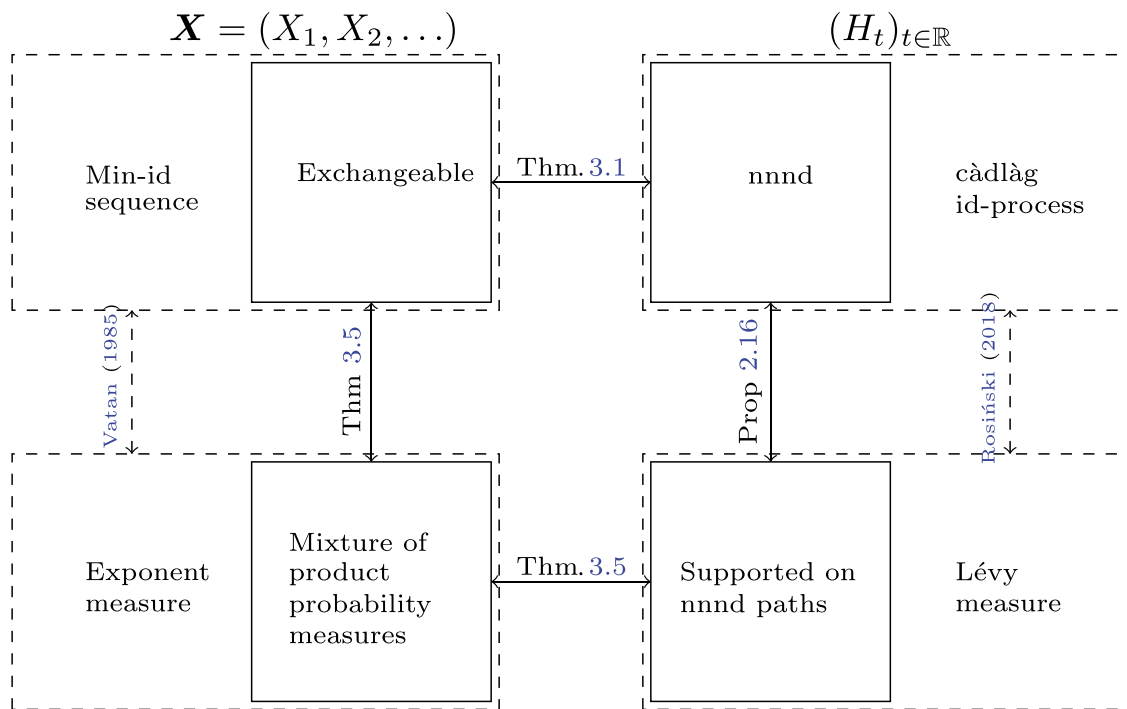
Fig. 1 Correspondences of exchangeable min-id sequences (top, left), exponent measures (bottom, left), non-decreasing and non-negative càdlàg id-processes (top, right), and Lévy measures (bottom, right)

of self-decomposability on the half-line in terms of multivariate distribution functions in Mai et al. (2017). Thirdly, if H is non-decreasing and strongly infinitely divisible with respect to time (aka time stable/strong-idt), see e.g. Mansuy (2005), Hakassou and Ouknine (2013), Kopp and Molchanov (2018), the finite-dimensional distributions of X are shown to be min-stable in Mai and Scherer (2014), Mai (2020). Figure 2 summarizes the correspondences.
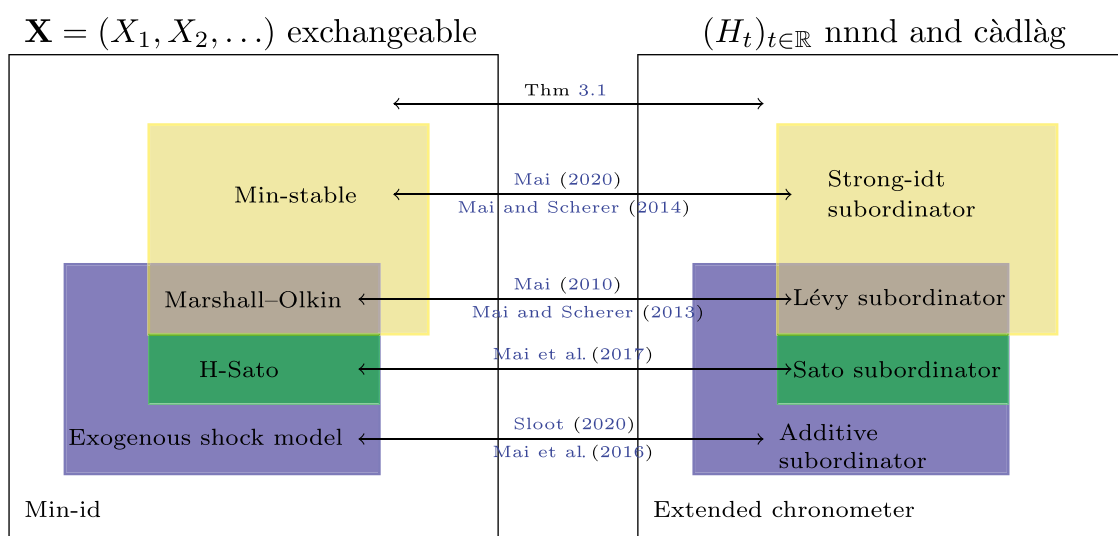
Fig. 2 Embedding of established correspondences of nnnd càdlàg processes and exchangeable sequences into the present framework

It is worth noting that our framework is general enough to include non-continuous min-id distributions, which correspond to id-processes H that are not stochastically continuous. While this stands in glaring contrast to most of the aforementioned references and might on first glimpse be accompanied by technical problems, our derivations show that a distinction between stochastically continuous and non-continuous processes is not a crucial technical obstacle, but rather a distinctive feature to study after a general theory is established.

1.1 Structure of the manuscript

In Sect. 2, we recall the most important results about min-id distributions and characterize their exchangeable subclass. Furthermore, we characterize nnnd infinitely divisible stochastic processes by their Lévy measure and drift. As a byproduct, we show that every càdlàg id-process can be represented as the sum of i.i.d. càdlàg processes, which solves an open problem posed by Barndorff-Nielsen et al. (2006). Section 3 provides the main contributions of this work. First, we show that each exchangeable min-id sequence uniquely corresponds to an nnnd infinitely divisible càdlàg process. Second, we show that the exponent measure of an exchangeable min-id sequence is given by the sum of a simple drift measure and a mixture of product probability measures. In Sect. 4 we present some important examples of exchangeable min-id sequences and embed them into our framework. A summary of the key findings is given in Sect. 5. All longer proofs are deferred to Appendix 1.

2 Preliminaries

2.1 Notation

The following notation is used throughout this paper. A topological space (S, τ) is always equipped with its Borel σ-algebra, denoted by $\mathcal{B}(S)$. Upon existence, we frequently use the notation 0_S to refer to the neutral element w.r.t. addition in S. The letters $\mathbb{R}, \overline{\mathbb{R}}, \mathbb{Q}, \mathbb{N}$, and $\mathbb{N}_0$ denote the real numbers, (extended) real numbers including $\pm\infty$, rational numbers, natural numbers, and natural numbers including 0, all equipped with their standard topologies. For $A \subset \overline{\mathbb{R}}$, $A^{\mathbb{N}}$ denotes the space of A-valued sequences equipped with the product (subspace-)topology. Vectors and sequences are written in bold letters to distinguish them from scalars. The symbols $>, \geq, <,$ and $\leq$ are understood componentwise, e.g. for $\boldsymbol{x}, \boldsymbol{y} \in \overline{\mathbb{R}}^d$, $\boldsymbol{x} < \boldsymbol{y}$ means that $x_i < y_i$ for all $1 \leq i \leq d$. Similarly, for every $\boldsymbol{x}, \boldsymbol{y} \in \overline{\mathbb{R}}^d$, the operators $\max$ (resp. $\min$)$\{\boldsymbol{x}, \boldsymbol{y}\}$ are applied componentwise. For every set $A \subset \overline{\mathbb{R}}^d$ we define $\max$ (resp. $\min$, $\sup$, $\inf$) A as the componentwise maximum (resp. minimum, supremum, infimum) of elements in A, where $\min \emptyset := \inf \emptyset := \infty$ and $\max \emptyset := \sup \emptyset := -\infty$. Moreover, for any $\boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{R}}^d$ we define $[\boldsymbol{a}, \boldsymbol{b}] := \times_{i=1}^d [a_i, b_i]$, where $[a_i, b_i]$ denotes a closed interval. The obvious modifications apply to $(\boldsymbol{a}, \boldsymbol{b}), [\boldsymbol{a}, \boldsymbol{b})$, and $[\boldsymbol{a}, \boldsymbol{b}]$. The function $\mathbf{1}_A(\boldsymbol{x})$ denotes the indicator function of a set A.

$D(T)$ (resp. $D^\infty(T)$) denotes the space of real-valued (resp. extended real-valued) càdlàg functions, i.e. right-continuous functions with left limits, which are indexed by a set $T \subset \mathbb{R}$. $D^\infty(T)$ is always equipped with the (Borel) σ-algebra generated by the finite dimensional

projections, i.e. $\mathcal{B}(D^\infty(T)) := \sigma\big(\big\{\{x \in D^\infty(T) \mid (x(t_1), \dots, x(t_d)) \in A\}, (t_i)_{1 \le i \le d} \subset T,$ $A \in \mathcal{B}(\overline{\mathbb{R}}^d), d \in \mathbb{N}\big\}\big)$. Note that this σ-algebra is a Borel σ-algebra, since it can be generated as the Borel σ-algebra of a topology on $D^\infty(T)$. Similarly, $D(T)$ is equipped with the subspace (Borel) σ-algebra $\mathcal{B}(D(T)) := \mathcal{B}(D^\infty(T)) \cap D(T)$. The function $0_{D^\infty(T)} := (0)_{t \in T}$ denotes the function which vanishes everywhere. A càdlàg process indexed by T denotes a random element $H \in D^\infty(T)$. Sometimes H is also referred to as $(H_t)_{t \in T}$ to emphasize the stochastic process character. H_t refers to the extended real-valued random variable obtained by projecting H at "time" t.

We write $X \sim Y$ to denote that two random elements X and Y are identical in distribution, even though X and Y do not need to be defined on the same probability space. If the probability space is not explicitly specified, we adopt the usual notation and denote the probability measure as $\mathbb{P}$. The distribution function F of a random vector $X_d \in \overline{\mathbb{R}}^d$ is defined as $F(x) := \mathbb{P}(X_d \le x)$. The survival function $\overline{F}$ of X_d is defined as $\overline{F}(x) := \mathbb{P}(X_d \in \times_{i=1}^d \{x_i, \infty])$, where "{" is interpreted as "(" if $x_i > -\infty$ and "{" is interpreted as "[" if $x_i = -\infty$. We frequently write $X_d \sim F$ (resp. $X_d \sim \overline{F}$) to denote that the random vector X_d has distribution (resp. survival) function F (resp. $\overline{F}$). A random variable has exponential distribution with mean $1/\lambda \ge 0$ if it has survival function $\overline{F}(x) = \exp\big(-\lambda x \mathbf{1}_{[0,\infty)}(x)\big)$. The terms min- (resp. max-) id random vector and min- (resp. max-)id distribution will be used synonymously, depending on whether we refer to a random vector or its associated distribution.

In slight abuse of the common terminology we say that a measure μ is supported on a set $A \subset \mathcal{B}\big(\overline{\mathbb{R}}^d\big)$ if $\mu\big(\overline{\mathbb{R}}^d \setminus A\big) = 0$.

2.2 Exchangeable min- and max-id distributions

First, let us recall the definition of exchangeable and extendible random vectors and exchangeable sequences.

Definition 2.1 (Exchangeable and extendible random vector/sequence)

1. A random vector $X_d = (X_{d,1}, \dots, X_{d,d}) \in \overline{\mathbb{R}}^d$ is exchangeable if $X_d \sim \big(X_{d,\pi(1)}, \dots, X_{d,\pi(d)}\big)$ for all permutations π on $\{1, \dots, d\}$. A random sequence $X = \big(X_i\big)_{i \in \mathbb{N}} \in \overline{\mathbb{R}}^\mathbb{N}$ is exchangeable if all its finite dimensional marginal distributions are exchangeable.
2. An exchangeable random vector $X_d = (X_{d,1}, \dots, X_{d,d}) \in \overline{\mathbb{R}}^d$ is extendible if there exists an exchangeable sequence $X = \big(X_i\big)_{i \in \mathbb{N}} \in \overline{\mathbb{R}}^\mathbb{N}$ (possibly defined on a different probability space) satisfying $X_d \sim (X_1, \dots, X_d)$.

Extendible random vectors form a proper subclass of exchangeable random vectors, since there exist many exchangeable random vectors which are not extendible. For example, consider a random vector $X_2 \in \mathbb{R}^2$ which follows a bivariate normal distribution with negative correlation.[2] (Aldous 1985, p. 7) shows that exchangeable

---

[2] Another example being an Archimedean copula with a d-monotone generator that is not completely monotone.

sequences of random variables necessarily have non-negative correlation. Therefore, X_2 is exchangeable but not extendible to an exchangeable sequence. It is also worth noting that an extendible random vector X may be extendible to more than one exchangeable sequence. On the other side, de Finetti's Theorem provides a unique stochastic representation of exchangeable sequences and thus also characterizes extendible random vectors.

Theorem 2.2 (De Finetti's Theorem, De Finetti (1931))

A sequence $X = (X_i)_{i \in \mathbb{N}} \in \overline{\mathbb{R}}^{\mathbb{N}}$ is exchangeable if and only if there exists an nnnd stochastic process $H \in D^\infty(\mathbb{R})$ such that

$$(X_i)_{i \in \mathbb{N}} \sim \left(\inf\{t \in \mathbb{R} \mid H_t \geq E_i\} \right)_{i \in \mathbb{N}}, \tag{2}$$

where $(E_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. unit exponential random variables independent of H. X has survival function

$$\mathbb{P}\left(X_1 > t_1, X_2 > t_2, \dots\right) = \mathbb{E}\left[\prod_{i \in \mathbb{N}} e^{-H_{t_i}} \right], \quad t = (t_1, t_2, \dots) \in [-\infty, \infty)^{\mathbb{N}}.$$

Moreover, the distribution of the process H is uniquely determined by the distribution of X and vice versa.

De Finetti's Theorem can be refined for exchangeable sequences with continuous marginal distribution.

Corollary 2.3 (Exchangeable sequences with continuous marginal distribution correspond to stochastically continuous càdlàg processes)

The sequence $X \in \overline{\mathbb{R}}^{\mathbb{N}}$ from Eq. (2) has continuous marginal distributions if and only if H is stochastically continuous.

Proof The proof can be found in Appendix 1. □

Next, we recall the most important results about min- and max-id distributions. We mainly follow (Resnick 2007, Chapter 5) and translate the results from max-id random vectors on $\mathbb{R}^d$ to min-id random vectors on $(-\infty, \infty]^d$. Many of these translations are straightforward, but some lurking technical subtleties need to be considered and are emphasized in the upcoming paragraphs. We start with the formal definition of min- (resp. max-)id random vectors/sequences. There are at least two equivalent definitions that will be used frequently throughout the paper. Therefore, they are presented jointly.

Definition 2.4 (Min-id distribution / Max-id distribution)

1. A random vector $X_d \in \overline{\mathbb{R}}^d$ is minimum-infinitely divisible (min-id) if for every $n \in \mathbb{N}$ there exist i.i.d. random vectors $(\underline{X}_d^{(i,1/n)})_{1 \leq i \leq n}$[3] such that $X_d \underset{d}{\sim} \min_{1 \leq i \leq n} \underline{X}_d^{(i,1/n)}$. Equivalently, the survival function $\overline{F}$ of a random vector $X_d \in \overline{\mathbb{R}}^d$ is min-id if $\overline{F}^t$ is a survival function for every $t > 0$.

2. A random vector $X_d \in \overline{\mathbb{R}}^d$ is maximum-infinitely divisible (max-id) if for every $n \in \mathbb{N}$ there exist i.i.d. random vectors $(X_d^{(i,1/n)})_{1 \leq i \leq n}$ such that $X_d \underset{d}{\sim} \max_{1 \leq i \leq n} X_d^{(i,1/n)}$. Equivalently, the distribution function F of a random vector $X_d \in \overline{\mathbb{R}}^d$ is max-id if F^t is a distribution function for every $t > 0$.

Similarly, a sequence $X \in \overline{\mathbb{R}}^{\mathbb{N}}$ is called min- (resp. max-)id if $(X_{i_1}, \ldots, X_{i_d})$ is a min- (resp. max-)id random vector for every $(i_1, \ldots, i_d) \in \mathbb{N}^d$.

Every univariate random variable $X \in \overline{\mathbb{R}}$ is min- and max-id. However, already for $d \geq 2$ it is certainly not trivial to decide whether a given random vector or survival (resp. distribution) function is min- (resp. max-)id. For example, a bivariate normal distribution with negative correlation is not min- (resp. max-)id, since min- (resp. max-)id random vectors necessarily exhibit non-negative correlation (Resnick 2007, Proposition 5.29).

An important property of the class of min- and max-id distributions is their distributional closure under monotone marginal transformations. This fact is summarized in the following lemma, which is a slight extension of (Resnick 2007, Proposition 5.2 iii).

Lemma 2.5 (Class of min- and max-id distributions is closed under monotone transformations)

Let X_d (resp. Y_d) $\in \overline{\mathbb{R}}^d$ denote a min- (resp. max-)id random vector. The following statements are valid.

1. *Let $(f_i)_{1 \leq i \leq d} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be non-decreasing. Then $f(X_d) := (f_1(X_1), \ldots, f_d(X_d))$ (resp. $f(Y_d) := (f_1(Y_1), \ldots, f_d(Y_d))$) is min- (resp. max-)id.*
2. *Let $(f_i)_{1 \leq i \leq d} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be non-increasing. Then $f(X_d)$ (resp. $f(Y_d)$) is max- (resp. min-)id.*

Lemma 2.5 shows that studying the entire class of min- and max-id distributions is equivalent to studying min-id distributions supported on an arbitrary non-empty subset of $A \subset \overline{\mathbb{R}}^d$, since for every min- or max-id random vector X_d there exist some strictly monotone functions $(f_i)_{1 \leq i \leq d}$ such that $f(X_d) \in A$ is min-id. A convenient choice for our analysis is the set $A = (-\infty, \infty]^d$. This choice allows us to assume that a min-id distribution does not have mass on $\{x \in \overline{\mathbb{R}}^d \mid x_i = -\infty \text{ for some } 1 \leq i \leq d\}$. Such min-id distributions are uniquely determined by their survival function restricted to $\mathbb{R}^d$. Thus, we can avoid the

---

[3] The notation $X_d^{(i,1/n)}$ shall emphasize that every $X_d^{(i,1/n)}$ can be interpreted as a $1/n$ contribution to X, since the $X_d^{(i,1/n)}$ are all equally likely to establish the minimum in $\min_{1 \leq i \leq n} X_d^{(i,1/n)} \sim X_d$.

technical subtleties involving survival functions on $\overline{\mathbb{R}}^d$. The convenience of this seemingly artificial fact will become clear in Sect. 3.1, since the choice $X \in (-\infty, \infty]^{\mathbb{N}}$ allows us to restrict our attention to càdlàg processes vanishing at $-\infty$. Therefore, without loss of generality, we only consider min-id random vectors on $(-\infty, \infty]^d$ in the remainder of the paper, if not explicitly mentioned otherwise.

Similar to (Resnick 2007, Proposition 5.8), we can characterize min-id distributions on $(-\infty, \infty]^d$ by a so-called exponent measure. To define the exponent measure of a min-id distribution we need to introduce some notation. For $\boldsymbol{\ell} \in (-\infty, \infty]^d$ define the set $E_{\boldsymbol{\ell}}^d := [-\infty, \boldsymbol{\ell}] \setminus \{\boldsymbol{\ell}\}$, which is equipped with the subspace topology inherited from $\overline{\mathbb{R}}^d$. Furthermore, for $\boldsymbol{x} < \boldsymbol{\ell}$, define $(\boldsymbol{x}, \infty]^{\complement} := E_{\boldsymbol{\ell}}^d \setminus (\boldsymbol{x}, \infty]$.

Definition 2.6 (Exponent measure of min-id distributions)

A Radon measure μ_d on $E_{\boldsymbol{\ell}}^d$ is called exponent measure if it satisfies

$$\mu_d \left(\bigcup_{i=1}^d \{\boldsymbol{x} \in E_{\boldsymbol{\ell}}^d \mid x_i = -\infty\} \right) = 0.$$

This definition of an exponent measure ensures that $X_d \sim \exp\left(-\mu_d\left((\cdot, \infty]^{\complement}\right)\right)$ is min-id on $(-\infty, \infty]^d$, since

$$\mathbb{P}(X_i = -\infty \text{ for some } 1 \le i \le d)$$

$$= 1 - \exp\left(-\mu_d\left(\bigcup_{i=1}^d \{\boldsymbol{x} \in E_{\boldsymbol{\ell}}^d \mid x_i = -\infty\}\right)\right) = 0.$$

The next proposition, which is similar to (Resnick 2007, Proposition 5.8), shows that every min-id random vector has a survival function of the form $\exp\left(-\mu_d(\cdot, \infty]^{\complement}\right)$.

Proposition 2.7 (Characterization of min-id distributions)

The following are equivalent:
1. $X_d \in (-\infty, \infty]^d$ *is min-id.*
2. *There exist $\boldsymbol{\ell} \in (-\infty, \infty]^d$ and an exponent measure μ_d defined on $E_{\boldsymbol{\ell}}^d$ such that*

$$X_d \sim \overline{F}(\boldsymbol{x}) = \begin{cases} \exp\left(-\mu_d\left((\boldsymbol{x}, \infty]^{\complement}\right)\right) & \boldsymbol{x} < \boldsymbol{\ell}, \\ 0 & else. \end{cases}$$

Moreover, the exponent measure μ_d associated with X_d is unique.
Proof The proof is a translation of the proof of (Resnick 2007, Proposition 5.8) to the min-id case. $\square$

We have already mentioned that it is sufficient to restrict our study of min-id $X_d \in \overline{\mathbb{R}}^d$ to that of min-id $X_d \in (-\infty, \infty]^d$. Since $E_{\boldsymbol{\ell}}^d$ is almost rectangular, we can restrict the study of min-id distributions on $(-\infty, \infty]^d$ to an even smaller and

more convenient class of min-id distributions, because Lemma 2.5 allows to transform every min-id random vector X_d to a min-id random vector $\tilde{X}_d$ which satisfies $\ell = \times_{i=1}^d \{\infty\}$. Thus, from now on, we always assume that a min-id random vector X_d satisfies the following condition, if not explicitly mentioned otherwise.

Condition ($\Diamond$) A min-id random vector $X_d \in \overline{\mathbb{R}}^d$ satisfies Condition ($\Diamond$) if the following two assumptions are satisfied
($\Diamond$1) $X_d \in (-\infty, \infty]^d$ and
($\Diamond$2) $\ell = \infty$.

The purpose of Condition ($\Diamond$) is to simplify our exposition in the remainder of the paper. It ensures that we can always assume that a min-id random vector X_d satisfies $\mathbb{P}(X_d > x) > 0$ for all $x \in \mathbb{R}^d$ and $\mathbb{P}(X_i > -\infty) = 1$ for all $1 \leq i \leq d$. These constraints allow to avoid splitting the proofs of our main theorems into several cases.

Up to this point we have mainly introduced and reformulated existing results. In the following, we characterize exchangeable and extendible min-id random vectors, which have not yet been discussed in the literature.

A natural question is whether the exchangeability (resp. extendibility) of the min-id random vector X_d is equivalent to the exchangeability (resp. extendibility) of its associated exponent measure μ_d. To this purpose, let us properly define exchangeability of an exponent measure. For any $x \in \overline{\mathbb{R}}^d$ and permutation π on $\{1, \ldots, d\}$ define $\pi(x) = (x_{\pi(1)}, \ldots, x_{\pi(d)})$. Similarly, for any set $A \subset \overline{\mathbb{R}}^d$, define $\pi(A) := \{\pi(x) \mid x \in A\}$. An exponent measure μ_d on E_∞^d is called exchangeable if $\mu_d(A) = \mu_d(\pi(A))$ for every $A \in \mathcal{B}(E_\infty^d)$ and every permutation π on $\{1, \ldots, d\}$.

Proposition 2.8 (Exchangeable exponent measure)

The following are equivalent for every min-id random vector $X_d \in (-\infty, \infty]^d$:
1. *X_d is exchangeable.*
2. *μ_d is exchangeable.*

It may seem obvious that μ_d is exchangeable if X_d is exchangeable, since Proposition 2.7 implies exchangeability of μ_d on sets of the type $(x, \infty]^\complement$. However, μ_d may have infinite mass and exchangeability on sets of the type $(x, \infty]^\complement$ is not a sufficient criterion for exchangeability of infinite measures in general.

Proof The proof can be found in Appendix 1. $\qquad\square$

Proposition 2.8 shows that exchangeability of X_d is in one-to-one correspondence with exchangeability of μ_d. A similar statement holds for extendibility of X_d, but the precise formulation of this result is tedious. The problem arises from the definition of E_∞^d, since E_∞^d is not a product space and this may lead to projections to the point $\infty \notin E_\infty^{d'}$, $d' < d$. Vatan has investigated this problem in the context of exponent measures of max-id sequences in Vatan (1985). Unfortunately, his definition of an exponent measure slightly differs from the original definition in Resnick

(2007) and does not directly translate to Definition 2.6. Vatan puts infinite mass on the lower boundary $-\boldsymbol{\ell}$ of the support of the exponent measure of a max-id sequence $-\boldsymbol{X}$ to ensure that exponent measures are projective. Under this constraint he proved the existence of a unique projective exponent measure on $[-\boldsymbol{\ell}, \infty] := \times_{i \in \mathbb{N}}[-\ell_i, \infty]$ associated with $-\boldsymbol{X}$. Removing the point $-\boldsymbol{\ell}$ from the support of Vatan's exponent measure allows us to translate his results to our setting and we obtain the exponent measure of a min-id sequence $\boldsymbol{X}$. Note that this modified exponent measure is generally not projective, which is due to the removal of $-\boldsymbol{\ell}$. We denote this global exponent measure on $E_{\boldsymbol{\ell}}^{\mathbb{N}} := [-\infty, \boldsymbol{\ell}] \setminus \{\boldsymbol{\ell}\}$ by μ. Similar to the study of min-id random vectors we can restrict the study of min-id sequences to min-id sequences whose d-dimensional margins satisfy Condition ($\Diamond$). Thus, from now on we will assume that the d-dimensional margins of a min-id sequence satisfy Condition ($\Diamond$), i.e. $\boldsymbol{X} \in (-\infty, \infty]^{\mathbb{N}}$ and $\boldsymbol{\ell} = \infty$.

Define $\mu_{i_1,\ldots,i_d}$ as the unique exponent measure of the d-dimensional margin $(X_{i_1}, \ldots, X_{i_d})$ of the min-id sequence $\boldsymbol{X}$. The results of Vatan (1985) are summarized in the following proposition and clarify the projective properties of extendible exponent measures μ_d.

Proposition 2.9 (Vatan (1985), *Exponent measure of a sequence*)

The following are equivalent for every min-id random vector $\boldsymbol{X}_d \in (-\infty, \infty]^d$ satisfying Condition ($\Diamond$):
1. *$\boldsymbol{X}_d$ is extendible to a min-id sequence $\boldsymbol{X}$.*
2. *There exists an exchangeable global exponent measure μ on $E_{\infty}^{\mathbb{N}}$ such that*

$$\mu\big(\{\boldsymbol{x} \in E_{\infty}^{\mathbb{N}} \mid (x_{i_1}, \ldots, x_{i_d}) \in A\}\big) = \mu_{i_1,\ldots,i_d}(A) = \mu_d(A),$$

for all distinct $(i_1, \ldots, i_d) \in \mathbb{N}^d$ and $A \in \mathcal{B}\big(E_{\infty}^d\big)$.

Proof The proof follows from an application of Proposition 2.8 and a translation of the results of Vatan (1985) to the framework of Resnick (2007) and the min-id case. □

Propositions 2.8 and 2.9 allow to characterize the upper tail dependence coefficients of an exchangeable min-id random vector via its exponent measure. To this purpose, for $2 \leq d' \leq d$, define the d'-variate upper tail dependence coefficient of an exchangeable min-id random vector $\boldsymbol{X}_d \in (-\infty, \infty]^d$ as $\rho_{d'}^u := \lim_{t \to \infty} \mathbb{P}(X_1 > t \mid X_2 > t, \ldots, X_{d'} > t)$.

Corollary 2.10 (Upper tail dependence of exchangeable min-id distribution)

The upper tail dependence coefficient of an exchangeable min-id random vector $\boldsymbol{X}_d \in (-\infty, \infty]^d$ satisfying Condition ($\Diamond$) is given by

$$\rho_{d'}^u = \exp\left(-\lim_{t \to \infty} \mu_{d'}\big([-\infty, t] \times (t, \infty]^{d'-1}\big)\right).$$

In particular $\rho_{d'}^u \leq \rho_{d'+1}^u$ for all $2 \leq d' < d$.

Proof The proof can be found in Appendix 1. $\qquad\qquad\qquad\qquad\qquad\square$

As a caveat we want to remark that in general $\lim_{t\to\infty} \mu_{d'}\big([-\infty, t] \times (t, \infty]^{d'-1}\big)$ may not be equal to $\mu_{d'}\big([-\infty, \infty) \times \{\infty\}^{d'-1}\big)$, which is due to the (possibly) infinite mass of $\mu_{d'}$.

2.3 Extended chronometers

In this subsection we characterize the class of nnnd infinitely divisible càdlàg processes. In particular, we show that the Lévy measure of such processes is concentrated on nnnd paths.

First, we recall the definition of infinitely divisible random vectors and infinitely divisible stochastic processes. For the sake of well-definedness of sums of (possibly) infinite quantities we only allow for random vectors and stochastic processes which cannot assume the values infinity and negative infinity with positive probability at the same time. Our specifications are formalized in the following definition.

Definition 2.11 (Infinitely divisible)

1. A random vector $\boldsymbol{H} \in \overline{\mathbb{R}}^d$ such that for all $1 \leq i \leq d$ we either have $P(H_i = \infty) = 0$ or $\mathbb{P}(H_i = -\infty) = 0$ is infinitely divisible (id) if for all $n \in \mathbb{N}$ there exist i.i.d. random vectors $(\boldsymbol{H}^{(i,1/n)})_{1 \leq i \leq n}$[4] such that $\boldsymbol{H} \sim \sum_{i=1}^n \boldsymbol{H}^{(i,1/n)}$.
2. A stochastic process $(H_t)_{t\in\mathbb{R}} \in \overline{\mathbb{R}}^{\mathbb{R}}$ such that for all $t \in \mathbb{R}$ we either have $P(H_t = \infty) = 0$ or $\mathbb{P}(H_t = -\infty) = 0$ is infinitely divisible (id) if for all $n \in \mathbb{N}$ there exist i.i.d. stochastic processes $\left(\left(H_t^{(i,1/n)}\right)_{t\in\mathbb{R}}\right)_{1 \leq i \leq n} \in \left(\overline{\mathbb{R}}^{\mathbb{R}}\right)^n$ such that $H \sim \sum_{i=1}^n H^{(i,1/n)}$.

An excellent textbook treatment of infinitely divisible distributions is Sato (1999). Infinitely divisible stochastic processes have been investigated, among others, by Lee (1967), Maruyama (1970), Barndorff-Nielsen et al. (2006), Rosiński (2018). We mainly follow the pathwise approach of Rosiński (2018). However, in contrast to Rosiński (2018), we only consider càdlàg id-processes and allow jumps to ∞. One can show that all relevant results of (Rosiński 2018, Sects. 1-3) remain valid under this slight change of the general framework.

In this paper we focus on nnnd càdlàg id-processes. For our purpose this is not a loss of generality, since de Finetti's Theorem implies that H in the construction method of Eq. (2) can be chosen as an nnnd càdlàg process. These processes can be viewed as an extension of chronometers, which were introduced in Barndorff-Nielsen et al. (2006).

---

[4] Again, the notation $\boldsymbol{H}^{(i,1/n)}$ shall emphasize that every $\boldsymbol{H}^{(i,1/n)}$ can be interpreted as a $1/n$ contribution to $\boldsymbol{H}$.

Definition 2.12 (Extended chronometer)

An nnnd infinitely divisible process $(H_t)_{t \in \mathbb{R}} \in D^\infty(\mathbb{R})$ is called an extended chronometer.

In contrast to chronometers in Barndorff-Nielsen et al. (2006), we do not require the stochastic continuity of extended chronometers. This generalization is necessary to account for non-continuous min-id distributions in Sect. 3.1. To simplify our developments we focus on extended chronometers that are compatible with Condition ($\Diamond$). To this purpose we need to impose two additional constraints on extended chronometers, which are specified in the following condition.

Condition ($\Diamond'$) An extended chronometer $(H_t)_{t \in \mathbb{R}}$ satisfies Condition ($\Diamond'$) if the following two conditions are satisfied
($\Diamond'1$) $\mathbb{P}\left(\lim_{t \to -\infty} H_t = 0\right) = 1$ and
($\Diamond'2$) $\mathbb{P}(H_t = \infty) < 1$ for all $t \in \mathbb{R}$.

Remark 1 (Equivalence of Conditions ($\Diamond$) and ($\Diamond'$)) Consider an extended chronometer $(H_t)_{t \in \mathbb{R}}$ and an i.i.d. unit exponential sequence $(E_i)_{i \in \mathbb{N}}$. By virtue of Eq. (2) we construct an exchangeable sequence $X := \left(\inf\{t \in \mathbb{R} \mid H_t \geq E_i\}\right)_{i \in \mathbb{N}}$. It is easy to see that $X > -\infty$ almost surely if and only if Condition ($\Diamond'1$) is satisfied. Moreover, assuming that X is min-id, one can check that Condition ($\Diamond'2$) is equivalent to the constraint $\ell = \infty$. Therefore, under the assumption that X is min-id, X satisfies Condition ($\Diamond$) if and only if $(H_t)_{t \in \mathbb{R}}$ satisfies Condition ($\Diamond'$).

Recall that the distribution of a non-negative stochastic process in $D^\infty(\mathbb{R})$ is uniquely determined by its Laplace transform. Translating (Rosiński 2018, Theorems 2.8 and 3.4) to non-negative càdlàg id-processes shows that the Laplace transform of a non-negative càdlàg id-process can be uniquely characterized by a function $b : \mathbb{R} \to \mathbb{R}$ and a measure ν on $D^\infty(\mathbb{R})$. To be precise, for all $d \in \mathbb{N}$, the Laplace transform of the d-dimensional margins of a non-negative id-process $(H_t)_{t \in \mathbb{R}} \in D^\infty(\mathbb{R})$ satisfying Condition ($\Diamond'2$) can be written as

$$L : [0, \infty)^d \times \mathbb{R}^d \to [0, 1]; (z, t) \mapsto \mathbb{E}\left[\exp\left(-\sum_{i=1}^{d} z_i H_{t_i}\right)\right]$$

$$= \exp\left(-\sum_{i=1}^{d} z_i b(t_i)\right. \tag{3}$$

$$+ \int_{D^\infty(\mathbb{R})} \left(\exp\left(-\sum_{i=1}^{d} z_i x(t_i)\right) - 1 - \sum_{i=1}^{d} z_i x(t_i) \mathbf{1}_{\{|x(t_i)| < \epsilon\}}\right) \nu(\mathrm{d}x)\right),$$

where $\epsilon > 0$ is arbitrary, $b : \mathbb{R} \to \mathbb{R}$ is a unique function, and ν is a unique measure on $D^\infty(\mathbb{R})$ satisfying

1. $\nu(0_{D^\infty(\mathbb{R})}) = 0$, and
2. $\int_{D^\infty(\mathbb{R})} \min\{x(t), 1\}\nu(\mathrm{d}x) < \infty$ for all $t \in \mathbb{R}$.

The function b is called drift of $(H_t)_{t\in\mathbb{R}}$ and the measure ν is called Lévy measure of $(H_t)_{t\in\mathbb{R}}$. More generally, every measure ν on $D^\infty(\mathbb{R})$ satisfying Conditions 1. and 2. is called a Lévy measure.

The literature usually treats id-processes as stochastic processes in $\mathbb{R}^\mathbb{R}$. Therefore, if $H \in D^\infty(\mathbb{R})$ is id, it is not clear whether the $H^{(i,1/n)}$ appearing in Definition 2.11 can also be chosen as elements of $D^\infty(\mathbb{R})$, as was already pointed out in (Barndorff-Nielsen et al. 2006, Remark 4.5). If H is stochastically continuous, non-negative, and almost surely non-decreasing the answer is positive, as can be deduced from (Barndorff-Nielsen et al. 2006, Proposition 6.2).

In our framework it is important that all $H^{(i,1/n)}$ can be chosen as càdlàg processes, since we need to interpret $\exp\left(-H^{(i,1/n)}\right)$ as a random survival function in Sect. 3. The next lemma generalizes the results of (Barndorff-Nielsen et al. 2006, Proposition 6.2) and shows that all càdlàg id-processes have a representation as a sum of càdlàg id-processes.

Lemma 2.13 (Càdlàg id-processes are distributed as i.i.d. sum of càdlàg id-processes)

Consider an id-process $H \in D^\infty(\mathbb{R})$. Then, for every $n \in \mathbb{N}$, we can find i.i.d. id-processes $\left(H^{(i,1/n)}\right)_{1\leq i\leq n} \in (D^\infty(\mathbb{R}))^n$ such that $H \sim \sum_{i=1}^n H^{(i,1/n)}$.

Proof The proof can be found in Appendix 1. □

Lemma 2.13 verifies that Definition 2.11 could also be formulated solely for processes which are id in the space $D^\infty(\mathbb{R})$, since it excludes the possibility of the existence of a stochastic process $H \in D^\infty(\mathbb{R})$ which is id in $\overline{\mathbb{R}}^\mathbb{R}$, but not id in $D^\infty(\mathbb{R})$.

To interpret $\exp\left(-H^{(i,1/n)}\right)$ as a random distribution function, when H is an extended chronometer, it remains to ensure that each $H^{(i,1/n)}$ is nnnd, i.e. an extended chronometer. The following corollary shows that extended chronometers are infinitely divisible in the space of extended chronometers, i.e. that every extended chronometer can be represented as the sum of arbitrarily many extended chronometers.

Corollary 2.14 (Extended chronometers are distributed as i.i.d. sum of extended chronometers)

Consider an extended chronometer H. Then, for every $n \in \mathbb{N}$, we can find i.i.d. extended chronometers $\left(H^{(i,1/n)}\right)_{1\leq i\leq n}$ such that $H \sim \sum_{i=1}^n H^{(i,1/n)}$.

Proof The proof can be found in Appendix 1. □

Our next goal is to connect the path properties of an extended chronometer with the support of its Lévy measure. To get an intuition about the correspondences of

path properties of an id-process and properties of its Lévy measure we recall the following example from Rosiński (2018).

Example 2.15 (Finite Lévy measure (Rosiński 2018, Example 2.26))

Consider a sequence of i.i.d. càdlàg processes $(h^{(i)})_{i \in \mathbb{N}}$ with marginal distribution $\mathbb{P}_h$, a Poisson random variable N with mean λ, and define a stochastic process $(H_t)_{t \in \mathbb{R}} := \left(\sum_{i=1}^{N} h_t^{(i)} \right)_{t \in \mathbb{R}}$. (Rosiński 2018, Example 2.26) proves that H is infinitely divisible with drift 0 and Lévy measure $\lambda \mathbb{P}_h \left(\cdot \cap \{0_{D^\infty(\mathbb{R})}\}^{\complement} \right)$. Obviously, H is nnnd if and only if the Lévy measure $\lambda \mathbb{P}_h \left(\cdot \cap \{0_{D^\infty(\mathbb{R})}\}^{\complement} \right)$ is supported on nnnd functions.

Unfortunately, the Lévy measure of an id-process is infinite in most cases of interest. Thus, the construction method in Example 2.15 is rather limited and we cannot immediately draw the same conclusions as in Example 2.15 for general extended chronometers.

In the following we show that the observations of Example 2.15 remain valid for extended chronometers, i.e. if a càdlàg id-process is nnnd then its Lévy measure is concentrated on nnnd càdlàg functions. More specifically, we show that the Lévy measure of an extended chronometer satisfying Condition $(\Diamond')$ is supported on the nnnd functions in $D^\infty(\mathbb{R})$ satisfying $\lim_{t \to -\infty} x(t) = 0$. A weaker version of this statement was stated (without proof) in (Lee 1967, Sect. 4), who followed a technically different approach in comparison to the pathwise approach of Rosiński (2018). The author claimed that a proof of his statement works similar to other proofs given in Lee (1967). However, all of the referred proofs are not very detailed and are not compatible with the pathwise approach of Rosiński (2018). An alternative approach to prove our claim would use an application of (Rosiński 2018, Theorem 3.4), which provides a tool to restrict the Lévy measure of an id-process to a smaller domain. Unfortunately, the theorem cannot be applied in our setting, since our favored domain, nnnd càdlàg functions, does not form an algebraic group under addition. Therefore, we provide a formal proof of our claims in the following proposition.

Proposition 2.16 (Laplace transform of an extended chronometer satisfying Condition $(\Diamond')$)

Let $d \in \mathbb{N}$ and let $(H_t)_{t \in \mathbb{R}}$ denote an extended chronometer satisfying Condition $(\Diamond')$. Then, for $z \in [0, \infty)^d$, $t \in \mathbb{R}^d$, we have

$$L(t, z) = \mathbb{E}\left[e^{-\sum_{i=1}^{d} z_i H_{t_i}} \right]$$

$$= \exp\left(-\sum_{i=1}^{d} z_i b(t_i) - \int_{M_\infty^0} \left(1 - e^{-\sum_{i=1}^{d} z_i x(t_i)} \right) \nu(\mathrm{d}x) \right),$$

where ν is a Lévy measure on

$$M_\infty^0 := \{ x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}, \lim_{t \to -\infty} x(t) = 0 \},$$

and $b \in M_\infty^0 \cap D(\mathbb{R})$.

Proof The proof can be found in Appendix 1. □

Remark 2 (Id-process with Lévy measure on M_∞^0 has extended chronometer version) By a similar reasoning as in the proof of Proposition 2.16 it is also possible to prove that every driftless càdlàg id-process with Lévy measure concentrated on nnnd càdlàg functions has a version that is non-decreasing and non-negative. We omit a proof of this statement, since this fact will not be used in our paper.

Remark 3 (Condition ($\Diamond'1$) corresponds to vanishing functions in Lévy measure) The proof of Proposition 2.16 shows that the Lévy measure of an id-process satisfying Condition ($\Diamond'1$) is concentrated on càdlàg paths vanishing at $-\infty$. If we omit Condition ($\Diamond'1$) this is no longer the case, which is the reason why we later need to omit this condition in Corollary 3.7.

Interestingly, we can infer $\mathbb{P}(H_t = \infty)$ from the associated Lévy measure, as the next example shows.

Example 2.17 (Probability of a jump to ∞) Consider an extended chronometer $(H_t)_{t \in \mathbb{R}}$ satisfying Condition ($\Diamond'$) with Lévy measure v and drift b. We want to investigate $\mathbb{P}(H_t = \infty)$ for every $t \in \mathbb{R}$, which is equivalent to investigating $\mathbb{P}(X_i \leq t \; \forall \; i \in \mathbb{N})$ for every $t \in \mathbb{R}$, where X denotes the exchangeable sequence constructed via Eq. (2).

The (one-dimensional) Lévy–Khintchine representation of the infinitely divisible random variable H_t yields $\mathbb{P}(H_t = \infty) = 1 - \exp(-v_t(\infty))$, where v_t denotes the (one-dimensional) Lévy measure of H_t. Proposition 2.16 shows that $v_t(\infty) = v(\{x \in M_\infty^0 \mid x(t) = \infty\})$. Therefore,

$$\mathbb{P}(H_t = \infty) = 1 - \exp\left(-v(\{x \in M_\infty^0 \mid x(t) = \infty\})\right).$$

We emphasize that, in contrast to e.g. additive processes, H cannot be decomposed into $H = H^{(1)} + H^{(2)}$, where $H^{(1)}$ is always finite and independent of $H^{(2)} \in \{0, \infty\}^{\mathbb{R}}$. Therefore, jumps to ∞ do not occur independently of the path behavior of the process in general. Let us verify this claim by an application of Example 2.15. Decompose v into $v = v_\infty + v_f$, where $v_f := v(\cdot \cap \{x \in M_\infty^0 \mid x(t) < \infty \text{ for all } t \in \mathbb{R}\})$ is concentrated on finite paths and $v_\infty := v(\cdot \cap \{x \in M_\infty^0 \mid x(t) = \infty \text{ for some } t \in \mathbb{R}\})$ is concentrated on paths that jump to ∞. Now, assuming that v_∞ is a finite measure with total mass c, we define $H^{(2)}$ as the id-process with Lévy measure v_∞ and $H^{(1)}$ as an independent id-process with Lévy measure v_f and drift b. Obviously, $H \sim H^{(1)} + H^{(2)}$, where $H^{(1)} \in D(\mathbb{R})$ has finite sample paths and

$$H^{(2)} \sim \sum_{i=1}^{N} h^{(2,i)},$$

where $(h^{(2,i)})_{i \in \mathbb{N}}$ denotes an i.i.d. sequence of càdlàg processes with distribution v_∞/c and N denotes an independent Poisson random variable with mean c. Since the paths

of $h^{(2,1)}$ can follow every increasing (càdlàg) path we observe that $H_t^{(2)}$ may take all finite values. Thus, in general, H cannot be decomposed into a finite process $H^{(1)}$ and a "killing" process $H^{(2)} \in \{0, \infty\}^{\mathbb{R}}$. The case of infinite v_∞ follows from $v_{t,\infty} := v\big(\cdot \cap \{x \in M_\infty^0 \mid x(t) = \infty\}\big)$, where $v_{t,\infty}$ is a finite measure for all $t \in \mathbb{R}$ and $v_\infty = \lim_{t \to \infty} v_{t,\infty}$.

3 Main results: Linking exchangeable min-id sequences to extended chronometers

After having collected all auxiliary results, we now formulate the main contributions of this paper.

3.1 Extendible min-id distributions satisfying $(\Diamond)$ are in one-to-one correspondence with extended chronometers satisfying $(\Diamond')$

We start with the characterization of extendible min-id distributions satisfying Condition $(\Diamond)$. Recall that de Finetti's Theorem implies that every exchangeable sequence $X \in \overline{\mathbb{R}}^{\mathbb{N}}$ is in one-to-one correspondence with an nnnd càdlàg process. We show that the class of stochastic processes corresponding to min-id sequences satisfying Condition $(\Diamond)$ is precisely the class of extended chronometers satisfying Condition $(\Diamond')$.

Theorem 3.1 (Extendible min-id distributions correspond to extended chronometers)

The following are equivalent:
1. *X is an exchangeable min-id sequence satisfying Condition $(\Diamond)$.*
2. *There exists an extended chronometer $(H_t)_{t \in \mathbb{R}} \in D^\infty(\mathbb{R})$ satisfying Condition $(\Diamond')$ such that $X \sim \big(\inf\{t \in \mathbb{R} \mid H_t \geq E_i\}\big)_{i \in \mathbb{N}}$, where $(E_i)_{i \in \mathbb{N}}$ are i.i.d. unit exponential and independent of H.*

Moreover, the law of H is uniquely associated to the law of X.

Proof The proof can be found in Appendix 1. $\square$

It is worth noting that Lemma 2.5 can be translated into a time-change of the extended chronometer.

Corollary 3.2 (Marginal transformation of exchangeable min-id sequence is time-change of the chronometer)

Consider an exchangeable min-id sequence $X \in (\infty, \infty]^{\mathbb{N}}$ and a left-continuous non-decreasing transformation f. Let H^X denote the extended chronometer

corresponding to X and $H^{f(X)}$ denote the chronometer corresponding to $f(X)$. Then $H^{f(X)} \sim H^X \circ f^{\leftharpoonup}$, where $f^{\leftharpoonup}(t) := \inf\{s \in \mathbb{R} \mid f(s) > t\}$.

Proof The claim follows from the identity

$$\mathbb{P}\big(f(X_1) > t_1, \ldots, f(X_d) > t_d\big) = \mathbb{P}\big(X_1 > f^{\leftharpoonup}(t_1), \ldots, X_d > f^{\leftharpoonup}(t_d)\big).$$

$\square$

The following examples present two interesting applications of Corollary 3.2.

Example 3.3 ($\mathbb{N}_0$-valued exchangeable min-id sequences) Consider the non-decreasing left-continuous transformation $x \mapsto \lceil x \rceil := \min\{n \in \mathbb{N}_0 \mid x \leq n\}$. Lemma 2.5 implies that each exchangeable min-id sequence X can be transformed into an $\mathbb{N}_0$-valued exchangeable min-id sequence $\lceil X \rceil$. Corollary 3.2 shows that the extended chronometer $H^{\lceil X \rceil}$ associated with $\lceil X \rceil$ can be obtained via a time-change of the extended chronometer H^X associated with X. Thus, $H^{\lceil X \rceil} \sim H^X \circ \lfloor \cdot \rfloor$, where $\lfloor x \rfloor := \lceil x \rceil^{\leftharpoonup} = \max\{n \in \mathbb{N}_0 \mid x \geq n\}$. Defining $(J_i)_{i \in \mathbb{N}} := (H_i^X - H_{i-1}^X)_{i \in \mathbb{N}}$ we observe that

$$H_t^{\lceil X \rceil} = H_{\lfloor t \rfloor}^X = \left(H_0^X + \sum_{i=1}^{\lfloor t \rfloor} J_i\right) \mathbf{1}_{\{t \geq 0\}}$$

can be represented as a pure jump process. If $H_0^X = 0$ and H^X has stationary and independent increments $H^{\lceil X \rceil}$ is known as a random walk and the sequence $\lceil X \rceil$ follows a multivariate narrow-sense geometric distribution, see Mai et al. (2013). In this case X has d-dimensional marginal distributions

$$(X_1, \ldots, X_d) \sim \big(\inf\{E_I \mid i \in I\}\big)_{1 \leq i \leq d},$$

where $(E_I)_{I \subset \{1, \ldots, d\}}$ is a collection of independent geometrically distributed random variables with parameters $(1 - p_I)$ such that p_I only depends on $|I|$. Moreover, the associated extended chronometer $H^{\lceil X \rceil}$ is a random walk with infinitely divisible i.i.d. jumps $J_i \sim H_1^X$.

Example 3.4 (From min- to max-id) Consider an exchangeable min-id sequence X with associated chronometer H^X and a continuous strictly decreasing transformation f with its corresponding inverse f^{-1}. Lemma 2.5 shows that $Y^{(f)} := f(X)$ is an exchangeable max-id sequence. According to de Finetti's Theorem there exists a random distribution function $F^{(f)}$ such that

$$Y^{(f)} \sim \big(\inf\{t \in \mathbb{R} \mid F_t^{(f)} \geq U_i\}\big)_{i \in \mathbb{N}}$$

for an i.i.d. sequence of Uniform(0, 1) distributed random variables $(U_i)_{i \in \mathbb{N}}$. Noting that

$$\mathbb{P}(Y_1^{(f)} \le t_1, \dots, Y_d^{(f)} \le t_d) = \mathbb{P}\left(X_1 \ge f^{-1}(t_1), \dots, X_d \ge f^{-1}(t_d)\right)$$

$$= \lim_{z \searrow t} \mathbb{P}\left(X_1 > f^{-1}(z_1), \dots, X_d > f^{-1}(z_d)\right) = \lim_{z \searrow t} \mathbb{E}\left[\exp\left(-\sum_{i=1}^d H_{f^{-1}(z_i)}^X\right)\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{i=1}^d \lim_{z_i \searrow t_i} H_{f^{-1}(z_i)}^X\right)\right]$$

yields that $\left(F_t^{(f)}\right)_{t \in \mathbb{R}} \sim \left(\exp\left(-\lim_{z \searrow t} H_{f^{-1}(z)}^X\right)\right)_{t \in \mathbb{R}}$. Therefore,

$$\boldsymbol{Y}^{(f)} \sim \left(\inf\left\{t \in \mathbb{R} \;\middle|\; -\log\left(1 - \exp\left(-\lim_{z \searrow t} H_{f^{-1}(z_i)}^X\right)\right) \ge E_i\right\}\right)_{i \in \mathbb{N}},$$

where $(E_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence of unit exponential random variables.

3.2 The exponent measure of an exchangeable min-id sequence satisfying $(\Diamond)$ is a mixture of product probability measures

We present an analogue of de Finetti's Theorem for exponent measures of exchangeable min-id sequences satisfying $(\Diamond)$.

First, we need to introduce some notation. For any distribution function G of a random variable on $(-\infty, \infty]$, define $\mathbb{P}_G$ as the probability measure associated with the distribution function G. Furthermore, $\otimes_{i=1}^d \mathbb{P}_G$ denotes the probability measure on $(-\infty, \infty]^d$ associated with $d \in \mathbb{N} \cup \{\infty\}$ i.i.d. copies of random variables with distribution $\mathbb{P}_G$. Define the space of distribution functions of random variables on $(-\infty, \infty]$ as $\overline{M}_\infty^0 := \{G : \mathbb{R} \to [0, 1] \mid G \text{ is distribution function of a random variable on } (-\infty, \infty]\}$. Let γ denote a measure on $\overline{M}_\infty^0$ and define a mixture of product probability measures by

$$\mu_{\gamma, d}(\cdot) := \int_{\overline{M}_\infty^0} \otimes_{i=1}^d \mathbb{P}_G(\cdot)\, \gamma(\mathrm{d}G). \tag{4}$$

Additionally, for any non-decreasing function $b \in M_\infty^0$, define $\mu_{b,d}$ as the exponent measure of $d \in \mathbb{N} \cup \{\infty\}$ i.i.d. copies of random variables on $(-\infty, \infty]$ with survival function $\exp(-b(\cdot))$ whose existence is ensured by Proposition 2.9.

Theorem 3.5 (Exponent measure of exchangeable min-id sequence)

The following are equivalent:
1. *X is an exchangeable min-id sequence satisfying $(\Diamond)$.*
2. *There exists a unique function $b \in M_\infty^0$ and a unique measure γ on $\overline{M}_\infty^0$ satisfying $\gamma\left(0_{D^\infty(\mathbb{R})}\right) = 0$ and $\int_{\overline{M}_\infty^0} G(t)\gamma(\mathrm{d}G) < \infty$ for all $t \in \mathbb{R}$ such that the exponent measure of X is given by*

$$\mu(A) = \mu_{b,\infty}(A) + \mu_{\gamma,\infty}(A)$$

for every $A \in \mathcal{B}(E_\infty^{\mathbb{N}})$.

Moreover, the Lévy measure and drift of the extended chronometer associated with X are given by $\nu(A) = \gamma(\{G \in \overline{M}_\infty^0 \mid G = 1 - \exp(-x(\cdot))$ for some $x \in A\})$ for every $A \in \mathcal{B}(D^\infty(\mathbb{R}))$ and $b(t) = \mu_{b,1}([-\infty, t])$.

Proof The proof can be found in Appendix 1.   □

Recall that a min-id sequence has exponent measure supported on $A^\perp :=$ $\{x \in \overline{\mathbb{R}}^{\mathbb{N}} \mid x_i = \infty$ for all but one $i \in \mathbb{N}\}$ if and only if it is an i.i.d. sequence. Theorem 3.5 yields a decomposition of the global exponent measure μ into $\mu_b + \mu_\gamma$, where $\mu_b := \mu_{b,\infty}$ is supported on the set $A^\perp$ and $\mu_\gamma := \mu_{\gamma,\infty}$. Therefore, $X \sim \min\{X^{(1)}, X^{(2)}\}$, where $X^{(1)}$ is an i.i.d. sequence with exponent measure μ_b and $X^{(2)}$ is an exchangeable min-id sequence with exponent measure μ_γ. This raises the question whether the decomposition from Theorem 3.5 separates μ into an independence part μ_b and a dependence part μ_γ, which would be a desired feature for modeling purposes. Mathematically this translates to μ_b and μ_γ being singular.

If μ_γ does not have mass on $A^\perp$, then $X^{(2)}$ does not contain an independent sequence, i.e. $X^{(2)}$ cannot be further decomposed into the minimum of a non-trivial i.i.d. sequence and an exchangeable min-id sequence. On the level of the associated extended chronometer this would correspond to the fact that exponent measures μ_γ associated to driftless extended chronometers do not put mass on $A^\perp$. Indeed, the next corollary shows that μ_γ never puts mass on $A^\perp$, which implies that the decomposition of μ into μ_b and μ_γ separates dependence from independence.

Corollary 3.6 (Decomposition of an exponent measure into dependence and independence)

Let X denote the exchangeable min-id sequence associated to the nnnd id-process $H = b + \tilde{H}$, where b denotes the drift of H and $\tilde{H}$ denotes the driftless random component of H. Let $\mu^{\tilde{H}}$ denote the exponent measure associated to $\tilde{H}$. Then $\mu_\gamma = \mu^{\tilde{H}}$ and the exponent measure μ of X is given by $\mu = \mu_b + \mu_\gamma$, where μ_b and μ_γ are singular.

Proof Note that $\mathbb{P}_G\big((-\infty, \infty)\big) \prod_{i \in \mathbb{N}} \mathbb{P}_G(\{\infty\}) = 0$ for all $G \in \overline{M}_\infty^0$, since $\lim_{t \to \infty} G(t) > 0$ for all $0_{D^\infty(\mathbb{R})} \neq G \in \overline{M}_\infty^0$. Thus,

$$\mu_\gamma\big(A^\perp\big) = \int_{\overline{M}_\infty^0} \sum_{i \in \mathbb{N}} \big(\otimes_{k \in \mathbb{N}} \mathbb{P}_G\big)\big(\{x_i < \infty \text{ and } x_j = \infty \text{ for all } j \neq i\}\big)\gamma(\mathrm{d}G)$$

$$= \int_{\overline{M}_\infty^0} \sum_{i \in \mathbb{N}} \mathbb{P}_G\big((-\infty, \infty)\big)\bigg(\prod_{j \in \mathbb{N}} \mathbb{P}_G(\{\infty\})\bigg)\gamma(\mathrm{d}G) = 0$$

Thus, μ_b and μ_γ are singular.   □

Remark 4 (Non-separability of dependence and independence for finite dimensional margins) It is educational to observe that $\mu_{b,d}$ and $\mu_{\gamma,d}$ are singular for a fixed $d \in \mathbb{N}$ if and only if γ is concentrated on the set $\{G \in \overline{M}_\infty^0 \mid \lim_{t \to \infty} G(t) = 1\}$, since

$$\mu_{\gamma,d}\big(\cup_{1\le i\le d}\{x_i < \infty \text{ and } x_j = \infty \text{ for all } j \ne i\}\big)$$

$$= \sum_{1\le i\le d} \int_{\overline{M}_\infty^0} \mathbb{P}_G\big((-\infty, \infty)\big)\big(\mathbb{P}_G(\{\infty\})\big)^{d-1} \gamma(\mathrm{d}G)$$

$$= d \int_{\overline{M}_\infty^0} \Big(\lim_{t\to\infty} G(t)\Big)\Big(1 - \lim_{t\to\infty} G(t)\Big)^{d-1} \gamma(\mathrm{d}G).$$

Therefore, $\mu_{b,d}$ and $\mu_{\gamma,d}$ usually do not separate dependence from independence, meaning that $X_d^{(2)} \sim \min\big\{X_d^{(2,\perp)}, X_d^{(2,\not\perp)}\big\} \sim \exp\big(-\mu_{\gamma,d}\big((\cdot,\infty]^\complement\big)\big)$, where $X_d^{(2,\perp)}$ has components which are d i.i.d. copies of a non-trivial random variable and $X_d^{(2,\not\perp)}$ denotes a non-trivial min-id random vector. Intuitively, this may be interpreted as follows: the finite dimensional exponent measure $\mu_{\gamma,d}$ smears around independence and this effect can only be distinguished from independence in the limit.

Remark 5 (Dependence structure of exchangeable min-id sequences) (Mai and Scherer 2014, Lemma 4.4) shows that exchangeable min-stable sequences admit positive ρ_2^u if and only if the exchangeable min-stable sequence is given by the comonotonic sequence $X = (\bar{X}, \bar{X}, \dots)$, where $\bar{X} \in \overline{\mathbb{R}}$ is some univariate random variable. Therefore, exchangeable min-stable sequences satisfy $\rho_2^u \in \{0, 1\}$ and $\rho_2^u = 1$ implies $X = (\bar{X}, \bar{X}, \dots)$, which raises the question whether the same result holds for exchangeable min-id sequences.

The question can be answered by the following example: Define an exchangeable min-id sequence via the extended chronometer $(H_t)_{t\in\mathbb{R}} = -\log\big(1 - \Gamma_t\big)$, where $(\Gamma_t)_{t\in\mathbb{R}}$ denotes a stochastically continuous Dirichlet process, see Ferguson (1973). The Dirichlet process naturally appears in Bayesian statistics when the distribution function $(\Gamma_t)_{t\in\mathbb{R}}$ of an i.i.d. sequence is viewed as the random quantity of interest. The distribution of the Dirichlet process is then specified as the prior distribution on the space of distribution functions, which arises under the assumption that $\big(\int_{A_1} \mathrm{d}\Gamma, \dots, \int_{A_d} \mathrm{d}\Gamma\big)$ follows a Dirichlet distribution for every $d \in \mathbb{N}$ and every measurable disjoint partition $(A_i)_{1\le i\le d}$ of $\mathbb{R}$. The Dirichlet process is particularly convenient in Bayesian statistics, since its posterior distribution is again a (non-stochastically continuous) Dirichlet process.

The authors of Ferguson (1974), Doksum (1974) have shown that H is infinitely divisible, which implies that the associated exchangeable sequence X is min-id. Moreover, Mai et al. (2015) show that the upper and lower bivariate tail dependence coefficients ρ_2^u and ρ_2^l of X can take any value in $(0, 1)$. Thus, exchangeable min-id sequences can exhibit arbitrary positive bivariate upper and lower tail dependence, which shows that the dependence structure of exchangeable min-id sequences is much richer than the dependence structure of exchangeable min-stable sequences.

To verify that $\rho_2^u = 1$ does not imply $X = (\bar{X}, \bar{X}, \dots)$ when X is an exchangeable min-id sequence it is easy to see that every extended chronometer with finite Lévy measure concentrated on the set $\{x \in M_\infty^0 \mid x(t) < \infty, \lim_{t\to\infty} x(t) = \infty\}$ yields an exchangeable min-id sequence X which satisfies $\rho_2^u = 1$ but not $X = (\bar{X}, \bar{X}, \dots)$.

3.3 Characterization of general exchangeable min-id sequences

Even though we have mentioned that studying min-id distributions satisfying Condition ($\lozenge$) is not a loss of generality, we feel the need to translate the results of Theorems 3.1 and 3.5 to arbitrary exchangeable min-id sequences on $[-\infty, \infty]^{\mathbb{N}}$. The appearance of the following corollary is slightly more technical than that of Theorems 3.1 and 3.5, which is due to the subtleties in the definition of a survival function of random vectors in $[-\infty, \infty]^d$ and explains why we preferred to develop the preliminary results under Conditions ($\lozenge$) and ($\lozenge'$).

Corollary 3.7 (Characterization of general exchangeable min-id sequences)

Assume that $\mathbb{P}(X = -\infty) < 1$. Then, the following are equivalent:
1. *$X \in \overline{\mathbb{R}}^{\mathbb{N}}$ is an exchangeable min-id sequence.*
2. *There exists $\ell = (\ell, \ell, \ldots) \in (-\infty, \infty]^{\mathbb{N}}$ such that*

$$(X_i)_{i \in \mathbb{N}} \sim \left(\inf\{t \in (-\infty, \ell) \mid H_t \geq E_i\}\right)_{i \in \mathbb{N}},$$

where $\inf \emptyset =: \ell$, $(E_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. unit exponential random variables independent of a unique extended chronometer $(H_t)_{t \in (-\infty, \ell)} \in D^{\infty}((-\infty, \ell))$ satisfying

(a) *$\mathbb{P}(H_t = 0) = 1$ for all $t < \sup\left\{x \in \mathbb{R} \mid \mathbb{P}(X_1 \leq x) = 0\right\} =: w,$*
(b) *The Lévy measure ν of H is supported on*

$$M_\ell := \{x \in D^{\infty}((-\infty, \ell)) \mid x \text{ is nnnd and } x(t) = 0 \text{ for all } t < w\},$$

(c) *H has real-valued drift $b \in M_\ell \cap D((-\infty, \ell))$.*

3. *There exists $\ell = (\ell, \ell, \ldots) \in (-\infty, \infty]^{\mathbb{N}}$ and an exchangeable Radon measure μ on $E_\ell^{\mathbb{N}}$ such that*

$$\mathbb{P}\left((X_{i_1}, \ldots, X_{i_d}) \in \times_{i=1}^d \{x_i, \ell]\right)$$
$$= \begin{cases} \exp\left(-\mu\left(\{y \in E_\ell^{\mathbb{N}} \mid (y_{i_1}, \ldots, y_{i_d}) \in \left(\times_{i=1}^d \{x_i, \infty]\right)^{\complement}\}\right)\right) & x < \ell \\ 0 & \text{otherwise} \end{cases},$$

where $\{x_i, \infty]$ is interpreted as $(x_i, \infty]$ if $x_i > -\infty$ and $\{-\infty, \infty]$ is interpreted as $[-\infty, \infty]$. Moreover, for all $A \in \mathcal{B}\left(E_\ell^{\mathbb{N}}\right)$, we have

$$\mu(A) = \mu_b(A) + \int_{\overline{M}_\ell} \otimes_{i \in \mathbb{N}} \mathbb{P}_G(A) \gamma(dG),$$

where b is a unique nnnd càdlàg function and γ is a unique measure

$$\overline{M}_\ell := \left\{G : (-\infty, \ell) \to [0, 1] \mid G \text{ distr. fct. of random variable on } [w, \ell]\right\}$$

satisfying

$$\gamma(0_{D^\infty(\mathbb{R})}) = 0 \ and \ \int_{\overline{M}_\ell} G(t)\gamma(\mathrm{d}G) < \infty \ for \ all \ t \in (-\infty, \ell).$$

The relation of γ and ν is given by

$$\gamma(A) = \nu\left(\left\{x \in M_\ell \mid \left(1 - \exp(-x(\cdot))\right) \in A\right\}\right) \ \text{for every} \ A \in \mathcal{B}\left(\overline{M}_\ell\right).$$

Proof The proof of Proposition 2.16 shows that the Lévy measure of an nnnd càdlàg id-process is concentrated on nnnd càdlàg functions. The rest of the proof is a combination of Theorems 3.1, 3.5, and Lemma 2.5. $\qquad\square$

A version of *Corollary* 3.7 for exchangeable max-id sequences can be easily deduced from Lemma 2.5 and Example 3.4.

4 Established families unified under the present umbrella

In this section we present several important examples of exchangeable min-id sequences. Moreover, we investigate the dependence structure of exchangeable min-id sequences and characterize extendible min-id random vectors with finite exponent measure.

4.1 Independence and comonotonicity

Corollary 3.7 shows that the random sequence $X \in [-\infty, \infty]^\mathbb{N}$ with i.i.d. components distributed according to the survival function $\exp(-b(\cdot))$ corresponds to the deterministic process $H_t = b(t)$. The exponent measure of X is given by $\mu = \mu_b$.

Consider the comonotonic case $X = (\bar{X}, \bar{X}, \ldots)$, where the random variable $\bar{X} \in \overline{\mathbb{R}}$ satisfies $\mathbb{P}(\bar{X} = -\infty) < 1$. The corresponding driftless extended chronometer H is given by $\left(H_t\right)_{t \in (-\infty, \ell)} = \left(\infty \mathbf{1}_{\{\bar{X} \leq t\}}\right)_{t \in (-\infty, \ell)}$. The Lévy measure ν of H is supported on

$$\left\{x \in D^\infty((-\infty, \ell)) \mid x(\cdot) = \infty \mathbf{1}_{\{\cdot \geq a\}} \ \text{for some} \ a \in (-\infty, \ell)\right\}.$$

Moreover, for $t < \ell$,

$$\nu\left(\left\{\infty \mathbf{1}_{\{\cdot \geq a\}} \mid a \in (-\infty, t]\right\}\right) = -\log\left(\mathbb{P}(\bar{X} > t)\right).$$

Therefore,

$$\mu\left((t, \infty]^{\complement}\right) = \int_{M_\ell} \left(1 - \exp\left(-\sum_{i \in \mathbb{N}} x(t_i)\right)\right)\nu(\mathrm{d}x) = -\log\left(\mathbb{P}\left(\bar{X} > \max_{i \in \mathbb{N}} t_i\right)\right)$$
$$= -\log\left(\mathbb{P}(X > t)\right).$$

4.2 Exogenous shock models / additive processes

Mai et al. (2016), Sloot (2020) prove that extendible exogenous shock models constitute a proper subclass of extendible min-id distributions. Fix $d \in \mathbb{N}$ and define a family of independent random variables $(\tau_I)_{I \subset \{1,\ldots,d\}} \in [0,\infty)^{2^d}$. Moreover, let the distribution of τ_I be continuous and solely dependent on $|I|$, i.e. the cardinality of the subset I of $\{1,\ldots,d\}$. Then the d-dimensional random vector

$$(X_i)_{1 \le i \le d} \sim \left(\min_{i \in I \subset \{1,\ldots,d\}} \tau_I \right)_{1 \le i \le d} \tag{5}$$

is exchangeable and interpreted as an exogenous shock model. The random variable τ_I models the arrival time of an exogenous shock destroying all components I and X_i equals the first time point at which component i is affected by some shock. Exchangeability boils down to our assumption that the shock arrival time distributions only depend on the number of components affected by the respective shocks. Thus, the model is parametrized by d distribution functions, since there are d different "shock sizes". If we let $d \to \infty$ in this construction, Kolmogorov's extension theorem guarantees the existence of an exchangeable sequence $X \in (0,\infty)^{\mathbb{N}}$ with the just described d-dimensional marginal distributions. Mai et al. (2016), Sloot (2020) prove that the associated extended chronometer is a stochastically continuous càdlàg process with independent increments. Such processes are known as (possibly killed) additive subordinators, see Sato (1999), Bertoin (1999) for a detailed treatment. The corresponding Lévy measure ν of H is supported on the class of one-step functions $\{u \mathbf{1}_{\{s \le \cdot\}} \mid s \in (0,\infty), u \in (0,\infty]\}$, see Rosiński (2018). Assuming that the extended chronometer is driftless, this implies that the associated exponent measure μ is an infinite mixture of (probability) distributions in the set

$$\left\{ \otimes_{i \in \mathbb{N}} \left((1 - \exp(-u))\delta_s + \exp(-u)\delta_\infty \right) \mid u \in (0,\infty], s \in (0,\infty) \right\},$$

where δ_s denotes the Dirac measure at s.

An important subclass of exogenous shock models is the class of Marshall–Olkin distributions. It is obtained by restricting the distribution of τ_I in (5) to exponential distributions. Furthermore, Mai (2010) shows that the class of extended chronometers corresponding to extendible Marshall–Olkin distributions is precisely the class of killed Lévy subordinators. Killed Lévy subordinators $(H_t)_{t \in \mathbb{R}}$ can be characterized as the class of extended chronometers with stationary and independent increments, which satisfy $H_0 = 0$ and have an independent exponential killing rate, see Bertoin (1999) for more details. Example 2.23 in Rosiński (2018) shows that the Lévy measure of a killed Lévy subordinator H is the image measure of the map

$$f : ([0,\infty) \times (0,\infty]; \mathcal{B}((0,\infty) \times (0,\infty]); \lambda \otimes \upsilon) \to M_\infty^0, \ (s,u) \mapsto u\mathbf{1}_{\{s \le \cdot\}} \tag{6}$$

on M_∞^0, where λ denotes the Lebesgue measure and υ denotes the Lévy measure of the infinitely divisible random variable H_1. The drift of H is given by $b(t) = b(1)t$, where $b(1) \ge 0$ denotes the drift of H_1.

In comparison to Mai et al. (2016), Sloot (2020), our framework allows for some additional flexibility, since we neither assume that H is indexed by $[0, \infty)$, nor stochastically continuous, nor that $\lim_{t \to \infty} H_t = \infty$. More precisely, we allow for non-continuously distributed failure times on $[-\infty, \infty]$ instead of continuously distributed failure times on $(0, \infty)$. An important observation in this regard is that the behavior of the stochastic model (5) under monotone, componentwise transformations of the X_i is not necessarily well-behaved in the framework of Mai et al. (2016), Sloot (2020). For instance, if H is a Lévy subordinator, a componentwise transformation of the X_i corresponds to a change from H_t to $H_{f(t)}$, which is no longer a Lévy subordinator unless f is linear with $f(0) = 0$. In contrast, Corollary 3.2 constitutes that under the more general umbrella of id-processes such marginal transformations are well-behaved.

As a final note of caution we remark that non-decreasing càdlàg processes with independent increments are not necessarily infinitely divisible. More precisely, our framework includes all càdlàg processes with independent id increments, meaning that if $\left(\lim_{s \nearrow t}(H_t - H_s)\right)_{t \in \mathbb{R}}$ is a collection of independent and id random variables then H is id. Thus, our framework does not incorporate all non-decreasing càdlàg processes with independent increments, but only extended chronometers with independent increments.

4.3 A common framework for extreme-value copulas and (reciprocal) Archimedean copulas

We consider a pair (κ, ρ) of a Radon measure κ on $[0, \infty)$ and a probability measure ρ on $\{G \mid G \text{ is a distribution function of a random variable on } [0, \infty]\} \subset \overline{M}_\infty^0$ with $\rho\left(0_{D^\infty(\mathbb{R})}\right) = 0$. Define the measure $\gamma_{\kappa,\rho}$ on $\overline{M}_\infty^0$ via

$$\gamma_{\kappa,\rho}(A) := \int_0^\infty \rho\left(\{G \ : \ G(s\cdot) \in A\}\right) \kappa(\mathrm{d}s).$$

Moreover, assume that $\gamma_{\kappa,\rho}$ satisfies

$$\int_{\overline{M}_\infty^0} G(t)\, \gamma_{\kappa,\rho}(\mathrm{d}G) = \int_{\overline{M}_\infty^0} \int_0^\infty G\left(\frac{t}{s}\right) \kappa(\mathrm{d}s)\rho(\mathrm{d}G) < \infty \text{ for all } t \geq 0.$$

Theorem 3.5 implies that

$$\mu_{\kappa,\rho} := \int \otimes_{i \in \mathbb{N}} \mathbb{P}_G\, \gamma_{\kappa,\rho}(\mathrm{d}G) \tag{7}$$

defines a valid global exponent measure on $E_\infty^{\mathbb{N}}$. The following proposition provides a series representation of the associated non-decreasing id-process.

Proposition 4.1 (*X* and *H* associated with $\gamma_{\kappa,\rho}$)

A series representation of the id-process H, associated with the exponent measure $\mu_{\kappa,\rho}$ in (7), is given by

$$(H_t)_{t \geq 0} = \left(\sum_{k \geq 1} - \log \left(1 - G_k \left(\frac{t}{S_k} \right) \right) \right)_{t \geq 0},$$

where $N := \sum_{k \geq 1} \delta_{(S_k, G_k)}$ denotes a Poisson random measure on $[0, \infty) \times \overline{M}_\infty^0$ with mean measure $\kappa \otimes \rho$. The survival function of the associated exchangeable min-id sequence X is given by

$$\mathbb{P}(X > t) = \exp \left(- \mathbb{E} \left[\kappa \left(\left[0, \max_{i \in \mathbb{N}} \frac{t_i}{Z_i} \right] \right) \right] \right), \tag{8}$$

where $Z = (Z_1, Z_2, \ldots)$ denotes an exchangeable sequence of random variables with associated random distribution function $G \sim \rho$.

Proof The proof can be found in Appendix 1. □

We find it educational to remark that the sequences $(S_k)_{k \in \mathbb{N}}$ and $(G_k)_{k \in \mathbb{N}}$ are independent and that $(G_k)_{k \in \mathbb{N}}$ is i.i.d. drawn from ρ and $S_k \sim f^{\leftarrow}(\epsilon_1 + \ldots + \epsilon_k)$, where (ϵ_k) are i.i.d. unit exponential and $f^{\leftarrow}(x) := \inf\{t \geq 0 \mid f(t) \geq x\}$ denotes the generalized inverse of the function $f(t) := \kappa([0, t])$.

Three prominent examples for the choice of the pair (κ, ρ) can be found in the literature.

4.3.1 Exchangeable min-stable sequences

Choosing $\kappa(ds) = ds$ yields

$$\gamma_{\kappa, \rho}(A) = \int_0^\infty \rho(\{G \ : \ G(s \cdot) \in A\}) ds.$$

By (Kopp and Molchanov 2018, Theorem 4.2), this defines the exponent measure of a driftless strong-idt process. This means that H satisfies $H_0 = 0$ and

$$\left(H_t \right)_{t \geq 0} \sim \left(\sum_{i=1}^n H_{\frac{t}{n}}^{(i)} \right)_{t \geq 0} \quad \text{for all } n \in \mathbb{N},$$

where $\left(H^{(i)} \right)_{i \in \mathbb{N}}$ are i.i.d. copies of H. The associated Lévy measure ν is given by

$$\nu(A) = \gamma_{\kappa, \rho} \left(\{G \mid G = 1 - \exp(-x(\cdot)) \text{ for some } x \in A\} \right).$$

Equation (8) simplifies to

$$\mathbb{P}(X > t) = \exp \left(- \mathbb{E} \left[\max_{i \geq 1} \frac{t_i}{Z_i} \right] \right), \tag{9}$$

which is a well known representation of the stable tail dependence function of an exchangeable min-stable sequence, see de Haan (1984). A result of Mai (2020) states that the probability law of $Y = (1/Z_1, 1/Z_2, \ldots)$ becomes unique, if we additionally postulate that ρ is concentrated on distribution functions G satisfying $\int_{(0,\infty]} s^{-1}\, \mathrm{d}G(s) = 1$. Survival functions of the form (9) are called min-stable multivariate exponential, since they imply $X \sim n \min_{i=1,\ldots,n}\{X^{(i)}\}$, where $n \in \mathbb{N}$ is arbitrary and $X^{(i)}$ denote independent copies of X. The min-stability property plays a fundamental role in multivariate extreme-value theory, since these are the only possible limiting distributions of componentwise minima of i.i.d. random vectors, after appropriate componentwise normalization, see Resnick (2007). Under the normalizing assumption $\mathbb{E}[1/Z_1] = 1$, Mai (2020) shows that the presented construction of the function $\ell(t) = \mathbb{E}[\max_{i \geq 1} t_i/Z_i]$ in (9) is general enough to comprise all possible stable tail dependence functions associated with exchangeable min-stable sequences. In other words, by (Mai and Scherer 2014, Theorem 5.3), all driftless nnnd strong-idt processes necessarily admit a series representation as in Proposition 4.1, where $S_k \sim \epsilon_1 + \ldots + \epsilon_k$ for an i.i.d. sequence of unit exponential random variables $(\epsilon_i)_{i \in \mathbb{N}}$ and $\kappa(\mathrm{d}s) = \mathrm{d}s$. However, it should be noted that not all finite-dimensional exchangeable min-stable random vectors can be obtained by this construction, since some exchangeable min-stable random vectors cannot be embedded into an exchangeable sequence (Mai and Scherer 2017, Example 3.2).

Regarding related examples from the literature let us mention α-idt processes, see Davydov et al. (2008), Hakassou and Ouknine (2012), aggregate self-similar processes, see Iglói and Barczy (2012), Barczy et al. (2015), and translatively stable processes, see Hakassou and Ouknine (2013). All these processes are strong-idt processes up to scaling and time change, which implies their infinite divisibility. As we have seen in Corollary 3.2 a deterministic time change of a non-decreasing strong-idt process solely changes the one dimensional marginal distribution of X, whereas a scaling of the strong-idt process corresponds to a scaling of the drift and a linear change of variables of the Lévy measure of the strong-idt process. Therefore, the nnnd instances of these processes uniquely correspond to an exchangeable min-id sequence X and their Lévy measure can be obtained as the image measure of the Lévy measure of a strong-idt process.

4.3.2 Reciprocal Archimedean copulas

Choose κ such that $\kappa(\{0\}) = 0$, $\kappa([0,\infty)) = \infty$, and $\rho = \delta_{\hat{G}}$, with $\hat{G}$ being the unit Fréchet distribution function $\hat{G}(t) = \exp(-1/t)\mathbf{1}_{\{t>0\}}$. This implies that

$$\gamma_{\kappa,\rho}(A) = \int_0^\infty \mathbf{1}_{\{\hat{G}(s\cdot)\in A\}}\kappa(\mathrm{d}s) = \kappa\left(\left\{ s \,\middle|\, \exp\left(-\frac{1}{s\cdot}\right)\mathbf{1}_{\{\cdot>0\}} \in A \right\}\right).$$

The associated exchangeable min-id sequence X has survival function

$$\mathbb{P}(X > t) = \exp\left(-\int_0^\infty 1 - \prod_{i\in\mathbb{N}}\left(1 - \exp\left(-\frac{s}{t_i}\right)\right)\kappa(\mathrm{d}s)\right), \tag{10}$$

which is exactly the representation of an exchangeable sequence with reciprocal Archimedean copula as introduced in Genest et al. (2018). Concretely, with the notation $\phi(t) = \int_0^\infty \exp(-tx)\kappa(\mathrm{d}x)$, we may rewrite

$$\mathbb{P}(X > t) = C_\phi\left(e^{-\phi(1/t_1)}, e^{-\phi(1/t_2)}, \dots\right),$$

where for $(u_1, u_2, \dots) \in [0,1]^{\mathbb{N}}$ the function

$$C_\phi(u_1, u_2, \dots) := \frac{\prod\limits_{\substack{A \subset \mathbb{N}, \\ |A| < \infty \text{ odd}}} \exp\left(-\phi\left(\sum_{k \in A} \phi^{-1}\left(-\log(u_k)\right)\right)\right)}{\prod\limits_{\substack{A \subset \mathbb{N}, \\ |A| < \infty \text{ even}}} \exp\left(-\phi\left(\sum_{k \in A} \phi^{-1}\left(-\log(u_k)\right)\right)\right)},$$

is the distribution function of an exchangeable sequence U and each component $U_i \sim \exp(-\phi(1/X_i))$ is uniformly distributed on [0,1]. The associated finite-dimensional margins of C_ϕ are called reciprocal Archimedean copulas, where the nomenclature is justified by some striking analogies with the concept of Archimedean copulas. For example, the Galambos copula is obtained by choosing

$$\kappa(\mathrm{d}s) = \frac{1}{\theta\,\Gamma(1 + 1/\theta)}\, s^{\frac{1}{\theta}-1}\, \mathrm{d}s$$

for some parameter $\theta > 0$, which yields $\phi(t) = t^{-1/\theta}$. We refer the interested reader to Genest et al. (2018) for more details about reciprocal Archimedean copulas.

4.3.3 Archimedean copulas with log-completely monotone generator

Let v_M denote the univariate Lévy measure of an infinitely divisible random variable $M \in (0, \infty)$. Define κ by $\kappa(A) := \int_0^\infty \mathbf{1}_A(1/s)v_M(\mathrm{d}s)$ for all $A \in \mathcal{B}([0, \infty))$ and choose $\rho = \delta_{\hat{G}}$, where $\hat{G}(t) = (1 - \exp(-t))\mathbf{1}_{\{t>0\}}$ denotes the distribution function of the unit exponential distribution. This implies that

$$\gamma_{\kappa,\rho}(A) = \int_0^\infty \mathbf{1}_{\{1-\exp(-\cdot/s)\in A\}} v_M(\mathrm{d}s)$$

and the associated exchangeable min-id sequence has survival function

$$\mathbb{P}(X > t) = \exp\left(-\mathbb{E}\left[\kappa\left(\left[0, \max_{i \in \mathbb{N}} \frac{t_i}{Z_i}\right]\right)\right]\right)$$

$$= \exp\left(-\int_0^\infty \int_{(0,\infty)^{\mathbb{N}}} 1 - \mathbf{1}_{\{z_i > st_i\ \forall\ i \in \mathbb{N}\}}\right.$$

$$\left. \otimes_{i \in \mathbb{N}} \mathbb{P}_{1-\exp(-\cdot)}(\mathrm{d}(z_1, z_2, \dots)) v_M(\mathrm{d}s)\right)$$

$$= \exp\left(-\int_0^\infty 1 - \exp\left(-s \sum_{i \in \mathbb{N}} t_i\right) v_M(\mathrm{d}s)\right),$$

which shows that X is associated to the nnnd id-process $(Mt)_{t\geq 0}$ with unique path Lévy measure

$$\nu\Big(\{x \in D^\infty(\mathbb{R}) \mid x(t) = a\,t\,1_{\{t\geq 0\}}, a \in A\}\Big) := \nu_M(A).$$

Moreover, X has marginal survival function $\psi(t) := \mathbb{E}\big[\exp(-tM)\big]$ and survival copula

$$C_\psi(\boldsymbol{u}) = \psi\left(\sum_{i\in\mathbb{N}} \psi^{-1}(u_i)\right), \ \boldsymbol{u} \in [0,1]^{\mathbb{N}},$$

which is an exchangeable Archimedean copula with completely monotone generator ψ according to Marshall and Olkin (1988), McNeil and Nešlehová (2009).

The definition of ψ in this context implies that ψ is even log-completely monotone, which means that it has the special representation $\psi(t) = \exp(-g(t))$, where $g : [0,\infty) \to [0,\infty)$; $g(t) = \int_0^\infty 1 - \exp(st)\nu_M(\mathrm{d}s)$ is called the Laplace exponent of ψ and g is a Bernstein function, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}g(t)$ is completely monotone. For example, the Gumbel copula, which is the only max-stable Archimedean copula, corresponds to $M \sim \alpha\text{-stable}$ with $\psi(t) = \exp\left(-t^\alpha\right)$, $\alpha \in (0,1]$, and $g(t) = t^\alpha$. The Lévy–Khintchine Theorem tells us that all log-completely monotone functions $\tilde{\psi}$ which satisfy $\tilde{\psi}(0) = 1$ are of the form $-\log\left(\tilde{\psi}(t)\right) = bt + \int_0^\infty 1 - \exp(-st)\nu_M(\mathrm{d}s)$ for some $b > 0$ and a Lévy measure ν_M of a non-negative infinitely divisible random variable M. Therefore, to obtain an exchangeable min-id sequence with exchangeable Archimedean survival copula with log-completely monotone generator $\tilde{\psi}$, we consider the exchangeable min-id sequence $X^{(\tilde{\psi})} \in (0,\infty)^{\mathbb{N}}$ corresponding to the extended chronometer $\left(H_t^{(\tilde{\psi})}\right)_{t\geq 0} := (bt + Mt)_{t\geq 0}$. We obtain that $X^{(\tilde{\psi})}$ has survival function

$$\mathbb{P}\left(X_1^{(\tilde{\psi})} > t_1, X_2^{(\tilde{\psi})} > t_2, \dots\right) = \mathbb{E}\left[\exp\left(-(M+b)\sum_{i\in\mathbb{N}} t_i\right)\right]$$
$$= C_{\tilde{\psi}}\big(\tilde{\psi}(t_1), \tilde{\psi}(t_2), \dots\big), \ \boldsymbol{t} \in [0,\infty)^{\mathbb{N}},$$

which shows that $X^{(\tilde{\psi})}$ has marginal survival function $\tilde{\psi}$ and Archimedean survival copula $C_{\tilde{\psi}}$.

It is worth noting that the Archimedean copula $C_{\tilde{\psi}}$ itself is max-id, since decreasing transformations of min-id sequences are max-id by Lemma 2.5. On the other hand, it is easy to see that the assumption of an Archimedean copula $C_{\tilde{\psi}}$ being max-id implies that $\tilde{\psi}$ corresponds to the Laplace transform of a non-negative infinitely-divisible random variable. Thus, an Archimedean copula on $[0,1]^{\mathbb{N}}$ is max-id if and only if $\tilde{\psi}$ is log-completely monotone.[5]

---

[5] Note that max-id Archimedean copulas are not the only positive lower orthant dependent (PLOD) Archimedean copulas. The set of completely monotone generators, a superset of log-completely monotone generators, corresponds to the class of extendible Archimedean copulas that are PLOD. Thus, a PLOD Archimedean copula is not necessarily max-id.

4.4 Subordination of Lévy processes by extended chronometers

Barndorff-Nielsen et al. (2006) prove that a Lévy process subordinated by an extended chronometer remains infinitely divisible. Formally, let $(L_t)_{t\geq0}$ denote a Lévy process and let $(H_t)_{t\geq0}$ denote an extended chronometer. If L is a subordinator $(Y_t)_{t\geq0} := (L_{H_t})_{t\geq0}$ defines an extended chronometer by (Barndorff-Nielsen et al. 2006, Theorem 7.1). It can be shown that Y has independent increments if H has independent increments and that Y is strong-idt if H is strong-idt.

Except for the quite simple example of (reciprocal) Archimedean copulas we have only seen examples of (killed) strong-idt and (killed) additive subordinators. An example of a càdlàg id-process which is (usually) neither strong-idt nor has independent increments is given by the non-negative solution to the Ornstein–Uhlenbeck type stochastic differential equation

$$\mathrm{d}V_t = -\lambda V_t \mathrm{d}t + \mathrm{d}Z_{\lambda t}, \tag{11}$$

where $\lambda > 0$ and $(Z_t)_{t\geq0}$ is a Lévy process satisfying

$$\int_{D^\infty\left([0,\infty)\right)} \max\{0, \log(|x(t)|)\} \nu_Z(\mathrm{d}x) < \infty$$

for all $t \geq 0$, where ν_Z denotes the Lévy measure of Z. In case $(Z_t)_{t\geq0}$ is a Lévy subordinator, V is the square of the stochastic volatility process of the Barndorff-Nielsen–Shepard model, see Barndorff-Nielsen and Shepard (2001), Carr et al. (2003). Moreover, the process $H_t := \int_0^t V_s \mathrm{d}s$ remains infinitely divisible (Barndorff-Nielsen et al. 2006, Sect. 4). H_t is known as the integrated or cumulated volatility up to time t, which is important when analyzing the realized variance or the quadratic variation of an option pricing model, see Kömm (2016). For instance, Duan and Yeh (2010) model the CBOE Volatility Index at time t via $const. + \frac{1}{t}\mathbb{E}_Q\left[\int_0^t V_s \mathrm{d}s\right]$, where $\mathbb{E}_Q$ denotes the expectation w.r.t. some risk neutral probability measure Q. Interestingly, we can generalize the ideas of Barndorff-Nielsen and Shepard (2001) and (Mansuy 2005, Example (2.2)) to id-processes to obtain that

$$\left(H_t^{(\kappa)}\right)_{t\geq0} := \left(\int_0^t V_s \kappa(\mathrm{d}s)\right)_{t\geq0}$$

defines an extended chronometer for every non-negative càdlàg id-process V and Radon measure κ on $[0, \infty)$, which is usually neither strong-idt nor has independent increments.

Proposition 4.2 (Extended chronometer via integration over non-negative id-process)

Let $(V_s)_{s\geq0} \in D([0, \infty))$ denote a non-negative càdlàg id-process with drift b_V and Lévy measure ν_V. Moreover, let κ denote a measure on $[0, \infty)$ such that $\kappa([0,t]) < \infty$ for all $t \geq 0$. Then,

$$\left(H_t^{(\kappa)}\right)_{t\geq 0} := \left(\int_0^t V_s\kappa(\mathrm{d}s)\right)_{t\geq 0}$$

defines an extended chronometer with Lévy measure

$$\nu_\kappa(A) := \nu_V\left(\left\{x \in D^\infty\left([0,\infty)\right) \;\middle|\; \int_0^{\cdot} x(s)\kappa(\mathrm{d}s) \in A, \int_0^\infty x(s)\kappa(\mathrm{d}s) > 0\right\}\right)$$

for all $A \in \mathcal{B}(D^\infty([0,\infty)))$ and drift $b^{(\kappa)}(\cdot) = \int_0^{\cdot} b_V(s)\kappa(\mathrm{d}s)$.

Proof The proof can be found in Appendix 1. $\square$

If we subordinate a Lévy subordinator L by $H^{(\kappa)}$ we can obtain an explicit representation of the survival function of the associated exchangeable min-id sequence $X^{(L,V,\kappa)}$. Let $\psi(a) := -\log\left(\mathbb{E}\left[\exp\left(-aL_1\right)\right]\right)$ denote the Laplace exponent of L. The min-id sequence $X^{(L,V,\kappa)}$ associated with $\left(L_{H_t^{(\kappa)}}\right)_{t\geq 0}$ has the following survival function:

$$\mathbb{P}\left(X_1 > t_1, \ldots, X_d > t_d\right) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^d L_{H_{t_i}^{(\kappa)}}\right)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\exp\left(-\sum_{i=1}^d L_{H_{t_i}^{(\kappa)}}\right)\middle| H\right]\right]$$

$$= \mathbb{E}\left[-\exp\left(\sum_{i=1}^d H_{t_i}^{(\kappa)}\left(\psi(d-i+1) - \psi(d-i)\right)\right)\right]$$

$$= \exp\left(-\sum_{i=1}^d \left(\psi(d-i+1) - \psi(d-i)\right)b_{t_i}^{(\kappa)}\right.$$

$$\left. - \int_{M_\ell} 1 - \exp\left(-\sum_{i=1}^d \left(\psi(d-i+1) - \psi(d-i)\right)x_{t_i}^{(\kappa)}\right)\nu_V(\mathrm{d}x)\right),$$

where $x_t^{(\kappa)} := \int_0^t x(s)\kappa(\mathrm{d}s)$. Thus, if ν_V and ψ are known, we obtain an explicit analytic representation of the survival function of $X^{(L,V,\kappa)}$. In particular this is the case if we start with "simple" processes L and V. E.g. choosing V as a Cox–Ingersoll–Ross process, see Cox et al. (1985) yields an explicit representation of the surival function of $X^{(L,V,\kappa)}$, since the Lévy measure of such processes can be obtained by an application of Proposition 4.2 to a scaled and time changed squared Bessel process, see (Rosiński 2018, Example 2.24) for more details on the Lévy measure of squared Bessel processes.

Generally, this approach yields a flexible way to combine two "simple" extended chronometers to an extended chronometer which is usually neither strong-idt nor additive.

4.5 Finite exponent measures

Let $X \in (-\infty, \infty]^{\mathbb{N}}$ denote a min-id sequence with $0 < \mathbb{P}(X = \infty) < 1$. The associated global exponent measure μ is a finite measure on $E_{\infty}^{\mathbb{N}}$ with total mass $c := -\log(\mathbb{P}(X = \infty))$. (Resnick 2007, Example 5.6) shows that there exists a sequence of i.i.d. sequences $(\mathbf{Z}^{(i)})_{i \in \mathbb{N}} \in (-\infty, \infty]^{\mathbb{N} \times \mathbb{N}}$ with $\mathbf{Z}^{(1)} \sim \mathbf{Z} \sim \mu/c =: \tilde{\mathbb{P}}$ and an independent Poisson random variable N with mean c such that $\min_{1 \leq i \leq N} \mathbf{Z}^{(i)} \sim X$. This can be easily verified by

$$
\mathbb{P}\left(\min_{1 \leq i \leq N} \mathbf{Z}^{(i)} > x\right) = \exp(-c) \sum_{i \in \mathbb{N}} \frac{c^i \tilde{\mathbb{P}}(\mathbf{Z} > x)^i}{i!} = \exp\left(-c\left(1 - \tilde{\mathbb{P}}\left((x, \infty]\right)\right)\right)
$$
$$
= \exp\left(-\mu\left((x, \infty]^{\complement}\right)\right) = \mathbb{P}(X > x).
$$

Obviously, X is exchangeable if and only if Z is exchangeable. Then, Theorem 3.5 tells us that μ can be decomposed into $\mu = \mu_b + \mu_\gamma$. Note that the existence of some $t \in \mathbb{R}$ such that

$$
\mu_{b,n}\left(\left(\underbrace{(t, \ldots, t)}_{n \text{ times}}, \infty\right]^{\complement}\right) = n b(t) \xrightarrow{n \to \infty} \infty
$$

is equivalent to $b \neq 0_{D^{\infty}(\mathbb{R})}$. Therefore, if μ is finite, $\mu_b = 0$ and the associated id-process H is driftless. Thus, μ is given by $\mu = \mu_\gamma = \int_{\overline{M}_{\infty}^0} \otimes_{i \in \mathbb{N}} \mathbb{P}_G \gamma(dG)$. Furthermore, Theorem 3.5 implies the existence of a unique Lévy measure ν such that $\nu(A) = \gamma(\{G \in \overline{M}_{\infty}^0 \mid G = 1 - \exp(-x(\cdot)) \text{ for some } x \in A\})$. An application of the monotone convergence theorem shows that

$$
c = \mu\left(E_{\infty}^{\mathbb{N}}\right) = \lim_{t \to \infty} \mu\left((t, \infty]^{\complement}\right) = \lim_{t \to \infty} \int_{M_{\infty}^0} 1 - \prod_{i \in \mathbb{N}} \exp(-x(t)) \nu(dx)
$$
$$
\underset{x \neq 0}{=} \nu\left(M_{\infty}^0\right),
$$

i.e. ν is finite as well. Now, Example 2.15 implies that the associated (driftless) id-process H is given by

$$
H \sim \sum_{i=1}^{N} h_i,
$$

where $(h_i)_{i \in \mathbb{N}}$ are i.i.d. stochastic processes on M_{∞}^0 with distribution ν/c and N is an independent Poisson random variable with mean c. Therefore, if μ is finite, X (resp. H) admits a simple construction method via Poisson maxima (resp. sums) of i.i.d. objects.

The correspondence of the exchangeable sequence Z with the finite exponent measure μ seems to imply that finite extendible exponent measures μ_d can be represented via extendible random vectors. However, as the following

paragraphs show, this is not always possible, since the global exponent measure μ can have infinite mass even if μ_d is finite for every $d \in \mathbb{N}$. To see this, assume that $X_d \in (-\infty, \infty]^d$ is exchangeable and min-id with finite exponent measure μ_d. Moreover, assume that X_d is extendible to an exchangeable min-id sequence X, which w.l.o.g. satisfies Condition ($\Diamond$). This ensures the existence of a global exponent measure μ such that μ_d satisfies the properties in Proposition 2.9.

(Resnick 2007, Example 5.6) and Proposition 2.8 imply that there exists a Poisson random variable N_d with mean $c_d = \mu_d(E_\infty^d)$ and i.i.d. exchangeable random vectors $(\mathbf{Z}_d^{(i)})_{i \in \mathbb{N}} \in (-\infty, \infty]^d$ with distribution μ_d/c_d such that $X_d \sim \min_{1 \le i \le N_d} \mathbf{Z}_d^{(i)}$. Therefore, μ_d can always be represented by a random vector $\mathbf{Z}_d^{(1)} \sim \mathbf{Z}_d \sim \mu_d/c_d$ and a constant c_d.

In the elaborations above we have seen that a finite global μ implies that $\mathbf{Z}_d$ is extendible to a sequence $\mathbf{Z}$. Since we know that μ_d is extendible to a global μ, it would be tempting to assume that $\mathbf{Z}_d$ is also extendible to a sequence $\mathbf{Z}$, independent of the total mass of μ. However, as the following calculations show, μ being finite is also a necessary condition for $\mathbf{Z}_d$ being extendible.

We begin with an analysis of μ_d. Independently of the total mass of μ it is always possible to decompose μ_d into $\mu_{b,d} + \mu_{\gamma,d}$ by Theorem 3.5. Therefore, we can define the constant $a_d = \mu_{\gamma,d}(E_\infty^d)/c_d = 1 - \mu_{b,d}(E_\infty^d)/c_d \in [0, 1]$. Now, μ_d can be generated as follows:

1. Draw a Bernoulli random variable B with success probability a_d.
2. If $B = 1$, draw a random variable $Y^{(d,1)}$ with distribution $\mu_{\gamma,d}(\cdot)/(c_d a_d)$.
3. If $B = 0$, draw a random variable $Y^{(d,2)}$ with distribution $\mu_{b,d}(\cdot)/(c_d(1 - a_d))$. Note that $Y^{(d,2)}$ is supported on $\{x \in (-\infty, \infty]^d \mid x_i = \infty \text{ for all but one } i\}$.
4. $\mu_d(A) = c_d \mathbb{P}(BY^{(d,1)} + (1 - B)Y^{(d,2)} \in A)$.

Taking a closer look at $Y^{(d,2)}$ reveals that $Y^{(d,2)}$ can never be embedded in a random vector on $(-\infty, \infty]^{d'}$, $d' > d$, if we do not allow for mass on $\times_{i=1}^d \{\infty\}$. Unfortunately, even if we allow for mass on $\times_{i=1}^d \{\infty\}$, an embedding of $Y^{(d,2)}$ in a sequence requires $\mu_d(\{\infty\}^d) = \infty$ and $\mu(E_\infty^{\mathbb{N}}) = \infty$, since $\mu_{b,d}(E_\infty^d) = d \lim_{t \to \infty} b(t) \xrightarrow{d \to \infty} \infty$ if and only if $b \ne 0_{D^\infty(\mathbb{R})}$. Therefore, $Y^{(d,2)}$ can never be extended to a sequence and the random variable $\mathbf{Z}_d = Y^{(d,1)} + Y^{(d,2)}$ can only be extendible if $a_d = 1$. Note that the necessity of $a_d = 1$ is not based on the fact that $Y^{(d,2)}$ is not extendible, but rather on the fact that the presence of $Y^{(d,2)}$ implies that the only possible extension of the distribution of $Y^{(d,2)}$ is an infinite measure. Intuitively, this resembles the fact that X cannot have finite exponent measure μ if it can be represented as the minimum of a non-trivial i.i.d. sequence (represented by $Y^{(d,2)}$) and an independent exchangeable min-id sequence (represented by $Y^{(d,1)}$).

It remains to investigate under which circumstances $Y^{(d,1)}$ is extendible. Observe that in case $a_d = 1$ the extendibility of $Y^{(d,1)}$ to a sequence $Y^{(1)}$ implies that μ is finite. Therefore, μ being finite is not only a sufficient but also necessary criterion for the extendibility of $\mathbf{Z}_d$. Note that the distribution of the first d components of $Y^{(1)}$ may not be exactly μ_d/c_d due to the removal of $\times_{i=1}^d \{\infty\}$ and the possibility of

$Y_1^{(1)} = \ldots = Y_d^{(1)} = \infty$. However, the distribution of Z_d can be obtained as the conditional distribution of $(Y_1^{(1)}, \ldots, Y_d^{(1)})$ given $(Y_1^{(1)}, \ldots, Y_d^{(1)}) \neq \infty$.

Our discussion is summarized in the following paragraph. Let $(E_i)_{i \in \mathbb{N}}$ denote a sequence of unit exponential random variables. The global extensions μ of μ_d can be classified into two possible cases:

1. Finite μ: In this case Z_d is extendible to an exchangeable sequence Z and the global exponent measure μ is in one-to-one correspondence with the tuple $(Z, \mathbb{P}(X = \infty))$. It is easy to see that the associated extended chronometer H is driftless with $\mathbb{P}(H = 0_{D^\infty(\mathbb{R})}) > 0$.
2. Infinite μ: In this case Z_d is not extendible. Nevertheless, we know that μ_d is extendible to a global μ. Thus, $\mathbb{P}(X = \infty) = \exp(-\mu(E_\infty^{\mathbb{N}})) = 0$ and the associated extended chronometer H satisfies

$$\mathbb{P}(H = 0_{D^\infty(\mathbb{R})}) = \mathbb{P}\left(\left\{E_i > \lim_{t \to \infty} H_t \text{ for all } i \in \mathbb{N}\right\}\right) = \mathbb{P}(X = \infty) = 0.$$

Additionally, the drift of H satisfies $\lim_{t \to \infty} b(t) < \infty$, since $\lim_{t \to \infty} b(t) = \infty$ would require that $\mathbb{P}(X_d = \infty) \leq \lim_{t \to \infty} \exp(-db(t)) = 0$. Moreover, we can deduce that

$$0 < \mathbb{P}(X_d = \infty) = \mathbb{P}\left(\left\{E_i > \lim_{t \to \infty} H_t \text{ for all } 1 \leq i \leq d\right\}\right)$$
$$\leq \mathbb{P}\left(0 < \lim_{t \to \infty} H_t < \infty\right).$$

Therefore, H is almost surely non-zero and bounded with positive probability.

4.6 Approximation of extended chronometer via extended chronometers with finite Lévy measure

Example 2.15 shows that a driftless extended chronometer H with finite Lévy measure ν may be simulated via finitely many i.i.d. copies $(h_i)_{i \in \mathbb{N}}$ of a stochastic process $h \sim \nu(\cdot)/\nu(D^\infty(\mathbb{R}))$. H may then be represented as $(H_t)_{t \in \mathbb{R}} \sim \left(\sum_{i=1}^{\tilde{N}} h_i(t)\right)_{t \in \mathbb{R}}$, where $\tilde{N}$ denotes a Poisson random variable with mean $\nu(D^\infty(\mathbb{R}))$.

In case ν is an infinite measure, Proposition 2.16 implies that there exists a Poisson random measure $N := \sum_{i \in \mathbb{N}} \delta_{h_i}$ on $\{x \in D^\infty(\mathbb{R}) \mid x \text{ nnnd}\}$ with intensity measure ν such that

$$(H_t)_{t \in \mathbb{R}} := \left(\sum_{i \in \mathbb{N}} h_i(t)\right)_{t \in \mathbb{R}}$$

defines an nnnd id-process with Lévy measure ν by (Rosiński 2018, Proposition 2.10). Clearly, for every $s \in \mathbb{R}$ and $\epsilon > 0$, the extended chronometer defined via

$$\left(H_t^{(s,\epsilon)}\right)_{t\in\mathbb{R}} := \left(\sum_{\substack{i\in\mathbb{N}\\ h_i(s)>\epsilon}} h_i(t)\right)_{t\in\mathbb{R}},$$

has finite Lévy measure $\nu_{s,\epsilon} := \nu(\cdot \cap \{x(s) > \epsilon\})$. This implies that the nnnd id-process

$$\left(H_t^{(s)}\right)_{t\in\mathbb{R}} := \left(\lim_{\epsilon\to 0} H_t^{(s,\epsilon)}\right)_{t\in\mathbb{R}} = \left(\sum_{\substack{i\in\mathbb{N}\\ h_i(s)>0}} h_i(t)\right)_{t\in\mathbb{R}}$$

has Lévy measure $\nu_{s,0} := \nu(\cdot \cap \{x(s) > 0\})$, which is generally an infinite measure that is not equal to ν. However, it is important to observe that

$$\left(H_t\right)_{t\leq s} = \left(H_t^{(s)}\right)_{t\leq s}.$$

Thus, if one is only interested in the path of the extended chronometer H up to time s, one may simulate $H^{(s,\epsilon)}$ for some small $\epsilon > 0$ to obtain an approximation of $\left(H_t\right)_{t\leq s}$.[6]

The just described approximation procedure of an extended chronometer is well-known when H is an additive process. In this case, a Poisson random measure with intensity $\nu_{s,\epsilon}$ may be represented as $\left(A_i \mathbf{1}_{\{\cdot \geq U_i\}}\right)_{1\leq i\leq \tilde{N}}$, where $\tilde{N}$ denotes a Poisson random variable with mean $\nu_{s,\epsilon}(D^\infty(\mathbb{R}))$ and $\left((U_i, A_i)\right)_{1\leq i\leq \tilde{N}}$ denotes a finite Poisson random measure on $[0,s]\times(\epsilon,\infty]$. This implies that $\left(H_t^{(s,\epsilon)}\right)_{t\in\mathbb{R}} = \left(\sum_{1\leq i\leq \tilde{N}} A_i \mathbf{1}_{\{t\geq U_i\}}\right)_{t\in\mathbb{R}} = \left(\sum_{U_i\leq t} A_i\right)_{t\in\mathbb{R}}$, which is a well-known representation of a (inhomogeneous) compound Poisson process. However, note that nnnd id-processes with independent increments are the only nnnd id-processes where $H^{(s,\epsilon)}$ can be represented as a (inhomogeneous) compound Poisson processes. This is due to the fact that (inhomogeneous) compound Poisson processes have independent increments and the independent increments property is preserved under the weak convergence of the finite dimensional margins of $H^{(s,\epsilon)}$ when $\epsilon \to 0$.

The exchangeable min-id sequences $X^{(s,\epsilon)}$, $X^{(s)}$, and X associated to $H^{(s,\epsilon)}$, $H^{(s)}$, and H obviously satisfy $\mathbb{P}\left(X^{(s,\epsilon)} > t\right) \geq \mathbb{P}\left(X^{(s,0)} > t\right) \geq \mathbb{P}(X > t)$ for all $t \in [-\infty,\infty)^{\mathbb{N}}$. Moreover, $\mathbb{P}\left(X^{(s,0)} > t\right) = \mathbb{P}(X > t)$ for all $t \in [-\infty, s]^{\mathbb{N}}$, since $\left(H^{(s)}\right)_{t\leq s} = \left(H_t\right)_{t\leq s}$. Therefore, one may approximate the exchangeable min-id sequence X associated to the extended chronometer H via a simulation of $X^{(s,\epsilon)}$ associated to the extended chronometer $H^{(s,\epsilon)}$ with finite Lévy measure.

---

[6] A similar reasoning also allows to approximate a general id-process by id-processes with finite Lévy measure.

4.7 Overview of established families under the present umbrella

Figure 3 provides a graphical overview of the established families of exchangeable min-id sequences introduced in this paper.

Section 4.2 establishes that the class of extendible exogenous shock models is contained in the class of exchangeable min-id sequences. Since extendible exogenous shock models are associated to the class of nnnd additive processes according to Mai et al. (2016) and Sloot (2020), we deduce that the class of extendible Marshall–Olkin distributions is contained in the class of extendible exogenous shock models as the class of extendible Marshall–Olkin distributions corresponds to the class of nnnd Lévy processes according to Mai (2010). Furthermore, the class of exchangeable Sato-frailty sequences is included in the class of extendible exogenous shock models as it is associated to the class of (additive) Sato-subordinators as shown in Mai et al. (2017). Moreover, the class of exchangeable Sato-frailty sequences only intersects with the class of Marshall–Olkin distributions when the associated Sato-subordinator is the α-stable subordinator, which is the only Sato-subordinator that is also a Lévy process according to (Sato 1999, Remark 16.2).

Section 4.3.1 shows that the class of exchangeable min-stable sequences with exponential margins is contained in the class of exchangeable min-id sequences and Mai and Scherer (2014) show that it corresponds to the class of nnnd strong-idt processes. Since the class of nnnd Lévy processes is precisely the intersection
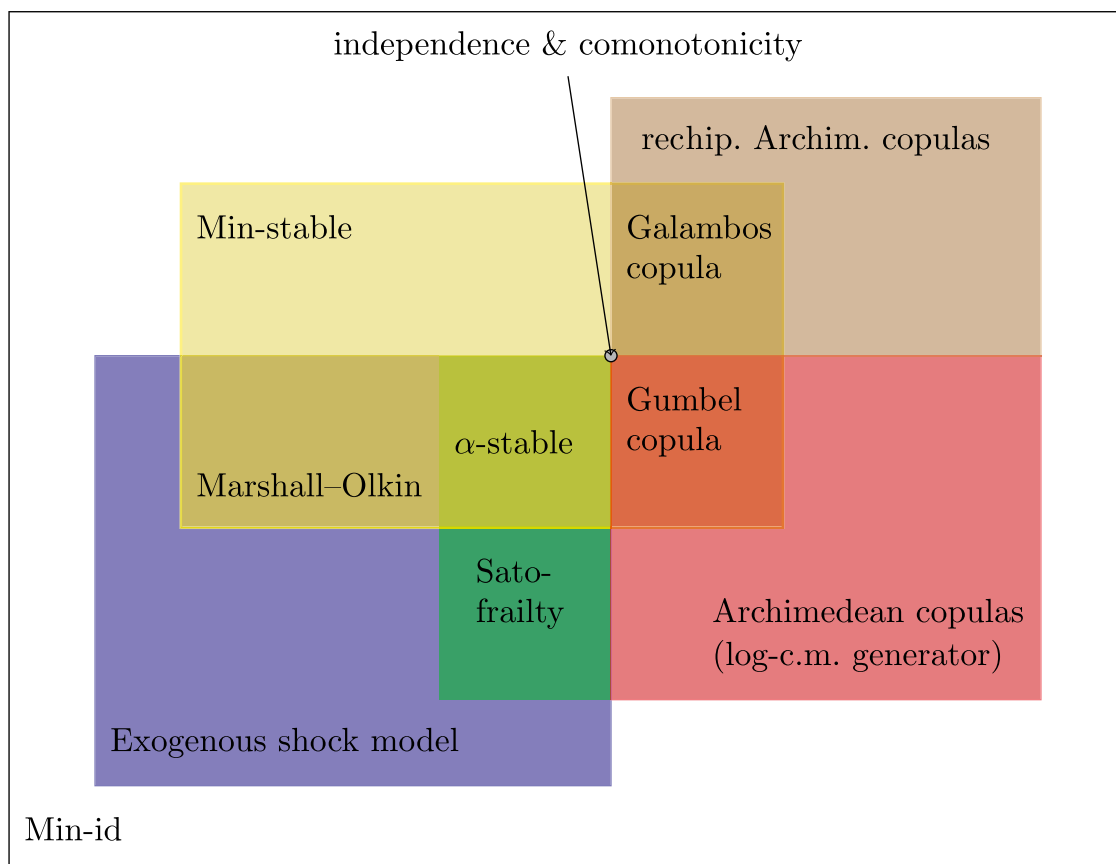
Fig. 3 Overview of established exchangeable min-id sequences under the present umbrella

of the classes of nnnd additive processes and nnnd strong-idt processes, the class of extendible Marshall–Olkin distributions is precisely the intersection of the class of extendible exogenous shock models and the class of exchangeable min-stable sequences with exponential margins. Furthermore, Genest et al. (2018) show that the Galambos copula is the only copula which belongs both to the class of copulas of exchangeable min-stable sequences (also called extreme value copulas) and to the class of reciprocal Archimedean copulas. Similarly, Genest and Rivest (1989) show that the Gumbel copula is the only copula which belongs both to the class of extreme value copulas and to the class of Archimedean copulas.

The class of extendible exogenous shock models does not intersect with the class of Archimedean copulas with log-completely monotone generator except for independence and comonotonicity, since the nnnd id-process associated to an Archimedean copula with log-completely monotone generator cannot have independent increments according to Sect. 4.3.3. A similar statement applies to reciprocal Archimedean copulas according to Sect. 4.3.2. Comparing the Lévy measures associated to reciprocal Archimedean copulas and Archimedean copulas with log-completely monotone generator, we obtain that their intersection can only consist of independence and comonotonicity as well.

5 Conclusion

We have shown that every exchangeable min-id sequence is in one-to-one correspondence with an nnnd infinitely divisible càdlàg process. Doing so, we have unified the work of Mai (2010), Mai and Scherer (2013, 2014), Mai et al. (2016, 2017), Sloot (2020) under one common umbrella. Furthermore, we have shown that the exponent measure of an exchangeable min-id sequence associated to a driftless nnnd id-process is a mixture of product probability measures. Therefore, de Finetti's Theorem is extended to exchangeable exponent measures. Several important examples of exchangeable min-id sequences have been presented and the existing literature has been embedded into our framework. A summary of these correspondences is given in Figs. 1, 2, and 3. As a byproduct we have shown that càdlàg id-processes can be represented as the sum of arbitrarily many i.i.d. càdlàg processes and that the Lévy measure of nnnd càdlàg id-processes is concentrated on nnnd càdlàg functions.

There are various well known subclasses of nnnd infinitely divisible stochastic processes, such as nnnd additive processes and strong-idt processes. However, it remains an interesting open problem to find nnnd infinitely divisible stochastic processes outside of these subclasses, which can be conveniently described analytically. Moreover, even if one would have an analytical characterization of such nnnd id-processes at hand, their simulation would probably only be feasible in an approximate fashion, similar to the approach discussed in Sect. 4.6. Nonetheless, even their approximate simulation would most likely still pose a quite challenging problem due to possible path dependencies of the involved stochastic processes. Thus, the simulation of the associated exchangeable min-id sequence remains challenging is such situations.

To circumvent this problem, Brück (2022) provides a simulation algorithm for the class of real-valued continuous max-id-processes, which ensures that a user-specified

number of locations of the max-id-process are simulated exactly. Since every min-id sequence X may be transformed into a continuous max-id-process $1/X$ with index set $\mathbb{N}$, the results of Brück (2022) may be applied to simulate an exchangeable min-id sequence X. The key ingredient of the proposed simulation algorithm in Brück (2022) is the exponent measure of X (or $1/X$), which can be easily constructed according to (4) or deduced from the Lévy measure of the associated nnnd id-process by Theorem 3.5. Thus, the results of this paper allow to construct exchangeable min-id sequences in terms of their associated exponent or Lévy measures, while Brück (2022) provides a method for the exact simulation of their d-dimensional margins.

Appendix 1

Proofs

Lemma 1.1 *The following properties are valid*:

1. $\pi(x) \in A \Leftrightarrow x \in \pi^{-1}(A)$ *and* $\pi^{-1}(x) \in A \Leftrightarrow x \in \pi(A)$.
2. $\pi(\pi^{-1}(A)) = A$.
3. $\pi(A) = \{\pi(x) \mid x \in A\} = \{y \mid \pi^{-1}(y) \in A\} = \{y \mid y \in \pi(A)\}$.
4. $\bar{\mathbb{R}}^d \setminus \pi(A) = \pi(\bar{\mathbb{R}}^d \setminus A)$ *and* $\pi(B) \setminus \pi(A) = \pi(B \setminus A)$.
5. $\pi(\cup_{i\in\mathbb{N}} A_i) = \cup_{i\in\mathbb{N}} \pi(A_i)$.
6. *Let E_ℓ be the support of an exchangeable exponent measure. Obviously, $E_\ell = \pi(E_\ell)$ and $\pi\left((x, \infty]^{\complement}\right) = (\pi(x), -\infty]^{\complement}$.*

Proof (Proof of Lemma 1.1)

   1. + 2. Obvious.
   3. Follows from 1.
   4. $\bar{\mathbb{R}}^d \setminus \pi(A) = \{x \mid x \notin \pi(A)\} = \{x \mid \pi^{-1}(x) \notin A\} = \{x \mid \pi^{-1}(x) \in \bar{\mathbb{R}}^d \setminus A\} = \pi(\bar{\mathbb{R}}^d \setminus A)$. The second assertion follows analogously.
   5. $\pi(\cup_{i\in\mathbb{N}} A_i) = \{x \mid \pi^{-1}(x) \in A_i \text{ for some } i \in \mathbb{N}\} = \{x \mid x \in \pi(A_i) \text{ for some } i \in \mathbb{N}\} = \cup_{i\in\mathbb{N}} \pi(A_i)$.
   6. Since E_ℓ is of the form $[-\infty, \ell] \setminus \ell$ with $\ell = (\ell, \ldots, \ell)$ we get $\pi\left((x, \infty]^{\complement}\right) = \pi\left(E_\ell \setminus (x, \infty]\right) = \{\pi(y) \mid y \in E_\ell, y_i \leq x_i \text{ for some } i\} = \{y \mid y \in E_\ell, y_i \leq \pi(x_i) \text{ for some } i\} = (\pi(x), \infty]^{\complement}$. $\qquad\square$

Proof of Corollary 2.3

Proof

"$\Leftarrow$" Assume that H is stochastically continuous. From the property that H is non-decreasing and càdlàg with $\lim_{s \nearrow t} H_s = H_t$ in probability, we obtain that $\lim_{s \nearrow t} H_s = H_t$ almost surely. This is due to the fact that there is at least one sequence $(s_n)_{n\in\mathbb{N}}$ with $s_n \nearrow t$ such that $\lim_{n\to\infty} H_{s_n} = H_t$ almost surely, since $\lim_{s \nearrow t} H_s = H_t$ in probability. This

immediately implies that every $(s'_n)_{n\in\mathbb{N}}$ with $s'_n \nearrow t$ satisfies $\lim_{n\to\infty} H_{s'_n} = H_t$ almost surely due to the non-decreasingness of H. Therefore, the Laplace transforms of $\lim_{s\nearrow t} H_s$ and H_t coincide, which implies that

$$\mathbb{P}(X_1 \geq t) = \lim_{s\nearrow t} \mathbb{P}(X_1 > s) = \lim_{s\nearrow t} \mathbb{E}\left[\exp(-H_s)\right] = \mathbb{E}\left[\exp(-H_t)\right] = \mathbb{P}(X_1 > t).$$

Thus, the distribution of X is continuous.

"$\Rightarrow$" Assume that X follows a continuous distribution. Seeking a contradiction, we assume that H is not stochastically continuous, i.e. there exists $t \in \mathbb{R}$ and $\delta, \epsilon > 0$ such that $\lim_{s\nearrow t} \mathbb{P}(H_t - H_s > \delta) > \epsilon$. Note that the limit exists, because H is increasing. Now, there exist $0 < q_1 < q_1 + \delta < q_2$ and $\eta > 0$ such that $\lim_{s\nearrow t} \mathbb{P}(H_s < q_1, H_t \geq q_2) > \eta$. This implies that

$$\mathbb{P}\left(X_1 = t\right) \geq \mathbb{P}\left(E_1 \in (q_1, q_2), \lim_{s\nearrow t} H_s < q_1, H_t \geq q_2\right) > \eta\mathbb{P}\left(E_1 \in (q_1, q_2)\right) > 0,$$

which is a contradiction. Therefore, H must be stochastically continuous. $\qquad\square$

Proof of Proposition 2.8

Proof Throughout the proof we frequently use properties of the permutation operator π derived in Lemma 1.1.

"$\Leftarrow$" The exchangeability of μ_d translates into the exchangeability of the survival function $\overline{F}$ of X_d, which in turn implies that X_d is exchangeable. We provide the precise reasoning behind this argument, since we need to carry out the same steps for finite exponent measures in "$\Rightarrow$". For ease of notation we write $\mathbb{P}(X_d \in A) =: \mathbb{P}(A)$. Let

$$\mathcal{A} := \left\{A \in \mathcal{B}\left((-\infty, \infty]^d\right) \mid \mathbb{P}(\pi(A)) = \mathbb{P}(A) \ \forall \text{ permutations } \pi \text{ on } \{1, \dots, d\}\right\}$$

denote the collection of $\mathbb{P}$-exchangeable sets. We show that $\mathcal{A}$ is a Dynkin system containing a Π-stable generator of $\mathcal{B}((-\infty, \infty]^d)$, which implies that $\mathcal{B}((-\infty, \infty]^d) \subset \mathcal{A}$. Obviously, the sets $(-\infty, \infty]^d$ and $\emptyset$ are both included in $\mathcal{A}$. Now consider some arbitrary set $A \in \mathcal{A}$. We get

$$\begin{aligned}\mathbb{P}\left((-\infty, \infty]^d \setminus A\right) &= 1 - \mathbb{P}(A) = 1 - \mathbb{P}(\pi(A)) = \mathbb{P}\left((-\infty, \infty]^d \setminus \pi(A)\right) \\ &= \mathbb{P}\left(\pi((-\infty, \infty]^d \setminus A)\right).\end{aligned} \tag{12}$$

Therefore, $(-\infty, \infty]^d \setminus A \in \mathcal{A}$. Next, consider $(A_i)_{i\in\mathbb{N}} \in \mathcal{A}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$. Using that every measure is continuous from below we obtain

$$\mathbb{P}\big(\cup_{i\in\mathbb{N}}A_i\big) = \lim_{n\to\infty}\mathbb{P}\big(\cup_{i=1}^n A_i\big) = \lim_{n\to\infty}\mathbb{P}\big(\cup_{i=1}^n\pi(A_i)\big)$$
$$= \mathbb{P}\big(\cup_{i\in\mathbb{N}}\pi(A_i)\big) = \mathbb{P}\big(\pi\big(\cup_{i\in\mathbb{N}}A_i\big)\big),$$

which establishes that $\mathcal{A}$ is a Dynkin system. From the exchangeability of μ_d we deduce that

$$\mathbb{P}\big((x,\infty]\big) = \overline{F}(x) = \exp\big(-\mu\big((x,\infty]^{\complement}\big)\big) = \exp\big(-\mu\big(\pi\big((x,\infty]^{\complement}\big)\big)\big)$$
$$= \exp\big(-\mu\big((\pi(x),\infty]^{\complement}\big)\big) = \overline{F}(\pi(x)) = \mathbb{P}\big((\pi(x),\infty]\big)$$
$$= \mathbb{P}\big(\pi((x,\infty])\big).$$

Therefore, $\mathcal{A}$ contains a Π-stable generator of $\mathcal{B}((-\infty,\infty]^d)$. This implies that $\mathbb{P}(A) = \mathbb{P}(\pi(A))$ for all $A \in \mathcal{B}((-\infty,\infty]^d)$ and all permutations π on $\{1,\dots,d\}$, which proves that $\overline{F}$ is the survival function of an exchangeable random vector. "$\Rightarrow$" If μ_d is a finite measure, the proof is simply a repetition of the arguments in "$\Leftarrow$" to show that

$$\mathcal{A} := \big\{A \in \mathcal{B}\big(E_\infty^d\big) \mid \mu_d(\pi(A)) = \mu_d(A)\ \forall\ \text{permutations}\ \pi\ \text{on}\ \{1,\dots,d\}\big\}$$

is a Dynkin system containing a Π-stable generator of $\mathcal{B}\big(E_\infty^d\big)$, since

$$\mu\big((x,\infty]^{\complement}\big) = -\log\big(\mathbb{P}\big((x,\infty]\big)\big) = -\log\big(\mathbb{P}\big(\pi((x,\infty])\big)\big)$$
$$= \mu\big((\pi(x),\infty]^{\complement}\big) = \mu\big(\pi\big((x,\infty]^{\complement}\big)\big).$$

Therefore, we may focus on the case $\mu_d(E_\infty^d) = \infty$. In this case the sets with infinite measure have non-empty intersection with an open neighborhood of ∞ implying μ_d to be σ-finite. Note that a similar reasoning as in (12) cannot be applied here, since $\mu_d\big(E_\infty^d\big) = \infty$, which is why we have to resort to a different reasoning to prove that μ_d is exchangeable. Instead, our goal is to show that μ_d is the pointwise limit of a sequence of exchangeable finite measures. The following paragraphs are dedicated to the proof of this statement. Let $(c_n)_{n\in\mathbb{N}} \in \mathbb{R}$, $c_n < \infty$ with $c_n \nearrow \infty$. Define

$$\mu^{(n)}(\cdot) := \mu_d\Big(\cdot \cap \big\{(c_n,\infty]^d\big\}^{\complement}\Big).$$

We claim that $\mu^{(n)}$ is exchangeable and that $\mu_d = \lim_{n\to\infty}\mu^{(n)}$. Note, that $\mu^{(n)}$ is an increasing sequence of measures with $\mu^{(n)}(E_\infty^d) = \mu_d\big(\{(c_n,\infty]^d\}^{\complement}\big) < \infty$. An application of (Doob 2012, Chapter 10, Theorem 1a) yields that $\tilde{\mu} := \lim_{n\to\infty}\mu^{(n)}$ is a measure. The construction of $\mu^{(n)}$ and the continuity from below of any measure show that

$$\tilde{\mu}(A) = \lim_{n\to\infty}\mu_d\Big(A \cap \big\{(c_n,\infty]^d\big\}^{\complement}\Big) = \mu_d\Bigg(\bigcup_{n\in\mathbb{N}}\Big\{A \cap \big\{(c_n,\infty]^d\big\}^{\complement}\Big\}\Bigg) = \mu_d(A)$$

for every $A \in \mathcal{B}(E_\infty^d)$, since $\infty \notin E_\infty^d$. Therefore $\mu_d = \tilde{\mu} = \lim_{n \to \infty} \mu^{(n)}$ is a point-wise limit of measures. Next, we show that each $\mu^{(n)}$ is exchangeable. Consider the collection of $\mu^{(n)}$-exchangeable sets

$$\mathcal{A}_{\mu^{(n)}} := \left\{ A \in \mathcal{B}(E_\infty^d) \mid \mu^{(n)}(\pi(A)) = \mu^{(n)}(A) \; \forall \text{ permutations } \pi \text{ on } \{1, \dots, d\} \right\}.$$

Similar to "$\Leftarrow$", we can show that $\mathcal{A}_{\mu^{(n)}}$ contains E_∞^d, $\emptyset$ and countable unions of pairwise disjoint sets from $\mathcal{A}_{\mu^{(n)}}$. Moreover, for any $A \in \mathcal{A}_{\mu^{(n)}}$, we have

$$\mu^{(n)}(E_\infty^d \setminus A) = \mu^{(n)}(E_\infty^d) - \mu^{(n)}(A) = \mu^{(n)}(E_\infty^d) - \mu^{(n)}(\pi(A))$$
$$= \mu^{(n)}(E_\infty^d \setminus \pi(A)) = \mu^{(n)}(\pi(E_\infty^d \setminus A)),$$

since $\mu^{(n)}(A) < \infty$. Thus, $\mathcal{A}_{\mu^{(n)}}$ is a Dynkin system. Moreover, for every $\boldsymbol{a} \in \mathbb{R}^d$ with $\boldsymbol{a} \leq c_n$, we have

$$\mu^{(n)}\big((\boldsymbol{a}, \infty]^\complement\big) = \mu\big((\boldsymbol{a}, \infty]^\complement\big) = \mu\big((\pi(\boldsymbol{a}), \infty]^\complement\big)$$
$$= \mu^{(n)}\big((\pi(\boldsymbol{a}), \infty]^\complement\big) = \mu^{(n)}\big(\pi\big((\boldsymbol{a}, \infty]^\complement\big)\big)$$

by the exchangeability of $\boldsymbol{X}_d$. One can invoke a similar argument for all $\boldsymbol{a} \in \mathbb{R}^d$ which do not satisfy $\boldsymbol{a} \leq c_n$ to obtain $\mu^{(n)}\big((\boldsymbol{a}, \infty]^\complement\big) = \mu^{(n)}\big(\pi\big((\boldsymbol{a}, \infty]^\complement\big)\big)$. An inclusion-exclusion principle argument yields that $A_{\mu^{(n)}}$ contains a Π-stable generator of $\mathcal{B}(E_\infty^d)$. Therefore, $\mu^{(n)}$ is exchangeable. Combining the arguments above we have

$$\mu_d(A) = \lim_{n \to \infty} \mu^{(n)}(A) = \lim_{n \to \infty} \mu^{(n)}(\pi(A)) = \mu_d(\pi(A)),$$

for any $A \in \mathcal{B}(E_\infty^d)$, which shows that μ_d is exchangeable. $\qquad \square$

Proof of Corollary 2.10

Proof We calculate

$$\mathbb{P}(X_1 > t \mid X_2 > t, \dots, X_{d'} > t) = \frac{\mathbb{P}(X_1 > t, X_2 > t, \dots, X_{d'} > t)}{\mathbb{P}(X_2 > t, \dots, X_{d'} > t)}$$
$$= \frac{\exp\big(-\mu_{d'}\big((\boldsymbol{t}, \infty]^\complement\big)\big)}{\exp\big(-\mu_{d'-1}\big((\boldsymbol{t}, \infty]^\complement\big)\big)}$$
$$= \frac{\exp\big(-\mu_{d'}\big((\boldsymbol{t}, \infty]^\complement\big)\big)}{\exp\Big(-\mu_{d'}\Big([-\infty, \infty] \times \big((t, \infty]^{d'-1}\big)^\complement\Big)\Big)}$$
$$= \exp\big(-\mu_{d'}\big([-\infty, t] \times (t, \infty]^{d'-1}\big)\big).$$

Now, $\rho_{d'}^u$ is obtained by letting t tend to ∞. The relation $\rho_{d'+1}^u \geq \rho_{d'}^u$ is obvious by Proposition 2.9 and

$$\exp\left(-\mu_{d'}\left([-\infty, t] \times (t, \infty]^{d'-1}\right)\right)$$
$$= \exp\left(-\mu_d\left([-\infty, t] \times (t, \infty]^{d'-1} \times (-\infty, \infty]^{d-d'}\right)\right).$$

Proof of Lemma 2.13

Proof Fix $m \in \mathbb{N}$. We know that $H \sim \sum_{i=1}^m H^{(i,1/m)}$ for some i.i.d. processes $H^{(i,1/m)} \in \overline{\mathbb{R}}^{\mathbb{R}}$. W.l.o.g. assume that H is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\left(H^{(i,1/m)}\right)_{1 \leq i \leq m}$ is a random element in $\left(\left(\overline{\mathbb{R}}^{\mathbb{R}}\right)^m, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$, where $\tilde{\mathcal{F}}$ denotes the product σ-algebra generated by the finite dimensional projections. The idea of the proof is as follows:

Prove that the càdlàg property of H transfers to $H^{(i,1/m)}$ if we restrict the processes to rational time indices and define i.i.d. càdlàg processes $\tilde{H}^{(i,1/m)}$ as rational time limits of $H^{(i,1/m)}$.

W.l.o.g. we can choose the metric $\bar{\tau}(x, y) = |\arctan(x) - \arctan(y)|$ to define distances (and thus continuity) on $\overline{\mathbb{R}}$. Denote the (measurable) set of $\mathbb{Q}$-right-continuous paths of $H^{(i,1/m)}$ as

$$A_{rc,i} := \{w \mid w_i \text{ is right-continuous as a function on } \mathbb{Q}\}.$$

We have

$$\tilde{\mathbb{P}}\left(A_{rc,i}\right) = \tilde{\mathbb{P}}\left(\bigcap_{\substack{q \in \mathbb{Q}}} \bigcap_{\substack{\epsilon > 0 \\ \epsilon \in \mathbb{Q}}} \bigcup_{\substack{\delta > 0 \\ \delta \in \mathbb{Q}}} \bigcap_{\substack{q_1 \in (q, q+\delta) \\ q_1 \in \mathbb{Q}}} \left\{\omega \mid \bar{\tau}\left(\omega_i(q_1), \omega_i(q)\right) < \epsilon\right\}\right) \overset{(\star)}{=} 1.$$

It remains to prove $(\star)$. Therefore, assume that $(\star)$ does not hold. In this case there exists $\bar{q} \in \mathbb{Q}$ such that $H^{(i,1/m)}$ is not right-continuous at $\bar{q}$ with positive probability, i.e. there exists $\bar{q} \in \mathbb{Q}$ and $\epsilon > 0$ such that

$$\tilde{\mathbb{P}}\left(\bigcap_{\substack{\delta > 0 \\ \delta \in \mathbb{Q}}} \bigcup_{\substack{q_1 \in (\bar{q}, \bar{q}+\delta) \\ q_1 \in \mathbb{Q}}} \left\{\omega \mid \bar{\tau}\left(\omega_i(q_1), \omega_i(\bar{q})\right) > \epsilon\right\}\right) > 0,$$

which is equivalent to

$$
\tilde{\mathbb{P}}\left(\bigcup_{\substack{(q_n)_{n\in\mathbb{N}}\,\in\,\mathbb{Q}^{\mathbb{N}} \\ q_n > \bar{q} \\ \lim_{n\to\infty} q_n = \bar{q}}} \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{ \omega \,\middle|\, \bar{\tau}\big(\omega_i(q_n), \omega_i(\bar{q})\big) > \epsilon \right\}\right) > 0.
$$

Every fixed sequence $(q_n)_{n\in\mathbb{N}}$ such that for all $N \in \mathbb{N}$ there exists some $n \geq N$ with $\bar{\tau}(\omega_i(q_n), \omega_i(\bar{q})) > \epsilon$ satisfies one of the following 4 cases:

1. $|\omega_i(\bar{q})| < \infty$ and there exists some $\bar{\epsilon} > 0$ such that $\omega_i(q_n) - \omega_i(\bar{q}) > \bar{\epsilon}$ for infinitely many n or
2. $|\omega_i(\bar{q})| < \infty$ and there exists some $\bar{\epsilon} > 0$ such that $\omega_i(q_n) - \omega_i(\bar{q}) < -\bar{\epsilon}$ for infinitely many n or
3. $\omega_i(\bar{q}) = \infty$ and $\omega_i(q_n) < C$ for infinitely many n and some constant $C \in \mathbb{R}$ or
4. $\omega_i(\bar{q}) = -\infty$ and $\omega_i(q_n) > C$ for infinitely many n and some constant $C \in \mathbb{R}$.

Thus, there exists an $\bar{\epsilon} > 0$ or $C > 0$ such that either

1.
$$
\tilde{\mathbb{P}}\left(\bigcup_{\substack{(q_n)_{n\in\mathbb{N}}\,\in\,\mathbb{Q}^{\mathbb{N}} \\ q_n > \bar{q} \\ \lim_{n\to\infty} q_n = \bar{q}}} \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{ \omega \,\middle|\, \omega_i(q_n) - \omega_i(\bar{q}) > \bar{\epsilon},\ |\omega_i(\bar{q})| < \infty \right\}\right) > 0 \text{ or}
$$

2.
$$
\tilde{\mathbb{P}}\left(\bigcup_{\substack{(q_n)_{n\in\mathbb{N}}\,\in\,\mathbb{Q}^{\mathbb{N}} \\ q_n > \bar{q} \\ \lim_{n\to\infty} q_n = \bar{q}}} \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{ \omega \,\middle|\, \omega_i(q_n) - \omega_i(\bar{q}) < -\bar{\epsilon},\ |\omega_i(\bar{q})| < \infty \right\}\right) > 0 \text{ or}
$$

3.
$$\tilde{\mathbb{P}}\left(\bigcup_{\substack{(q_n)_{n\in\mathbb{N}}\,\in\,\mathbb{Q}^{\mathbb{N}} \\ q_n > \bar{q} \\ \lim_{n\to\infty} q_n = \bar{q}}} \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\omega \,\Big|\, \omega_i(q_n) < C,\ \omega_i(\bar{q}) = \infty\right\}\right) > 0 \text{ or}$$

4.
$$\tilde{\mathbb{P}}\left(\bigcup_{\substack{(q_n)_{n\in\mathbb{N}}\,\in\,\mathbb{Q}^{\mathbb{N}} \\ q_n > \bar{q} \\ \lim_{n\to\infty} q_n = \bar{q}}} \bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\omega \,\Big|\, \omega_i(q_n) > C,\ \omega_i(\bar{q}) = -\infty\right\}\right) > 0.$$

W.l.o.g. we assume that the first assertion holds, since the other cases are treated similarly. The assertion implies that there exists a (fixed) sequence $(\bar{q}_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \bar{q}_n = \bar{q}$ such that

$$\tilde{\mathbb{P}}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\omega \,\Big|\, \omega_i(\bar{q}_n) - \omega_i(\bar{q}) > \bar{\epsilon},\ |\omega_i(\bar{q})| < \infty\right\}\right) > 0.$$

Since the ω_i are i.i.d. we obtain that

$$0 = \mathbb{P}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{H_{\bar{q}_n} - H_{\bar{q}} > m\bar{\epsilon},\ |H_{\bar{q}}| < \infty\right\}\right)$$

$$\geq \tilde{\mathbb{P}}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} \left\{\omega \,\Big|\, \omega_1(\bar{q}_n) - \omega_1(\bar{q}) > \bar{\epsilon},\ |\omega_1(\bar{q})| < \infty\right\}\right)^m > 0,$$

which is a contradiction. Therefore, $(\star)$ is valid and $A_{rc} := \cap_{i=1}^m A_{rc,i}$ satisfies $\tilde{\mathbb{P}}(A_{rc}) = 1$.

Next, define a finally one-sided Cauchy sequence as a Cauchy sequence $(q_n)_{n\in\mathbb{N}}$ for which there exists some $N \in \mathbb{N}$ such that $q_n > \lim_{n\to\infty} q_n$ or $q_n < \lim_{n\to\infty} q_n$ for all $n \geq N$. Denote the set of existing left and right limits of w_i for finally one-sided $\mathbb{Q}$-Cauchy sequences as

$$A_{fC,i} := \left\{ \omega \mid \left(\omega_i(q_n) \right)_{n\in\mathbb{N}} \text{ is a Cauchy sequence (w.r.t. } \bar{\tau}) \right.$$

$$\left. \text{for all finally one-sided Cauchy sequences } (q_n)_{n\in\mathbb{N}} \in \mathbb{Q}^{\mathbb{N}} \right\}$$

$$= \bigcap_{\substack{(q_n)\in\mathbb{Q}^{\mathbb{N}} \\ \text{finally one-} \\ \text{sided Cauchy}}} \quad \bigcap_{\substack{\epsilon > 0 \\ \epsilon \in \mathbb{Q}}} \bigcup_{N\in\mathbb{N}} \bigcap_{m,n\geq N} \left\{ \omega \mid \bar{\tau}\left(\omega_i(q_n), \omega_i(q_m) \right) < \epsilon \right\}.$$

Note that we explicitly allow for rational Cauchy sequences with irrational limit. Similar to the proof of $\mathbb{P}\left(A_{rc,i} \right) = 1$ we can show that $\mathbb{P}\left(A_{fC,i} \right) = 1$, which implies that $A_{fC} := \cap_{i=1}^m A_{fC,i}$ satisfies $\tilde{\mathbb{P}}\left(A_{fC} \right) = 1$.

Finally, define

$$\tilde{H}_t^{(i,1/m)}(\omega) := \lim_{\substack{q \to t \\ q \in \mathbb{Q} \\ q > t}} \omega_i(q) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega), \tag{13}$$

which is measurable as the pointwise limit of measurable functions, if we can show that the limit exists and is independent of the chosen sequence. Therefore, choose two sequences $\left(q_n^{(1)} \right)_{n\in\mathbb{N}}, \left(q_n^{(2)} \right)_{n\in\mathbb{N}} \in (\mathbb{Q} \cap (t,\infty))^{\mathbb{N}}$ with limit $t \in \mathbb{R}$. $q_n^{(1)}, q_n^{(2)} > t$ for all $n \in \mathbb{N}$ implies that both sequences are finally one-sided Cauchy sequences. Therefore, both limits $\lim_{n\to\infty} \omega_i\left(q_n^{(1)} \right) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega)$ and $\lim_{n\to\infty} \omega_i\left(q_n^{(2)} \right) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega)$ exist. Moreover the combined sequence

$$\left(\tilde{q}_n \right)_{n\in\mathbb{N}} := \left(\left\{ \begin{array}{ll} q_n^{(1)} & n \text{ even} \\ q_n^{(2)} & n \text{ odd} \end{array} \right. \right)_{n\in\mathbb{N}}$$

is a finally one-sided Cauchy sequence. Therefore, the limit

$$\lim_{n\to\infty} \omega_i(\tilde{q}_n) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega)$$

exists as well and

$$\lim_{n\to\infty} \omega_i\left(q_n^{(1)} \right) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega) = \lim_{n\to\infty} \omega_i\left(q_n^{(2)} \right) \mathbf{1}_{\{A_{rc}\}}(\omega) \mathbf{1}_{\{A_{fC}\}}(\omega).$$

Thus, the limit in Eq. (13) exists and is independent of the chosen sequences, which implies that $\left(\tilde{H}^{(i,1/m)} \right)_{1\leq i\leq m}$ define valid stochastic processes. Observe that

$$\tilde{\mathbb{P}}\left(\tilde{H}_q^{(i,1/m)}(\omega) = w_i(q) = H_q^{(i,1/m)}(\omega) \text{ for all } q \in \mathbb{Q} \right) = 1. \tag{14}$$

Thus, $\tilde{H}^{(i,1/m)}$ and $H^{(i,1/m)}$ almost surely coincide on $\mathbb{Q}$, which follows from the fact that ω_i is only considered to be non-zero on $\mathbb{Q}$-right-continuous paths. We claim that $\tilde{H}^{(i,1/m)} \in D^\infty(\mathbb{R})$.

Firstly, we prove that $\tilde{H}^{(i,1/m)}(\omega)$ is right-continuous for all $\omega \in \Omega$ and $1 \leq i \leq m$. To this purpose choose some strictly decreasing sequence $(t_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ with $\lim_{n\to\infty} t_n = t$ and let $\epsilon > 0$ be arbitrary. Choose $t_n < q_n = q_n(\epsilon,\omega) \in \mathbb{Q}$ such that $\bar{\tau}\left(\tilde{H}_{t_n}^{(i,1/m)}, \tilde{H}_{q_n}^{(i,1/m)} \right) < \epsilon$ and $|t_n - q_n| < 1/n$, which is possible since $\tilde{H}_{t_n}^{(i,1/m)}$ is

defined as the limit of a rational time evaluations of $H^{(i,1/m)}$. Since $(q_n)_{n\in\mathbb{N}}$ is a finally one-sided Cauchy sequence with limit t there exists some $N(\omega,\epsilon)\in\mathbb{N}$ such that for all $n\geq N$ we have that $\bar{\tau}\left(\tilde{H}_{q_n}^{(i,1/m)},\tilde{H}_t^{(i,1/m)}\right)<\epsilon$. Therefore, for $n\geq N$, we obtain

$$\bar{\tau}\left(\tilde{H}_t^{(i,1/m)},\tilde{H}_{t_n}^{(i,1/m)}\right)\leq\bar{\tau}\left(\tilde{H}_t^{(i,1/m)},\tilde{H}_{q_n}^{(i,1/m)}\right)+\bar{\tau}\left(\tilde{H}_{q_n}^{(i,1/m)},\tilde{H}_{t_n}^{(i,1/m)}\right)<2\epsilon,$$

which yields that $\lim_{n\to\infty}\tilde{H}_{t_n}^{(i,1/m)}=\tilde{H}_t^{(i,1/m)}$. Thus, $\tilde{H}^{(i,1/m)}$ is right-continuous.

Secondly, we show that $\tilde{H}^{(i,1/m)}(\omega)$ has left limits for all $\omega\in\Omega$ and $1\leq i\leq m$. To see this choose some arbitrary $t\in\mathbb{R}$, $\epsilon>0$ and a sequence $(t_n)_{n\in\mathbb{N}}\in(-\infty,t)^{\mathbb{N}}$ with limit t. Define $q_n(\epsilon,\omega)$ as some rational number in $[t_n,t)$ such that $\bar{\tau}\left(\tilde{H}_{t_n}^{(i,1/m)},\tilde{H}_{q_n}^{(i,1/m)}\right)<\epsilon$, which is possible since $\tilde{H}^{(i,1/m)}$ is right-continuous. If n, m are large enough such that t_n and t_m are close to t we have that q_n and q_m are also close to t. Therefore, $(q_n)_{n\in\mathbb{N}}$ is a finally one-sided Cauchy sequence. Thus, we can find an $N(\omega,\epsilon)\in\mathbb{N}$ such that for all $m,n\geq N$ we have $\bar{\tau}\left(\tilde{H}_{q_n}^{(i,1/m)},\tilde{H}_{q_m}^{(i,1/m)}\right)<\epsilon$, which implies

$$\bar{\tau}\left(\tilde{H}_{t_n}^{(i,1/m)},\tilde{H}_{t_m}^{(i,1/m)}\right)\leq\bar{\tau}\left(\tilde{H}_{t_n}^{(i,1/m)},\tilde{H}_{q_n}^{(i,1/m)}\right)+\bar{\tau}\left(\tilde{H}_{t_m}^{(i,1/m)},\tilde{H}_{q_m}^{(i,1/m)}\right)$$
$$+\bar{\tau}\left(\tilde{H}_{q_n}^{(i,1/m)},\tilde{H}_{q_m}^{(i,1/m)}\right)$$
$$<3\epsilon.$$

Since $\epsilon>0$ was arbitrary we have shown that $\tilde{H}^{(i,1/m)}$ has left limits for every ω.

Obviously, $\left(\tilde{H}^{(i,1/m)}\right)_{1\leq i\leq m}$ are i.i.d. as almost sure pointwise limits of i.i.d. objects. It remains to prove that $H^{(1,1/m)}\sim\tilde{H}^{(1,1/m)}$. The characteristic functional of $H^{(1,1/m)}$, denoted as $CF_{H^{(1,1/m)}}(z,t)$ and the characteristic functional of $\tilde{H}^{(1,1/m)}$, denoted as $CF_{\tilde{H}^{(1,1/m)}}(z,t)$, coincide for $z\in\mathbb{R}^d$ and $t\in\mathbb{Q}^d$ by Eq. (14). For arbitrary $z\in\mathbb{R}^d$ and $t\in\mathbb{R}^d$ let $CF_H(z,t)$ denote the characteristic functional of H. We use the fact that H is right-continuous to obtain

$$CF_{H^{(1,1/m)}}(z,t)=CF_H(z,t)^{\frac{1}{m}}=\lim_{\substack{s\searrow t\\ s\in\mathbb{Q}^d\\ s>t}}CF_H(z,s)^{\frac{1}{m}}=\lim_{\substack{s\searrow t\\ s\in\mathbb{Q}^d\\ s>t}}CF_{H^{(1,1/m)}}(z,s)$$

$$=\lim_{\substack{s\searrow t\\ s\in\mathbb{Q}^d\\ s>t}}\mathbb{E}_{\tilde{\mathbb{P}}}\left[\exp\left(\sum_{j=1}^d iz_jH_{s_j}^{(1,1/m)}\right)\right]=\lim_{\substack{s\searrow t\\ s\in\mathbb{Q}^d\\ s>t}}\mathbb{E}_{\tilde{\mathbb{P}}}\left[\exp\left(\sum_{j=1}^d iz_j\tilde{H}_{s_j}^{(1,1/m)}\right)\right]$$

$$=CF_{\tilde{H}^{(1,1/m)}}(z,t),$$

where the second to last equality uses that $\tilde{H}^{(1,1/m)}$ and $H^{(1,1/m)}$ almost surely coincide on $\mathbb{Q}$ and the last equality uses the fact that $\tilde{H}^{(1,1/m)}$ is right-continuous. This proves that $H^{(1,1/m)}\sim\tilde{H}^{(1,1/m)}$. $\qquad\square$

Proof of Corollary 2.14

Proof We use the same notation as in the proof of Lemma 2.13. Denote the set of $\mathbb{Q}$-non-negative paths as

$$A_{\geq 0} := \left\{ \omega \mid \omega_i(q) \geq 0 \text{ for all } q \in \mathbb{Q} \text{ and } 1 \leq i \leq m \right\}.$$

Furthermore, denote the set of $\mathbb{Q}$-non-decreasing paths of as

$$A_{\nearrow} := \{ \omega \mid \omega_i(\cdot) \text{ is non-decreasing on } \mathbb{Q} \text{ for all } 1 \leq i \leq m \}.$$

By similar arguments as in the proof of Lemma 2.13 we obtain that

$$\tilde{\mathbb{P}}(A_{\nearrow}) = \tilde{\mathbb{P}}\left(\bigcap_{1 \leq i \leq m} \bigcap_{\substack{q_1, q_2 \in \mathbb{Q} \\ q_1 \leq q_2}} \{ \omega \mid \omega_i(q_1) \leq \omega_i(q_2) \} \right) = 1 \text{ and}$$

$$\tilde{\mathbb{P}}(A_{\geq 0}) = \tilde{\mathbb{P}}\left(\bigcap_{1 \leq i \leq m} \bigcap_{q \in \mathbb{Q}} \{ \omega \mid \omega_i(q) \geq 0 \} \right) = 1,$$

since intersections of countably many sets with probability 1 have probability 1. Next, define

$$\hat{H}_t^{(i,1/m)}(\omega) := \tilde{H}_t^{(i,1/m)}(\omega) \mathbf{1}_{\{A_{\nearrow}\}}(\omega) \mathbf{1}_{\{A_{\geq 0}\}}(\omega). \tag{15}$$

Obviously, $\hat{H}^{(i,1/m)}$ is non-negative and non-decreasing. Similar to the proof of Lemma 2.13 we can show that the Laplace transforms of $\hat{H}^{(i,1/m)}$ and $H^{(i,1/m)}$ coincide and the claim follows. $\qquad\square$

Proof of Proposition 2.16

Proof Let $\tilde{b}$ and $\tilde{\nu}$ denote the drift and Lévy measure of H given by (Rosiński 2018, Theorem 2.8). Note that $\tilde{\nu}$ is a measure on $\overline{\mathbb{R}}^{\mathbb{R}}$ equipped with the σ-algebra generated by the finite dimensional projections, since Rosiński (2018) views id-processes as processes in $\mathbb{R}^{\mathbb{R}}$. There are two things that need to be shown:

1. $\tilde{\nu}$ can be restricted to a measure on M_∞^0.
2. The integral over $\tilde{\nu}$ in Eq. (3) can be defined without the compensating term $\sum_{i=1}^d z_i x(t_i) \mathbf{1}_{\{|x(t_i)| < \epsilon\}}$ and $b \in M_\infty^0 \cap D(\mathbb{R})$.

We start with the first statement. Similar to the proof of (Rosiński 2018, Theorem 3.4) we can show that there exists an exact representation ν of $\tilde{\nu}$ defined on

$$C_0 := \left\{ x \in D^\infty(\mathbb{R}) \mid \lim_{t \to -\infty} x(t) = 0 \right\},$$

since C_0 is an algebraic group under addition and a standard Borel space as a measurable subset of a standard Borel space. For additional information on exact representations of Lévy measures, see (Rosiński 2018, Definition 2.20). For our purposes it suffices to view an exact representation of a Lévy measure as a restriction of a Lévy measure to a smaller domain. To prove the first statement it suffices to show that ν vanishes on the set

$$C_1 := \{x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}\}^{\complement},$$

since C_1 is measurable (in $D^\infty(\mathbb{R})$). In particular, $M_\infty^0 = C_0 \cap C_1^{\complement}$.

For $I = (t_1, \dots, t_d) \in \mathbb{R}^d$ (w.l.o.g. $t_1 \leq \dots \leq t_d$) and $A \in \mathcal{B}\left(\overline{\mathbb{R}}^d\right)$ define

$$\nu_I(A) := \nu\left(\{x \in D^\infty(\mathbb{R}) \mid (x(t_1), \dots, x(t_d)) \in A\}\right).$$

$\tilde{\nu}_I(A)$ is defined analogously. Observe that $\nu_I(A) = \tilde{\nu}_I(A)$ for all $A \in \mathcal{B}\left(\overline{\mathbb{R}}^d\right)$ by construction. We show that

$$\nu(C_1) = \nu\left(\{x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}\}^{\complement}\right) = 0.$$

(Barndorff-Nielsen et al. 2006, Proposition 6.1) tells us that the Lévy measure $\tilde{\nu}_{(t_1,\dots,t_d)}$ of $(H_{t_1}, \dots, H_{t_d})$ is concentrated on the cone $K_d := \{\boldsymbol{x} \in \overline{\mathbb{R}}^d \mid 0 \leq x_1 \leq \dots \leq x_d\}$, which implies that $\nu_{(t_1,\dots,t_d)}$ is also concentrated on K_d. Now, assume that $\nu(C_1) = \nu\left(\{x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}\}^{\complement}\right) > 0$. Observe that

$$\{x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}\}^{\complement} = \bigcup_{\substack{t_1, t_2 \in \mathbb{Q} \\ t_1 < t_2}} \{x \in D^\infty(\mathbb{R}) \mid x(t_1) > x(t_2)\}.$$

Therefore, there exist $\bar{t}_1 < \bar{t}_2$ with $\nu\left(\{x \in D^\infty(\mathbb{R}) \mid x(\bar{t}_1) > x(\bar{t}_2)\}\right) > 0$. By the construction of ν we get

$$\nu\left(\{x \in D^\infty(\mathbb{R}) \mid x(\bar{t}_1) > x(\bar{t}_2)\}\right) = \nu_{(\bar{t}_1, \bar{t}_2)}\left(\cup_{s \in \mathbb{Q}} (s, \infty] \times [-\infty, s]\right)$$
$$= \tilde{\nu}_{(\bar{t}_1, \bar{t}_2)}\left(\cup_{s \in \mathbb{Q}} (s, \infty] \times [-\infty, s]\right) = 0,$$

which is a contradiction. Therefore, $\nu(C_1) = \nu\left(\{x \in D^\infty(\mathbb{R}) \mid x \text{ is non-decreasing}\}^{\complement}\right) = 0$ and ν is concentrated on non-decreasing functions which satisfy $\lim_{t \to -\infty} x(t) = 0$. Thus, ν is concentrated on $M_\infty^0 = C_0 \cap C_1^{\complement}$, i.e.

$$\nu(A) = \nu(A \cap M_\infty^0), \quad \forall A \in \mathcal{B}(D^\infty(\mathbb{R})).$$

Let us turn to the proof of the second statement. For every $d \in \mathbb{N}$ and $\boldsymbol{t} \in \mathbb{R}^d$ we obtain a non-negative drift vector $(b(t_1), \ldots, b(t_d))$ from the d-dimensional Lévy–Khintchine triplet of the non-negative random vector $(H_{t_1}, \ldots, H_{t_d})$ with truncation function 0. Condition $(\Diamond'2)$ and (Barndorff-Nielsen et al. 2006, Proposition 6.1) imply the existence of a unique non-decreasing finite drift $b : \mathbb{R} \to [0, \infty)$. Right-continuity and $\lim_{t \to -\infty} b(t) = 0$ follow from the right-continuity of H and $\lim_{t \to -\infty} H_t = 0$. Thus, $b \in M_\infty^0 \cap D(\mathbb{R})$.

Combining the above yields

$$\mathbb{E}\left[\exp\left(-\sum_{i=1}^d z_i H_{t_i}\right)\right] = \exp\left(-\sum_{i=1}^d z_i b(t_i) + \int_{M_\infty^0}\left(\exp\left(-\sum_{i=1}^d z_i x(t_i)\right) - 1\right)\nu(\mathrm{d}x)\right)$$

for every $\boldsymbol{z} \in [0, \infty)^d, \boldsymbol{t} \in \mathbb{R}^d$.

Remark 6 (Implications for strong-idt processes) An id-process H is called strong-idt, if

$$\left(H_t\right)_{t \geq 0} \sim \left(\sum_{i=1}^n H_{\frac{t}{n}}^{(i)}\right)_{t \geq 0}$$

for all $n \in \mathbb{N}$, where $\left(H^{(i)}\right)_{i \in \mathbb{N}}$ denote i.i.d. copies of H. Such processes are studied, among others, in Kopp and Molchanov (2018), Mai (2020). Kopp and Molchanov (2018) study the Lévy measure and series representations of real-valued strong-idt processes without focus on non-decreasing paths. Mai (2020) refines these results in the special case of non-decreasing H, which might possibly also take the value ∞. However, Mai (2020) does not formally prove the extension to extended real-valued processes, despite he uses the results of Kopp and Molchanov (2018). Proposition 2.16 fills this gap by formally justifying that the claimed extension is correct. Furthermore, whereas Kopp and Molchanov (2018) work on the space of càdlàg functions equipped with the Skorohod ($J1$) metric, Mai (2020) works with the Lévy metric defined for distribution functions. While it is known that the two metrics are not equivalent in general, one can actually prove that their induced Borel σ-algebras on the space of non-decreasing paths coincide. Thus, implicitly both references indeed work with the same objects. In particular, (Mai 2020, Lemma 1) implicitly shows with a tedious and probabilistic proof that the Lévy measure of a non-decreasing strong-idt process is concentrated on non-decreasing paths. In this regard, Proposition 2.16 provides a more direct proof of this fact. Moreover, it is even more general, since it holds for arbitrary non-decreasing and non-negative id-processes and not just strong-idt processes.

Proof of Theorem 3.1

Proof

"⟹" Let $n \in \mathbb{N}$. Since X is min-id, there exist i.i.d. sequences $(X^{(i,1/n)})_{1 \le i \le n}$ $\in (-\infty, \infty]^{\mathbb{N}}$ such that $X \sim \min_{1 \le i \le n} X^{(i,1/n)}$. First, we claim that the exchangeability of X implies the exchangeability of $X^{(1,1/n)}$. Seeking a contradiction, we assume that $X^{(1,1/n)}$ is not exchangeable. In this case, there exists $\{i_1, \dots, i_d\} \subset \mathbb{N}$ such that $(X^{(1,1/n)}_{1,i_1}, \dots, X^{(1,1/n)}_{1,i_d})$ is not exchangeable. By similar arguments as in the proof of Proposition 2.8 there exist $\boldsymbol{x} \in \mathbb{R}^d$ and a permutation π on $\{1, \dots, d\}$ such that $\mathbb{P}(X^{(1,1/n)}_d > \boldsymbol{x}) \ne \mathbb{P}(X^{(1,1/n)}_d > \pi(\boldsymbol{x}))$. This yields

$$\mathbb{P}(X_{i_1} > x_1, \dots, X_{i_d} > x_d) = \mathbb{P}\left(X^{(1,1/n)}_{i_1} > x_1, \dots, X^{(1,1/n)}_{i_d} > x_d\right)^n$$

$$\ne \mathbb{P}\left(X^{(1,1/n)}_{i_1} > \pi(\boldsymbol{x})_1, \dots, X^{(1,1/n)}_{i_d} > \pi(\boldsymbol{x})_d\right)^n$$

$$= \mathbb{P}\left(X_{i_1} > \pi(\boldsymbol{x})_1, \dots, X_{i_d} > \pi(\boldsymbol{x})_d\right),$$

which is a contradiction. Therefore, $X^{(1,1/n)}$ is exchangeable. Now, de Finetti's Theorem yields the existence of i.i.d. nnnd càdlàg processes $\left(H^{(i,1/n)}\right)_{1 \le i \le n}$ $\in D^\infty(\mathbb{R})$ such that

$$\left(X^{(i,1/n)}_j\right)_{j \in \mathbb{N}} \sim \inf\left\{t \in \mathbb{R} \mid H^{(i,1/n)}_t \ge E^{(i)}_j\right\},$$

where $((E^{(i)}_j)_{j \in \mathbb{N}})_{1 \le i \le n}$ are i.i.d. unit exponential independent of $\left(H^{(i,1/n)}\right)_{1 \le i \le n}$. Obviously, $\lim_{t \to -\infty} H^{(i,1/n)}_t = 0$ almost surely, since $\mathbb{P}\left(X^{(1/n)}_{i,j} = -\infty\right) > 0$ otherwise. Moreover, $\mathbb{P}\left(H^{(i,1/n)}_t = \infty\right) < 1$ for all $t \in \mathbb{R}$, since $\ell = \infty$. Therefore, $H^{(i,1/n)}$ satisfies Condition $(\lozenge')$. It remains to show that $H^{(1,1)} =: H$ is infinitely divisible and unique. Let $\overline{F}_k$ denote the survival function of $(X_1, \dots, X_k)$ and recall that $\left(X^{(1,1/n)}_1, \dots, X^{(1,1/n)}_k\right) \sim \overline{F}^{1/n}_k$. Choose $\boldsymbol{z} \in \mathbb{N}^d$ and $\boldsymbol{t} \in \mathbb{R}^d$, then

$$L(\boldsymbol{z}, \boldsymbol{t}) = \mathbb{E}\left[\exp\left(-\sum_{j=1}^d z_j H_{t_j}\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{j=1}^d \sum_{k=1}^{z_j} H_{t_j}\right)\right]$$

$$= \overline{F}_{\sum_{j=1}^d z_j}(\underbrace{t_1, \dots, t_1}_{z_1 \text{ times}}, \dots, \underbrace{t_d, \dots, t_d}_{z_d \text{ times}})$$

$$= \left(\overline{F}^{1/n}_{\sum_{j=1}^d z_j}\right)^n (\underbrace{t_1, \dots, t_1}_{z_1 \text{ times}}, \dots, \underbrace{t_d, \dots, t_d}_{z_d \text{ times}})$$

$$= \mathbb{E}\left[\exp\left(-\sum_{j=1}^d \sum_{k=1}^{z_j} H^{(1,1/n)}_{t_j}\right)\right]^n = \mathbb{E}\left[\prod_{i=1}^n \exp\left(-\sum_{j=1}^d z_j H^{(i,1/n)}_{t_j}\right)\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{j=1}^d z_j \sum_{i=1}^n H^{(i,1/n)}_{t_j}\right)\right].$$

Using the fact that the Laplace transform of a non-negative random vector is uniquely determined by its values on $\mathbb{N}^d$, see Kleiber and Stoyanov (2013) for more details, this shows that $H_t \sim \sum_{i=1}^n H_t^{(i,1/n)}$. Since t and n were arbitrary, we get that H is infinitely divisible. The uniqueness of H follows from

$$\mathbb{E}\left[\exp\left(-\sum_{j=1}^d z_j H_{t_j}\right)\right] = \overline{F}_{\sum_{j=1}^d z_j}(\underbrace{t_1, \ldots, t_1}_{z_1 \text{ times}}, \ldots, \underbrace{t_d, \ldots, t_d}_{z_d \text{ times}}).$$

"$\Leftarrow$" We refer to the survival function of $(X_1, \ldots, X_d)$ by $\overline{F}_d$. Since exchangeability of X is obvious by construction, it suffices to show that $\overline{F}_d^{1/n}$ is a survival function of a random variable on $(-\infty, \infty]^d$ for every $d, n \in \mathbb{N}$ and that $\ell = \infty$. By Corollary 2.14, there exist i.i.d. extended chronometers $(H^{(i,1/n)})_{1 \le i \le n}$ such that $H \sim \sum_{i=1}^n H^{(i,1/n)}$. It easily follows that $\lim_{t \to -\infty} H_t^{(1,1/n)} = 0$, which implies that $H^{(1,1/n)} \in M_\infty^0$. For $t \in \mathbb{R}^d$, we get

$$\overline{F}_d(t) = \mathbb{E}\left[\exp\left(-\sum_{j=1}^d H_{t_j}\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{j=1}^d \sum_{i=1}^n H_{t_j}^{(i,1/n)}\right)\right]$$

$$= \mathbb{E}\left[\exp\left(-\sum_{j=1}^d H_{t_j}^{(1,1/n)}\right)\right]^n.$$

Since $H_{t_j}^{(1,1/n)}$ is càdlàg we obtain that

$$\overline{F}_d^{1/n}(t) = \mathbb{E}\left[\exp\left(-\sum_{j=1}^d H_{t_j}^{(1,1/n)}\right)\right]$$

is the survival function of the first d components of the exchangeable sequence

$$X^{(1,1/n)} := \left(\inf\left\{t \in \mathbb{R} \mid H_t^{(1,1/n)} \ge E_i\right\}\right)_{i \in \mathbb{N}} \in (-\infty, \infty]^{\mathbb{N}}.$$

Therefore, X is min-id. Moreover, $\ell = \infty$, since H satisfies Condition $(\lozenge'2)$. Thus, X satisfies Condition $(\lozenge)$. Since d and n were arbitrary, the claim follows.

Proof of Theorem 3.5

Proof Theorem 3.1 and Proposition 2.9 provide a one-to-one correspondence between the min-id sequence X, a (unique) Lévy measure ν on M_∞^0 with drift b, and an exponent measure μ on $E_\infty^{\mathbb{N}}$. Choosing $t \in \mathbb{R}^d$, $d \in \mathbb{N} \cup \{\infty\}$, we can rewrite this correspondence as

$$\mathbb{P}\left(X_1 > t_1, \ldots, X_d > t_d\right) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{d} H_{t_i}\right)\right]$$

$$= \exp\left(-\sum_{i=1}^{d} b(t_i) - \int_{M_\infty^0} 1 - \exp\left(-\sum_{i=1}^{d} x(t_i)\right)\nu(\mathrm{d}x)\right)$$

$$= \exp\left(-\sum_{i=1}^{d} b(t_i) - \int_{M_\infty^0} 1 - \prod_{i=1}^{d} \exp\left(-x(t_i)\right)\nu(\mathrm{d}x)\right)$$

$$= \exp\left(-\sum_{i=1}^{d} b(t_i) - \int_{\overline{M}_\infty^0} 1 - \prod_{i=1}^{d} \overline{G}(t_i)\gamma(\mathrm{d}G)\right)$$

$$= \exp\left(-\sum_{i=1}^{d} b(t_i) - \int_{\overline{M}_\infty^0} \otimes_{i=1}^{d}\mathbb{P}_G\left((-\infty, \infty]^d \setminus (\boldsymbol{t}, \infty]\right)\gamma(\mathrm{d}G)\right),$$

where $\overline{G} = 1 - G$ is a survival function of a random variable on $(-\infty, \infty]$ and γ is the image measure of the Lévy measure ν under the transformation $h : M_\infty^0 \to \overline{M}_\infty^0$, $x \mapsto 1 - \exp(-x(\cdot))$. This implies that for every $d \in \mathbb{N} \cup \{\infty\}$

$$\mu_d\left(E_\infty^d \setminus (\boldsymbol{t}, \infty]\right) = \mu_{b,d}\left((-\infty, \infty]^d \setminus (\boldsymbol{t}, \infty]\right)$$

$$+ \int_{\overline{M}_\infty^0} \otimes_{i=1}^{d}\mathbb{P}_G\left((-\infty, \infty]^d \setminus (\boldsymbol{t}, \infty]\right)\gamma(\mathrm{d}G).$$

A similar argument as in the proof of Proposition 2.8 yields

$$\mu_d(A) = \mu_{b,d}(A) + \int_{\overline{M}_\infty^0} \otimes_{i=1}^{d}\mathbb{P}_G(A)\gamma(\mathrm{d}G) \quad \text{for all } A \in \mathcal{B}(E_\infty^d).$$

It remains to verify the properties of γ. Obviously, $\gamma\left(0_{D^\infty(\mathbb{R})}\right) = 0$. Applying the inequalities $1 - x \leq \min\{1, -\log(x)\}$ and $\min\{1, -\log(x)\} \leq e(1 - x)$ for any $x \in [0, 1]$ yields

$$\int_{\overline{M}_\infty^0} G(t)\gamma(\mathrm{d}G) = \int_{\overline{M}_\infty^0} 1 - \overline{G}(t)\gamma(\mathrm{d}G) \leq \int_{M_\infty^0} \min\{1, x(t)\}\nu(\mathrm{d}x) < \infty$$

as well as

$$\int_{M_\infty^0} \min\{1, x(t)\}\nu(\mathrm{d}x) \leq e \int_{\overline{M}_\infty^0} 1 - \overline{G}(t)\gamma(\mathrm{d}G) = e \int_{\overline{M}_\infty^0} G(t)\gamma(\mathrm{d}G) < \infty.$$

Therefore, the integrability condition of γ is equivalent to the integrability condition of ν. $\qquad\square$

Proof of Proposition 4.1

Proof We observe that

$$H_t = \sum_{k \geq 1} -\log\left\{1 - G_k\left(\frac{t}{S_k}\right)\right\} = \int_{[0,\infty)\times\overline{M}_\infty^0} -\log\left(1 - G\left(\frac{t}{S}\right)\right) N(\mathrm{d}(S,G))$$

is given by integration of the (measurable) function $(S,G) \mapsto -\log(1 - G(t/S))$ w.r.t. the Poisson random measure N. Similarly, for $z \in [0,\infty)^d$, $\sum_{i=1}^d z_i H_{t_i}$ is given by the integration of $(S,G) \mapsto \sum_{i=1}^d -z_i \log\left(1 - G(t_i/S)\right)$ w.r.t. the Poisson random measure N. Therefore, the Laplace transform of H is given by the (one-dimensional) Laplace transform of an integral over a Poisson random measure. An application of (Resnick 2007, Proposition 3.6) yields

$$L(z,t) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^d z_i H_{t_i}\right)\right]$$

$$= \mathbb{E}\left[\exp\left(-\int_{[0,\infty]\times\overline{M}_\infty^0} \sum_{i=1}^d -z_i \log\left(1 - G\left(\frac{t_i}{S}\right)\right) N(\mathrm{d}(S,G))\right)\right]$$

$$= \exp\left(-\int_{[0,\infty]} \int_{\overline{M}_\infty^0} 1 - \prod_{i=1}^d \left(1 - G\left(\frac{t_i}{s}\right)\right)^{z_i} \rho(\mathrm{d}G)\kappa(\mathrm{d}s)\right).$$

Thus, H is infinitely divisible and the exchangeable min-id sequence X associated with H has exponent measure $\mu_{\kappa,\rho}$. We can express the survival function of X as

$$\mathbb{P}(X > t) = \exp\left(-\int_{\overline{M}_\infty^0} \int_0^\infty 1 - \prod_{i\in\mathbb{N}} \left(1 - G\left(\frac{t_i}{s}\right)\right)\kappa(\mathrm{d}s)\rho(\mathrm{d}G)\right)$$

$$= \exp\left(-\int_{\overline{M}_\infty^0} \int_0^\infty \int_{[0,\infty]^\mathbb{N}} \mathbf{1}\left\{y_i \leq \frac{t_i}{s} \text{ for some } i \in \mathbb{N}\right\} \left(\otimes_{i\in\mathbb{N}}\mathbb{P}_G\right)(\mathrm{d}y)\kappa(\mathrm{d}s)\rho(\mathrm{d}G)\right)$$

$$= \exp\left(-\int_{\overline{M}_\infty^0} \int_{[0,\infty]^\mathbb{N}} \int_0^\infty \mathbf{1}\left\{s \leq \max_{i\in\mathbb{N}} \frac{t_i}{y_i}\right\} \kappa(\mathrm{d}s) \left(\otimes_{i\in\mathbb{N}}\mathbb{P}_G\right)(\mathrm{d}y)\rho(\mathrm{d}G)\right),$$

which finishes the argument.

Proof of Proposition 4.2

Proof It is easy to see that $H^{(\kappa)}$ is infinitely divisible, since $x(\cdot) \mapsto \int_0^\cdot$ is a measurable map in $D^\infty\left([0,\infty)\right)$. By the generalized Lévy–Itô representation (Rosiński 2018, Proposition 3.1 and Theorem 5.1) there exists a version V' of V such that

$$\left(V'_s\right)_{s\geq 0} = \left(b_V(s) + \int_{D^\infty(\mathbb{R})} x(s)N(\mathrm{d}x)\right)_{s\geq 0},$$

where N denotes a Poisson random measure on $D^\infty\big([0,\infty)\big)$ with intensity ν_V. Note that the compensating term in the generalized Lévy–Itô representation can be omitted by (Rosiński 2018, Theorem 5.1). Moreover, N can be chosen as a random measure on $D^\infty\big([0,\infty)\big)_+ := \{x \in D^\infty\big([0,\infty)\big) \mid x(t) \geq 0 \text{ for all } t \geq 0\}$, which follows by similar arguments as in the proof of Proposition 2.16. Thus,

$$\left(H_t^{(\kappa)}\right)_{t\geq0} = \left(\int_0^t V_s\kappa(\mathrm{d}s)\right)_{t\geq0}$$
$$\sim \left(\int_0^t \left(b_V(s) + \int_{D^\infty\big([0,\infty)\big)_+} x(s)N(\mathrm{d}x)\right)\kappa(\mathrm{d}s)\right)_{t\geq0}.$$

Since N is σ-finite and concentrated on non-negative functions we can use Fubini's Theorem to obtain

$$\left(\int_0^t \left(b_V(s) + \int_{D^\infty\big([0,\infty)\big)_+} x(s)N(\mathrm{d}x)\right)\kappa(\mathrm{d}s)\right)_{t\geq0}$$
$$=\left(\int_0^t b_V(s)\kappa(\mathrm{d}s) + \int_{D^\infty\big([0,\infty)\big)_+}\int_0^t x(s)\kappa(\mathrm{d}s)N(\mathrm{d}x)\right)_{t\geq0},$$

which is a decomposition of $H^{(\kappa)}$ into a non-decreasing deterministic drift $b^{(\kappa)}(t) := \int_0^t b(s)\kappa(\mathrm{d}s)$ and an integral over a Poisson random measure $\int_{D^\infty\big([0,\infty)\big)_+} x^{(\kappa)}(t)N(\mathrm{d}x)$, where $x^{(\kappa)}(t) := \int_0^t x(s)\kappa(\mathrm{d}s)$ is a non-decreasing function in $D^\infty\big([0,\infty)\big)_+$. Therefore, using the usual formula for the Laplace transform of an integral over a Poisson random measure (Resnick 2007, Proposition 3.6), we obtain for arbitrary $z \in [0,\infty]^d$ and $t \in [0,\infty)^d$

$$\mathbb{E}\left[\exp\left(-\sum_{i=1}^d z_i H_{t_i}^{(\kappa)}\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{i=1}^d z_i\left(b_{t_i}^{(\kappa)} + \int_{D^\infty\big([0,\infty)\big)_+} x^{(\kappa)}(t_i)N(\mathrm{d}x)\right)\right)\right]$$
$$= \exp\left(-\sum_{i=1}^d z_i b_{t_i}^{(\kappa)} - \int_{D^\infty\big([0,\infty)\big)_+} 1 - \exp\left(-\sum_{i=1}^d z_i x^{(\kappa)}(t_i)\right)\nu_V(\mathrm{d}x)\right).$$

Note that

$$\exp\left(-\sum_{i=1}^d z_i b_{t_i}^{(\kappa)} - \int_{D^\infty\big([0,\infty)\big)_+} 1 - \exp\left(-\sum_{i=1}^d z_i x^{(\kappa)}(t_i)\right)\nu_V(\mathrm{d}x)\right)$$
$$\stackrel{(\star)}{=} \exp\left(-\sum_{i=1}^d z_i b_{t_i}^{(\kappa)} - \int_{M_\ell^0} 1 - \exp\left(-\sum_{i=1}^d z_i x(t_i)\right)\nu_\kappa(\mathrm{d}x)\right),$$

since we use that $x^{(\kappa)} \in M_\ell^0$ and we only omit those terms in $(\star)$ for which $\exp\big(- \sum_{i=1}^d z_i x^{(\kappa)}(t_i)\big) = 1$.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Aldous, D.J.: Exchangeability and related topics. In: École d'Été de Probabilités de Saint-Flour XIII – 1983, pp. 1–198. Springer Berlin Heidelberg (1985)

Alzaid, A.A., Proschan, F.: Max-infinite divisibility and multivariate total positivity. J. Appl. Probab. **31**(3), 721–730 (1994). https://doi.org/10.2307/3215150

Balkema, A.A., Resnick, S.I.: Max-infinite divisibility. J. Appl. Probab. **14**(2), 309–319 (1977)

Barczy, M., Kern, P., Pap, G.: Dilatively stable stochastic processes and aggregate similarity. Aequationes Math. **89**(6), 1485–1507 (2015)

Barndorff-Nielsen, O.E., Maejima, M., Sato, K.I.: Infinite divisibility for stochastic processes and time change. J. Theor. Probab. **19**, 411–446 (2006)

Barndorff-Nielsen, O.E., Shepard, N.: Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. J. R. Stat. Soc. Ser. B Stat Methodol. **63**, 167–207 (2001)

Bertoin, J.: Subordinators: examples and applications. In: Lectures on probability theory and statistics, pp. 1–91. Springer (1999)

Bopp, G.P., Shaby, B.A., Huser, R.: A hierarchical max-infinitely divisible spatial model for extreme precipitation. J. Am. Stat. Assoc. (2020). https://doi.org/10.1080/01621459.2020.1750414

Brück, F.: Exact simulation of continuous max-id processes with applications to exchangeable max-id sequences. J. Multivar. Anal. (2022). https://doi.org/10.1016/j.jmva.2022.105117

Carr, P., Geman, H., Madan, D.B., Yor, M.: Stochastic volatility for Lévy processes. Math. Financ. **13**(3), 345–382 (2003)

Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**(2), 385–407 (1985). http://www.jstor.org/stable/1911242

Davydov, Y., Molchanov, I., Zuyev, S.: Strictly stable distributions on convex cones. Electron. J. Probab. **13**, 259–321 (2008)

de Finetti, B.: Funzione caratteristica di un fenomeno aleatorio. Atti della R. Academia Nazionale dei Lincei, Ser. 6. Memorie, Classe di Scienze Fisiche, Matematiche e Naturali **4**, 251–299 (1931)

de Haan, L.: A spectral representation for max-stable processes. Ann. Probab. **12**(4), 1194–1204 (1984). https://doi.org/10.1214/aop/1176993148

Doksum, K.: Tailfree and neutral random probabilities and their posterior distributions. Ann. Probab. **2**, 183–201 (1974). https://doi.org/10.1214/aop/1176996703

Doob, J.L.: Measure theory, vol. 143. Springer Science & Business Media (2012)

Duan, J.C., Yeh, C.Y.: Jump and volatility risk premiums implied by vix. J. Econ. Dyn. Control **34**(11), 2232–2244 (2010). https://doi.org/10.1016/j.jedc.2010.05.006

Epifani, I., Lijoi, A., Prünster, I.: Exponential functionals and means of neutral-to-the-right priors. Biometrika **90**(4), 791–808 (2003)

Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**, 209–230 (1973). https://doi.org/10.1214/aos/1176342360

Ferguson, T.S.: Prior distributions on spaces of probability measures. Ann. Stat. **2**, 615–629 (1974). https://doi.org/10.1214/aos/1176342752

Genest, C., Nešlehová, J.G., Rivest, L.P.: The class of multivariate max-id copulas with ℓ_1-norm symmetric exponent measure. Bernoulli **24**(4B), 3751–3790 (2018). https://doi.org/10.3150/17-BEJ977

Genest, C., Rivest, L.P.: A characterization of Gumbel's family of extreme value distributions. Statist. Probab. Lett. **8**(3), 207–211 (1989). https://doi.org/10.1016/0167-7152(89)90123-5

Hakassou, A., Ouknine, Y.: A note on α-idt processes. arXiv (2012)

Hakassou, A., Ouknine, Y.: Idt processes and associated Lévy processes with explicit constructions. Stochastics An International Journal of Probability and Stochastic Processes **85**(6), 1073–1111 (2013)

Hjort, N.L.: Nonparametric Bayes estimators based on beta processes in models for life history data. Ann. Stat. **18**(3), 1259–1294 (1990)

Huser, R., Opitz, T., Thibaud, E.: Penultimate modeling of spatial extremes: statistical inference for max-infinitely divisible processes. arXiv (2018)

Huser, R., Opitz, T., Thibaud, E.: Max-infinitely divisible models and inference for spatial extremes. Scand. J. Stat. **48**(1), 321–348 (2021)

Hüsler, J.: Limit properties for multivariate extreme values in sequences of independent, non-identically distributed random vectors. Stoch. Process. Their Appl. **31**(1), 105–116 (1989)

Iglói, E., Barczy, M.: Path properties of dilatively stable processes and singularity of their distributions. Stoch. Anal. Appl. **30**(5), 831–848 (2012). https://doi.org/10.1080/07362994.2012.704849

James, L.F.: Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. Ann. Stat. **33**(4), 1771–1799 (2005). https://doi.org/10.1214/009053605000000336

James, L.F., Lijoi, A., Prünster, I.: Posterior analysis for normalized random measures with independent increments. Scand. J. Stat. **36**(1), 76–97 (2009). https://doi.org/10.1111/j.1467-9469.2008.00609.x

Joe, H., Hu, T.: Multivariate distributions from mixtures of max-infinitely divisible distributions. J. Multivar. Anal. **57**(2), 240–265 (1996)

Kabluchko, Z., Stoev, S.: Stochastic integral representations and classification of sum- and max-infinitely divisible processes. Bernoulli **22**(1), 107–142 (2016). https://doi.org/10.3150/14-BEJ624

Kalbfleisch, J.D.: Non-parametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society. Series B (Methodological) **40**(2), 214–221 (1978). URL http://www.jstor.org/stable/2984758

Kleiber, C., Stoyanov, J.: Multivariate distributions and the moment problem. J. Multivar. Anal. **113**, 7–18 (2013)

Kömm, H.: Integrated Volatility, pp. 29–46. Springer Fachmedien Wiesbaden, Wiesbaden (2016). https://doi.org/10.1007/978-3-658-12596-7_3

Kopp, C., Molchanov, I.: Series representation of time-stable stochastic processes. Probab. Math. Stat. **38**, 299–315 (2018). https://doi.org/10.19195/0208-4147.38.2.4

Lee, P.M.: Infinitely divisible stochastic processes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **7**(2), 147–160 (1967)

Mai, J.F.: Extendibility of Marshall–Olkin distributions via Lévy subordinators and an application to portfolio credit risk. Dissertation, Technische Universität München (2010)

Mai, J.F.: Canonical spectral representation for exchangeable max-stable sequences. Extremes **23**(1), 151–169 (2020)

Mai, J.F., Schenk, S., Scherer, M.: Analyzing model robustness via a distortion of the stochastic root: A Dirichlet prior approach. Statistics & Risk Modeling **32**(3-4), 177 – 195 (2015). https://www.degruyter.com/view/journals/strm/32/3-4/article-p177.xml

Mai, J.F., Schenk, S., Scherer, M.: Exchangeable exogenous shock models. Bernoulli **22**, 1278–1299 (2016). https://doi.org/10.3150/14-BEJ693

Mai, J.F., Schenk, S., Scherer, M.: Two novel characterizations of self-decomposability on the half-line. J. Theor. Probab. **30**, 365–383 (2017)

Mai, J.F., Scherer, M.: Reparameterizing Marshall–Olkin copulas with applications to sampling. J. Stat. Comput. Simul. **81**(1), 59–78 (2011). https://doi.org/10.1080/00949650903185961

Mai, J.F., Scherer, M.: Extendibility of Marshall–Olkin distributions and inverse pascal triangles. Braz. J. Probab. Stat. **27**(3), 310–321 (2013)

Mai, J.F., Scherer, M.: Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time. Extremes **17**(1), 77–95 (2014)

Mai, J.F., Scherer, M.: Simulating Copulas, 2nd edn. World Scientific (2017). https://doi.org/10.1142/10265

Mai, J.F., Scherer, M.: Subordinators which are infinitely divisible w.r.t. time: Construction, properties, and simulation of max-stable sequences and infinitely divisible laws. Lat. Am. J. Probab. Math. Stat. **16**, 1–29 (2019). https://doi.org/10.30757/ALEA.v16-35

Mai, J.F., Scherer, M., Shenkman, N.: Multivariate geometric distributions, (logarithmically) monotone sequences, and infinitely divisible laws. J. Multivar. Anal. **115**, 457–480 (2013). https://doi.org/10.1016/j.jmva.2012.11.012

Mansuy, R.: On processes which are infinitely divisible with respect to time. arXiv preprint math/0504408 (2005)

Marshall, A.W., Olkin, I.: Families of multivariate distributions. J. Am. Stat. Assoc. **83**(403), 834–841 (1988). https://doi.org/10.1080/01621459.1988.10478671

Marshall, A.W., Olkin, I.: Multivariate distributions generated from mixtures of convolution and product families, *Lecture Notes–Monograph Series*, vol. 16, pp. 371–393. Institute of Mathematical Statistics (1990). https://doi.org/10.1214/lnms/1215457574

Maruyama, G.: Infinitely divisible processes. Theory of Probability & Its Applications **15**(1), 1–22 (1970)

McNeil, A.J., Nešlehová, J.: Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. Ann. Stat. **37**, 3059–3097 (2009). https://doi.org/10.1214/07-AOS556

Mulinacci, S.: In: Marshall–Olkin Distributions - Advances in Theory and Applications, chap. Marshall–Olkin Machinery and Power Mixing: The Mixed Generalized Marshall–Olkin Distribution., pp. 65–86. Springer (2015)

Padoan, S.A.: Extreme dependence models based on event magnitude. J. Multivar. Anal. **122**, 1–19 (2013). https://doi.org/10.1016/j.jmva.2013.07.009

Regazzini, E., Lijoi, A., Prünster, I.: Distributional results for means of normalized random measures with independent increments. Ann. Stat. **31**(2), 560–585 (2003). https://doi.org/10.1214/aos/1051027881

Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes, vol. 4. Springer Science & Business Media (2007)

Rosiński, J.: Representations and isomorphism identities for infinitely divisible processes. Ann. Probab. **46**(6), 3229–3274 (2018). https://doi.org/10.1214/17-AOP1246

Sato, K.I.: Lévy processes and infinitely divisible distributions. Cambridge University Press (1999)

Skorohod, A.: Random Processes with Independent Increments. Mathematics and its Applications. Springer, Netherlands (1991)

Sloot, H.: The deFinetti representation of generalised Marshall–Olkin sequences. Dependence Modeling **8**(1), 107–118 (2020). https://doi.org/10.1515/demo-2020-0006

Sun, Y., Mendoza-Arriaga, R., Linetsky, V.: Marshall–Olkin distributions, subordinators, efficient simulation, and applications to credit risk. Adv. Appl. Probab. **49**(2), 481–514 (2017). https://doi.org/10.1017/apr.2017.10

Vatan, P.: Max-infinite divisibility and max-stability in infinite dimensions. In: Probability in Banach Spaces V, pp. 400–425. Springer Berlin Heidelberg (1985)

A.2 Exact simulation of continuous max-id processes with applications to exchangeable max-id sequences [8]

Summary

This paper provides an exact simulation algorithm for stochastic processes of the type

$$\tilde{\boldsymbol{X}} := \left(\tilde{X}_t \right)_{t \in T} := \left(\max_{i \in \mathbb{N}} f_i(t) \right)_{t \in T}, \tag{16}$$

where $M = \sum_{i \in \mathbb{N}} \delta_{f_i}$ denotes a Poisson random measure on the space of continuous functions. Stochastic processes with representation (16) are known as continuous max-id processes. These processes frequently appear when investigating or modeling maxima of independent stochastic processes, since, under mild conditions, the weak limit (if it exists) of the scaled pointwise maxima of independent stochastic process is necessarily max-id.

The exact simulation algorithm presented in this paper is a generalization of the simulation algorithm for continuous max-stable processes, a subclass of continuous max-id processes, presented in [19]. The simulation algorithm is exact in the sense that it allows to exactly simulate the max-id process $\tilde{\boldsymbol{X}}$ at a finite, but arbitrary, number of locations. Additionally, it provides an approximation of the whole process $\tilde{\boldsymbol{X}}$ from below. Besides generalizing the class of stochastic processes that may be simulated, the simulation algorithm proposed in this paper has two additional major advantages. First, I consider non-continuous margins, which may be relevant for practical applications, but is a feature that cannot be captured by max-stable processes. Second, I avoid the computation of the conditional (on $f(t) = k$) probability distribution of the intensity measure of the Poisson random measure M. This is usually a challenging and non-standard task, since the intensity measure of the Poisson random measure M is an infinite measure in most cases.

Further, I provide a complexity analysis of the proposed simulation algorithm. It shows that the expected number of random functions that need to be simulated to obtain an exact simulation of $\tilde{\boldsymbol{X}}$ at a certain prespecified number of locations is essentially equal to the number of these locations, plus an additional penalty term, which penalizes non-continuous marginal distributions.

The motivation to develop a simulation algorithm for continuous max-id processes stems from [10]. The main result of [10] is that we provide a stochastic representation of an exchangeable min-id sequence

$$-\tilde{\boldsymbol{X}} := \left(\inf \left\{ t \in \mathbb{R} \mid H(t) \geq E_i \right\} \right)_{i \in \mathbb{N}},$$

where H denotes a non-negative and non-decreasing infinitely divisible càdlàg process and $(E_i)_{i \in \mathbb{N}}$ denotes a sequences of i.i.d. exponential random variables with mean 1. Even though this stochastic representation is appealing and suggests to simulate $-\tilde{\boldsymbol{X}}$ as the first passage time of H over the $(E_i)_{i \in \mathbb{N}}$, it turns out that even the approximate simulation of the infinitely

divisible process H is usually quite challenging. Moreover, standard simulation methods via the Lebesgue density of the finite-dimensional distributions cannot be applied, since such densities usually do not exist or are not available in closed form. Thus, it was an open problem to find a description of $-\tilde{\boldsymbol{X}}$ which is suitable for simulation purposes.

The central observation with respect to this regard is that simple calculations provide that $\tilde{\boldsymbol{X}} := -(-\tilde{\boldsymbol{X}})$ may be viewed as a continuous max-id process with index set $T = \mathbb{N}$. The elaborations in Section 4 of the core article [8] show that the simulation algorithm applied to $\tilde{\boldsymbol{X}}$ essentially boils down to simulating a finite number of "simple" conditionally i.i.d. sequences, which is usually much easier than (approximately) simulating H. To illustrate the general simulation procedure for $\tilde{\boldsymbol{X}}$, I particularly derive a representation of the family of Sato-frailty sequences $-\tilde{\boldsymbol{X}}$ in terms of (16). The family of Sato-frailty sequences was introduced in [42], but simulation algorithms of such sequences were previously only available for some particular representatives of this family. As a side result, I characterize the Lévy measure of self-similar additive processes. Moreover, the application of the simulation algorithm to more general classes of exchangeable min-id sequences is sketched.

As a side result, I also provide a simulation algorithm which is tailored to finite dimensional max-id random vectors, i.e. max-id processes with index set $T = \{1, \ldots, d\}$. The reason for the special treatment of max-id random vectors is that the intensity measure of the associated Poisson random measure is usually described by exploiting the geometric structure of $\mathbb{R}^d$, e.g. as a scale mixture of probability distributions on a unit-sphere. Viewing these exponent measures as measures on the space of continuous functions on $\{1, \ldots, d\}$ is not nicely compatible with the inherent geometric structure, which illustrates the need of a special treatment of such max-id random vectors.

Individual contributions

I was fully responsible for generating the ideas, proofs and writing of the paper, even though I benefited from feedback of Matthias Scherer and Jan-Frederik Mai with regards to the presentation of the content.

I am the main author of this article.

Permission to include the article

Permissions

As a general rule, permission should be sought from the rights holder to reproduce any substantial part of a copyrighted work. This includes any text, illustrations, charts, tables, photographs, or other material from previously published sources. Obtaining permission to re-use content published by Elsevier is simple. Follow the guide below for a quick and easy route to permission.

Permission guidelines   ScienceDirect content   ClinicalKey content   Tutorial videos   Help and support

Permission guidelines

For further guidelines about obtaining permission, please review our Frequently Asked Questions below:

When is permission required?	⌄
When is permission not required?	⌄
From whom do I need permission?	⌄
How do I obtain permission to use photographs or illustrations?	⌄
Do I need to obtain permission to use material posted on a website such as Blogs/Google images/e-commerce websites?	⌄
What rights does Elsevier require when requesting permission?	⌄
How do I obtain permission from another publisher?	⌄
What is RightsLink/CCC?	⌄
What should I do if I am not able to locate the copyright owner?	⌄
Can I obtain permission from a Reproduction Rights Organization (RRO)?	⌄
Is Elsevier an STM signatory publisher?	⌄
Do I need to request permission to re-use work from another STM publisher?	⌄
Do I need to request permission to text mine Elsevier content?	⌄
Can I include/use my article in my thesis/dissertation?	⌃

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.

For any further clarifications, you can submit your query via our online form (https://service.elsevier.com/app/contact/supporthub/permissions-helpdesk)

Which uses of a work does Elsevier view as a form of 'prior publication'?	⌄
How do I obtain permission to use Elsevier Journal material such as figures, tables, or text excerpts, if the request falls within the STM permissions guidelines?	⌄
How do I obtain permission to use Elsevier Journal material such as figures, tables, or text excerpts, if the amount of material I wish to use does not fall within the free limits set out in the STM permissions guidelines?	⌄
How do I obtain permission to use Elsevier Book material such as figures, tables, or text excerpts?	⌄
How do I obtain permission to use Elsevier material that is NOT on ScienceDirect or Clinical Key?	⌄
Can I use material from my Elsevier journal article within my thesis/dissertation?	⌃

As an Elsevier journal author, you have the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see the Copyright page (https://www.elsevier.com/about/policies/copyright)for more information. No written permission from Elsevier is necessary.

This right extends to the posting of your thesis to your university's repository provided that if you include the published journal article, it is embedded in your thesis and not separately downloadable.

For any further clarifications, you can submit your query via our online form (https://service.elsevier.com/app/contact/supporthub/permissions-helpdesk)

Can I modify a figure when I have received permission to use it?	⌄
Can I modify Lancet material?	⌄
How do I obtain copies of a book in a format suitable for a person with a disability?	⌄
What is the National Institute of Health's (NIH) policy for depositing articles to PubMed Central?	⌄
How should I acknowledge Elsevier material?	⌄
How do I set up a RightsLink account?	⌄
As an Elsevier book author can I post my chapter on a website?	⌄

Copyright

Overview   Author rights   Institution rights   Government rights   Find out more

Overview

In order for Elsevier to publish and disseminate research articles, we need certain publishing rights from authors, which are determined by a publishing agreement between the author and Elsevier.

For articles published open access, the authors license exclusive rights in their article to Elsevier where a CC BY-NC-ND end user license is selected, and license non-exclusive rights where a CC BY end user license is selected.

For articles published under the subscription model, the authors typically transfer copyright to Elsevier. In some circumstances, authors may instead grant us (or the learned society for whom we publish) an exclusive license to publish and disseminate their work.

Regardless of whether they choose to publish open access or subscription with Elsevier, authors have many of the same rights under our publishing agreement, which support their need to share, disseminate and maximize the impact of their research.

For open access articles, authors will also have additional rights, depending on the Creative Commons end user license that they select. This Creative Commons license sets out the rights that readers (as well as the authors) have to re-use and share the article: please see here (https://www.elsevier.com/about/policies/open-access-licenses) for more information on how articles can be re-used and shared under these licenses.
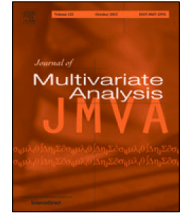
This page aims to summarise authors' rights when publishing with Elsevier; these are explained in more detail in the publishing agreement between the author and Elsevier.

Irrespective of how an article is published, Elsevier is committed to protect and defend authors' works and their reputation. We take allegations of infringement, plagiarism, ethical disputes, and fraud very seriously.

Author rights

The below table explains the rights that authors have when they publish with Elsevier, for authors who choose to publish either open access or subscription. These apply to the corresponding author and all co-authors.

Author rights in Elsevier's proprietary journals	Published open access	Published subscription
Retain patent and trademark rights	√	√
Retain the rights to use their research data freely without any restriction	√	√
Receive proper attribution and credit for their published work	√	√
Re-use their own material in new works without permission or payment (with full acknowledgement of the original article): 1. Extend an article to book length 2. Include an article in a subsequent compilation of their own work 3. Re-use portions, excerpts, and their own figures or tables in other works.	√	√
Use and share their works for scholarly purposes (with full acknowledgement of the original article): 1. In their own classroom teaching. Electronic and physical distribution of copies is permitted 2. If an author is speaking at a conference, they can present the article and distribute copies to the attendees 3. Distribute the article, including by email, to their students and to research colleagues who they know for their personal use 4. Share and publicize the article via Share Links, which offers 50 days' free access for anyone, without signup or registration 5. Include in a thesis or dissertation (provided this is not published commercially) 6. Share copies of their article privately as part of an invitation-only work group on commercial sites with which the publisher has a hosting agreement	√	√
Publicly share the preprint on any website or repository at any time.	√	√
Publicly share the accepted manuscript on non-commercial sites	√	√ using a CC BY-NC-ND license and usually only after an embargo period (see Sharing Policy (https://www.elsevier.com/about/policies/sharing) for more information)
Publicly share the final published article	√ in line with the author's choice of end user license	×

Exact simulation of continuous max-id processes with applications to exchangeable max-id sequences

Florian Brück

Technical University Munich, Chair of Mathematical Finance, Parkring 11, 85748 Garching, Germany

ABSTRACT

An algorithm for the unbiased simulation of continuous max-(resp. min-) infinitely divisible stochastic processes is developed. The algorithm only requires the simulation of finite Poisson random measures on the space of continuous functions and avoids the necessity of computing conditional distributions of infinite (exponent) measures. The complexity of the algorithm is characterized in terms of the expected number of simulated atoms of the Poisson random measures on the space of continuous functions. Special emphasis is put on the simulation of exchangeable max-(or min-) infinitely divisible sequences, in particular exchangeable Sato-frailty sequences. Additionally, exact simulation schemes of exchangeable exogenous shock models and exchangeable max-stable sequences are sketched.

© 2022 Elsevier Inc. All rights reserved.

1. Introduction

This paper provides an exact simulation algorithm for real-valued continuous stochastic processes $\boldsymbol{X} := (X_t)_{t \in T}$ with the property that for every given $n \in \mathbb{N}$ there exist independent and identically distributed (iid) stochastic processes $\left(\boldsymbol{X}^{(i,n)}\right)_{1 \le i \le n}$ such that

$$\boldsymbol{X} \sim \max_{1 \le i \le n} \boldsymbol{X}^{(i,n)}. \tag{1}$$

Such stochastic processes are called maximum-infinitely divisible (max-id) processes and they essentially constitute the class of possible weak limits of pointwise maxima of triangular arrays of independent stochastic processes [2]. Recently, max-id processes have attracted attention in the modeling of extreme events, see e.g. [6,16,29], while its subclass of max-stable processes is the central object of study in the extreme value theory of iid stochastic processes.

Under the assumption that $\boldsymbol{X}$ and $t \mapsto \sup\{x \in \mathbb{R} \mid \mathbb{P}(X_t > x) = 1\}$ are continuous, [2,14] show that $\boldsymbol{X}$ can be represented as the pointwise maximum of a (usually infinite) Poisson random measure (PRM) $N = \sum_{i \in \mathbb{N}} \delta_{f_i}$ on the space of continuous functions, i.e.,

$$\boldsymbol{X} \sim \max_{i \in \mathbb{N}} f_i. \tag{2}$$

The intensity measure $\mu(\cdot) := \mathbb{E}[N(\cdot)]$ of the PRM N is also called the exponent measure of $\boldsymbol{X}$ and it uniquely characterizes its distribution. The initial motivation for our simulation algorithm for $\boldsymbol{X}$ stems from [10, Algorithm 1], who have provided an exact simulation algorithm for continuous max-stable processes. In this paper, we generalize the ideas of [10] to a simulation algorithm for continuous max-id processes. The key ingredient of their and our simulation algorithm is the PRM

---

representation of X in (2) and its associated exponent measure. Basically, both simulation algorithms can be deduced from results of [11,12] about the conditional distribution of a specific decomposition of the PRM N. This specific decomposition of the PRM N allows to simulate only those functions which are relevant to determine the values of X at certain locations $t_1, \ldots, t_d$ and to approximate the whole sample path of X via the pointwise maximum over those finitely many functions. The mechanism of our simulation algorithm can be summarized as follows.

(i) Simulate only those functions $\left(f_j^{(1)}\right)_{1 \leq j \leq k_1}$ which maximize (2) at the first location t_1;

(ii) For $n \in \{2, \ldots, d\}$: Given the maximizing functions at locations $t_1, \ldots, t_{n-1}$, i.e. $\left\{f_j^{(i)} \mid 1 \leq i \leq n-1, \ 1 \leq j \leq k_i\right\}$, we only simulate those functions $\left(f_j^{(n)}\right)_{1 \leq j \leq k_n}$ which possibly contribute to the maximum in (2) at location t_n;

(iii) Use $\hat{X} = \max_{\{1 \leq j \leq k_n, 1 \leq n \leq d\}} f_j^{(n)}$ to approximate the sample path of X and additionally obtain $(X_{t_1}, \ldots, X_{t_d}) = (\hat{X}_{t_1}, \ldots, \hat{X}_{t_d})$.

Motivated by the recent results of [7], we apply the proposed simulation algorithm for continuous max-id processes to the simulation of exchangeable sequences of random variables $Y := (Y_i)_{i \in \mathbb{N}}$ with the property that for every $n \in \mathbb{N}$ there exist iid sequences of random variables $Y^{(i,n)} := \left(Y_j^{(i,n)}\right)_{j \in \mathbb{N}}$

$$Y \sim \min_{1 \leq i \leq n} Y^{(i,n)}. \tag{3}$$

Such sequences are known as minimum-infinitely divisible (min-id) sequences and are as well characterized by a so-called exponent measure [37]. It is obvious that $1/Y$ is a sequence of exchangeable random variables with stochastic representation (1), therefore simply being a particular example of a general continuous max-id process with index set $T = \mathbb{N}$. According to de Finetti's seminal theorem every exchangeable sequence of random variables admits the (unique) stochastic representation

$$Y \sim \left(\inf\left\{t \in \mathbb{R} \mid H_t \geq E_i\right\}\right)_{i \in \mathbb{N}}, \tag{4}$$

where $(E_i)_{i \in \mathbb{N}}$ is a sequence of independent and identically distributed (iid) Exponential random variables with unit mean and $(H_t)_{t \in \mathbb{R}}$ denotes a (unique in law) non-negative and non-decreasing (nnnd) stochastic process with càdlàg paths. [7] show that when Y has the stochastic representation (3) then the associated nnnd càdlàg process H satisfies the property that for every given $n \in \mathbb{N}$ there exist iid stochastic processes $\left(H^{(i,n)}\right)_{1 \leq i \leq n}$ such that

$$H \sim \sum_{1 \leq i \leq n} H^{(i,n)}. \tag{5}$$

Such processes are called infinitely divisible (id) and were extensively investigated in [32]. In analogy to the Lévy–Khintchine triplet of id random vectors on $\mathbb{R}^d$, nnnd id càdlàg processes are characterized by a so-called (path) Lévy measure on the space of càdlàg functions and a deterministic càdlàg (drift-)function [32].

In theory, the stochastic representation (4) immediately suggests a simulation algorithm for Y as the first passage times of the id process H over iid Exponential barriers. In practice, however, even the approximate simulation of the associated id process H is usually a challenging task. For instance, when the d-dimensional marginal distributions of Y becomes a multivariate Exponential distribution [26], then H must belong to the class of Lévy processes [22], i.e. H must have stationary and independent increments. Unfortunately, even for Lévy processes, exact simulation algorithms are only known for specific families and approximate simulation algorithms are extensively discussed in the literature, e.g. see [1,5,9]. Thus, the lack of the ability to simulate general id processes H limits the practical use of the stochastic representation (4), even though one may be able to analytically characterize the law of the id process H.

To overcome this challenge, we exploit the stochastic representation of $1/Y$ in terms of maxima over points of a Poisson random measure, which can be derived from (2) and the Lévy measure and drift of the associated id process H. More specifically, [7, Corollary 3.7] shows that the exponent measure of $1/Y$ can be uniquely characterized as a (possibly infinite) mixture of iid sequences in terms of the Lévy measure and drift of the associated id process H. This will allow us to construct an exact simulation algorithm for Y via $1/Y$, while essentially simulating a finite number of conditionally iid sequences.

The rather general theoretical results about the simulation of exchangeable min-id sequences are then used to derive an exact simulation algorithm for the class of exchangeable Sato-frailty sequences, which have been fully characterized analytically in [21]. Exchangeable Sato-frailty sequences can be characterized as the class of exchangeable min-id sequences associated to self-similar additive processes, i.e. they are associated to stochastically continuous càdlàg processes with independent increments which have the additional property that there exists some $\gamma > 0$ such that for all $a \geq 0$ we have $(H_{at})_{t \geq 0} \sim (a^\gamma H_t)_{t \geq 0}$, see e.g. [33, Section 3] for more details on self-similar additive processes. Even though analytical expressions of their multivariate marginal distributions are available, the simulation of such sequences has so far only been feasible for small sample sizes or some particular cases, which is due to the fact that the simulation of the associated self-similar additive process is generally complicated. We characterize the exponent measure of an exchangeable Sato-frailty

sequence in terms of the Lévy measure of the associated self-similar additive process and illustrate that our simulation algorithm essentially boils down to the simulation of two-dimensional random vectors.

In a recent article [38] have independently developed a simulation algorithm for continuous max-id processes on compact non-empty real domains T under the additional assumption of continuous marginal distributions. Their algorithm follows similar ideas as [10, Algorithm 1] translated to the max-id case. However, both of these algorithms require the computation of certain conditional distributions of the (infinite) exponent measure, which is usually a challenging task. Moreover, our framework is more general than that of [38], since we will explicitly consider arbitrary locally compact metric spaces T as index sets and non-continuous marginal distributions. This level of generality is necessary for our purposes, since we put special emphasis on simulation algorithms for exchangeable max-id sequences which have locally compact (but not compact) index sets and possibly non-continuous marginal distributions.

The remainder of the paper is organized as follows. Section 2 summarizes the theoretical background on continuous max-id processes. Section 3 introduces the exact simulation algorithm for continuous max-id processes and characterizes the complexity of the algorithm. In Section 4 we illustrate how our simulation algorithm for continuous max-id processes can be used to simulate exchangeable max-id sequences and we derive a particular exact simulation algorithm for exchangeable Sato-frailty sequences in Section 5. Section 6 provides a short example of how our simulation algorithm for exchangeable Sato-frailty sequences could be used in practice. Appendix A provides a general exact simulation algorithm tailored to max-id random vectors. Technical lemmas and proofs can be found in Appendix B.

2. Continuous max-id processes

Let us first introduce some notation. The index set T always denotes a locally compact metric space. Moreover, let $C(T) := \{f \mid f : T \to \mathbb{R} \text{ is continuous}\}$ denote the space of real-valued continuous functions on T equipped with the Borel σ-algebra generated by the topology of uniform convergence on compact sets. For some given function $h \in C(T)$ let $C_h(T) := \{f \mid f \in C(T), f \geq h, f \neq h\}$ denote the space of continuous functions dominating h. A real-valued stochastic process defined on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is denoted by $\boldsymbol{X} := (X_t)_{t \in T}$. Vectors in $\mathbb{R}^d$ are denoted in lower case bold letters. The projection of $\boldsymbol{X}$ to $\boldsymbol{t} := (t_1, \ldots, t_d)$ is denoted as $\boldsymbol{X_t} := (X_{t_1}, \ldots, X_{t_d})$. The operators max, min, inf, sup are always interpreted as pointwise operators, e.g. $\sup_{i \in \mathbb{N}} f_i$ is interpreted as the pointwise supremum of the functions $(f_i)_{i \in \mathbb{N}}$. The Dirac measure at a point f is denoted as δ_f. For a (random) point measure $N = \sum_{i \in \mathbb{N}} \delta_{f_i}$ we frequently use the notation $f \in N$ to denote that N has an atom at f, i.e. to denote that $N(\{f\}) \geq 1$. With this notation at hand we can state the definition of max-id processes and their associated vertices.

Definition 1 (*Max-id Process*). A stochastic process $\boldsymbol{X} \in \mathbb{R}^T$ is called max-id if for all $n \in \mathbb{N}$ there exist iid stochastic processes $\left(\boldsymbol{X}^{(i,n)}\right)_{1 \leq i \leq n}$ such that

$$\boldsymbol{X} \sim \max_{1 \leq i \leq n} \boldsymbol{X}^{(i,n)}.$$

The vertex of $\boldsymbol{X}$ is defined as the function

$$(h_{\boldsymbol{X}}(t))_{t \in T} := (\sup\{x \in \mathbb{R} \mid \mathbb{P}(X_t > x) = 1\})_{t \in T} \in [-\infty, \infty)^T.$$

The most common choices for the index set T of a max-id process are subsets of $\mathbb{R}^d$ and $\mathbb{Z}^d$. However, since requiring additional structure for T does not yield any simplifications in the following derivations, we keep the discussion as general as possible.

It is obvious that $g(\boldsymbol{X}) := (g(X_t))_{t \in T}$ defines a max-id process for every non-decreasing real-valued function g whenever $\boldsymbol{X}$ is a max-id process. This implies that, e.g., $\exp(\boldsymbol{X}) - \exp(h_{\boldsymbol{X}})$ defines a non-negative max-id process with vertex $\boldsymbol{0} := (0)_{t \in T}$. In this paper, we restrict the discussion to continuous max-id processes with continuous vertex, meaning that $h_{\boldsymbol{X}}$ and $t \mapsto X_t(\omega)$ are continuous functions for every $\omega \in \Omega$. Thus, we can assume that a continuous max-id process $\boldsymbol{X}$ with continuous vertex is non-negative with vertex $h_{\boldsymbol{X}} = \boldsymbol{0}$, since every continuous max-id processes $\boldsymbol{X}'$ with continuous vertex $h_{\boldsymbol{X}'}$ can be transformed to a continuous max-id process $\boldsymbol{X}$ with vertex $h_{\boldsymbol{X}} = \boldsymbol{0}$ by setting $\boldsymbol{X} := \exp(\boldsymbol{X}') - \exp(h_{\boldsymbol{X}'})$.

Under the assumption of a continuous and finite vertex, [11,14] have shown that a continuous max-id process $\boldsymbol{X}$ can be represented as the pointwise maxima of atoms of a Poisson random measure (PRM) on $C_{h_{\boldsymbol{X}}}(T)$. We summarize their results in the following theorem with the convention $\max_\emptyset := \boldsymbol{0}$.

Theorem 1 (*Spectral Representation of Continuous Max-ID Process [11,14]*).

1. *If $\boldsymbol{X}$ is a continuous max-id process with vertex $h_{\boldsymbol{X}} = \boldsymbol{0}$ then there exists a PRM N on $C_{\boldsymbol{0}}(T)$ with locally finite intensity measure μ, called exponent measure, which satisfies*

$$\mu\left(\left\{f \in C_{\boldsymbol{0}}(T) \;\middle|\; \sup_{k \in K} f(k) > \epsilon\right\}\right) < \infty \text{ for all compact } K \subset T \text{ and } \epsilon > 0 \tag{6}$$

such that

$$\boldsymbol{X} \sim \max_{f \in N} f.$$
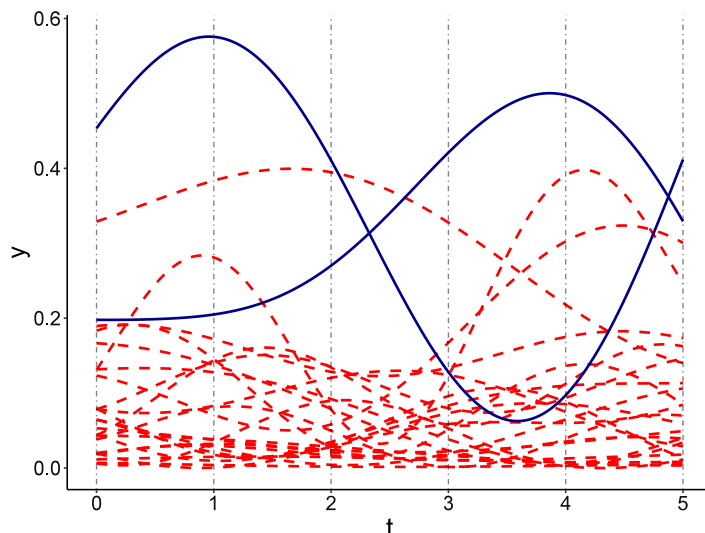
Fig. 1. Illustration of extremal and subextremal functions of a PRM N. Functions in solid-blue belong to $N^+_{(0,\ldots,5)}$, functions in dashed-red belong to $N^-_{(0,\ldots,5)}$.

2. *Conversely, given a locally finite measure μ on $C_{\mathbf{0}}(T)$ which satisfies* (6), *there exists a PRM N on $C_{\mathbf{0}}(T)$ with intensity μ such that*

$$\boldsymbol{X} := \max_{f \in N} f$$

defines a continuous max-id process with vertex $h_{\boldsymbol{X}} = \mathbf{0}$.

It is easy to see that $\mathbb{P}\big(N(C_{\mathbf{0}}(T)) = \infty\big) = 1$ if and only if μ is an infinite measure. For example, this is the case if $\mathbb{P}(X_t > 0) = 1$ for some $t \in T$. Since this is a desired property in many applications, a simulation of $\boldsymbol{X}$ via the simulation of the infinite PRM N is usually practically infeasible. However, it is crucial to observe that the value of $\boldsymbol{X_t} := (X_{t_1}, \ldots, X_{t_d})$ is fully determined by the atoms of the random measure of extremal functions at $\boldsymbol{t}$

$$N^+_{\boldsymbol{t}} := \sum_{f \in N} \delta_f \mathbf{1}_{\left\{ f(t_i) = X_{t_i} \text{ for some } 1 \leq i \leq d \right\}}. \tag{7}$$

Thus, all atoms of the random measure of subextremal functions at $\boldsymbol{t}$

$$N^-_{\boldsymbol{t}} := \sum_{f \in N} \delta_f \mathbf{1}_{\left\{ f(t_i) < X_{t_i} \text{ for all } 1 \leq i \leq d \right\}} \tag{8}$$

are irrelevant when we are solely interested in $\boldsymbol{X_t}$. $N^+_{\boldsymbol{t}}$, resp. $N^-_{\boldsymbol{t}}$, are called the extremal, resp. subextremal, point measure at $\boldsymbol{t}$. Fig. 1 illustrates the extremal and subextremal functions of a continuous max-id process on $\mathbb{R}$ with $\boldsymbol{t} = (0, 1, \ldots, 5)$. [12, Section 2] analyze the extremal and subextremal random point measures of a continuous max-id process and show that they are indeed well-defined. Moreover, they show that

$N^+_{\boldsymbol{t}}$ *is an almost surely finite random measure if and only if one of the following conditions is satisfied*:

(i) $\mu(C_{\mathbf{0}}(T)) < \infty$, *or*

(ii) $\mu(C_{\mathbf{0}}(T)) = \infty$ *and* $\min_{1 \leq i \leq d} X_{t_i} > 0$ *almost surely.* $\tag{9}$

If one of the conditions in (9) is satisfied, it is sufficient to determine the finite number of atoms of the random measure $N^+_{\boldsymbol{t}}$ in order to obtain $\boldsymbol{X_t}$ via

$$\boldsymbol{X_t} = \left(\max_{f \in N^+_{\boldsymbol{t}}} f(t_1), \ldots, \max_{f \in N^+_{\boldsymbol{t}}} f(t_d) \right).$$

Additionally, the random measure $N^+_{\boldsymbol{t}}$ also yields an approximation (from below) of the whole sample path of $\boldsymbol{X}$ via

$$\boldsymbol{X} \approx \hat{\boldsymbol{X}} := (\hat{X}_t)_{t \in T} := \left(\max_{f \in N^+_{\boldsymbol{t}}} f(t) \right)_{t \in T}.$$

Thus, to obtain an exact simulation of $\boldsymbol{X_t}$ and to approximate the sample path of $\boldsymbol{X}$ via $\hat{\boldsymbol{X}}$ we simply need to focus on simulation algorithms of the finite random point measure $N^+_{\boldsymbol{t}}$.

The main ingredient of our simulation algorithm for $N_{\boldsymbol{t}}^+$ will be based on the conditional distribution of $N_{\boldsymbol{t}}^-$ given $N_{\boldsymbol{t}}^+$, which is derived in [11, Lemma 3.2]. More specifically, it is shown that the conditional distribution of $N_{\boldsymbol{t}}^-$ given $N_{\boldsymbol{t}}^+$ is given by the distribution of a PRM with intensity $\mathbf{1}_{\{f(t_i)<X_{t_i},\,1\le i\le d\}}\mathrm{d}\mu(f)$. To illustrate the implications of this result, let us assume we are given an initialization $N_{t_1}^+ \subset N_{\boldsymbol{t}}^+ \ne N$ of $N_{\boldsymbol{t}}^+$. To obtain $N_{(t_1,t_2)}^+$ we only need to consider those atoms of $N_{t_1}^-$ which belong to $N_{t_2}^+$. Given $N_{t_1}^+$, the random measure $N_{t_2}^+ \setminus N_{t_1}^+$ is the restriction of $N_{t_1}^-$ to the (measurable) set

$$\left\{\tilde{N} \text{ extremal point measure on } C_{\mathbf{0}}(T) \text{ at location } t_2 \text{ and concentrated on } \left\{f(t_2) \ge \max_{\tilde{f}\in N_{t_1}^+} \tilde{f}(t_2)\right\}\right\}.$$

Now, [11, Lemma 3.2] implies that, conditional on $N_{t_1}^+$, the random measure $N_{t_2}^+ \setminus N_{t_1}^+$ has the same distribution as $\mathrm{argmax}_{f\in\bar{N}} f(t_2)$, where $\bar{N}$ is a PRM with intensity

$$\mathbf{1}_{\left\{f(t_1)<\max_{\tilde{f}\in N_{t_1}^+}\tilde{f}(t_1) \text{ and } f(t_2)\ge\max_{\tilde{f}\in N_{t_1}^+}\tilde{f}(t_2)\right\}}\mathrm{d}\mu(f).$$

Assuming that $\max_{\tilde{f}\in N_{t_1}^+}\tilde{f}(t_2)$ is positive, (6) implies that $\bar{N}$ is a **finite** PRM. Therefore, one may simulate $N_{\boldsymbol{t}}^+$ by iterative simulation of finite PRMs with intensities

$$\mathbf{1}_{\left\{f(t_i)<\max_{\tilde{f}\in N_{(t_1,\dots,t_n)}^+}\tilde{f}(t_i) \text{ for all } 1\le i\le n \text{ and } f(t_{n+1})\ge\max_{\tilde{f}\in N_{(t_1,\dots,t_n)}^+}\tilde{f}(t_{n+1})\right\}}\mathrm{d}\mu(f), \quad 1\le n\le d-1. \tag{10}$$

From a practical perspective one should note that it is sufficient to be able to simulate from a finite PRM with intensity $\mathbf{1}_{\{f(t)\ge c\}}\mathrm{d}\mu(f)$ for all $t\in T$ and $c>0$ to simulate the PRMs with intensities (10), assuming that $\max_{\tilde{f}\in N_{(t_1,\dots,t_n)}^+}\tilde{f}(t_{n+1})>0$. To verify the claim, recall that the restriction of any PRM $\hat{N}$ with intensity $\hat{\mu}$ to an arbitrary measurable set A again defines a PRM with intensity $\mathbf{1}_{\{f\in A\}}\mathrm{d}\hat{\mu}(f)$. Thus, to simulate a PRM with intensity (10), one can simulate a finite PRM with intensity

$$\mathbf{1}_{\left\{f(t_{n+1})\ge\max_{\tilde{f}\in N_{(t_1,\dots,t_n)}^+}\tilde{f}(t_{n+1})\right\}}\mathrm{d}\mu(f)$$

and simply ignore those atoms which do not satisfy the constraints in (10).

Remark 1 (*Infinite $N_{\boldsymbol{t}}^+$*). It is easy to see that the event $X_{t_i}=0$ implies $N_{t_i}^+=N$. Thus, when μ is an infinite measure, the simulation of $N_{t_i}^+$ requires the simulation of infinitely many atoms with probability $\mathbb{P}\left(X_{t_i}=0\right)=\exp(-\mu(\{f\in C_{\mathbf{0}}(T)\mid f(t_i)>0\}))$. However, one may avoid this unpleasant situation by discarding finite exponent measures from μ. Consider the set of possibly 0-valued locations

$$J_0 := \left\{j\in\{1,\dots,d\}\mid \mathbb{P}\left(X_{t_j}=0\right)>0\right\}$$

and consider the exponent measures of the form

$$\mu_j(\cdot) = \mu\left(\cdot\cap\{f\in C_{\mathbf{0}}(T)\mid f(t_j)>0, f(t_k)=0,\ k<j,\ k\in J_0\}\right),\ j\in J_0. \tag{11}$$

Note that the μ_j are supported on disjoint sets and that each μ_j is finite, since $0<\mathbb{P}\left(X_j=0\right)\le\exp\left(-\mu_j\left(C_{\mathbf{0}}(T)\right)\right)$. Therefore, it is possible to (exactly) simulate independent max-id processes $(\hat{\boldsymbol{X}}_j)_{j\in J_0}$ by the simulation of independent PRMs with finite exponent measures $(\mu_j)_{j\in J_0}$. It remains to consider the residual of the exponent measure μ, given by $\tilde{\mu}:=\mu-\sum_{j\in J_0}\mu_j$, which is more easily described as

$$\tilde{\mu}(\cdot) = \mu\left(\cdot\cap\{f\in C_{\mathbf{0}}(T)\mid f(t_j)=0, j\in J_0\}\right). \tag{12}$$

Let $\tilde{N}$ denote a PRM with intensity $\tilde{\mu}$ and let $\tilde{\boldsymbol{X}}$ denote the continuous max-id process associated with the exponent measure $\tilde{\mu}$. It is not difficult to show that $\tilde{\mu}$ is either vanishing or an infinite measure and that $\tilde{\boldsymbol{X}}$ satisfies $\mathbb{P}\left(\tilde{X}_{t_j}=0\right)=1$ for all $j\in J_0$. Therefore, the exact simulation of $\tilde{\boldsymbol{X}}_{\boldsymbol{t}}$ only involves the simulation of the finite random measure $\tilde{N}_{(t_i)_{i\notin J_0}}^+$. Moreover, it is easily seen that $\boldsymbol{X}$ admits the representation

$$\boldsymbol{X} = \max\left\{\tilde{\boldsymbol{X}}; \max_{j\in J_0}\hat{\boldsymbol{X}}_j\right\},$$

which shows that $\boldsymbol{X}_t$ can be determined by the pointwise maxima of finitely many finite random point measures.

So far, we have assumed that we are given a finite initialization $N_{t_1,\dots,t_n}^+$ of $N_{\boldsymbol{t}}^+$ with $\max_{f\in N_{t_1,\dots,t_n}^+} f(t_{n+1})>0$ and, under this assumption, we have shown that we only need to simulate from finite PRMs to obtain $N_{t_1,\dots,t_{n+1}}^+$, resp. $\hat{\boldsymbol{X}}$. In Section 3 we show that the ability to simulate from a PRM with intensity $\mathbf{1}_{\{f(t)\ge c\}}\mathrm{d}\mu(f)$ for every $t\in T$ and $c>0$ is also sufficient to obtain such initializations of $N_{\boldsymbol{t}}^+$. Thus, we construct an algorithm for the exact simulation of $\boldsymbol{X}_t$ and approximation of $\boldsymbol{X}$ via $\hat{\boldsymbol{X}}$, which solely requires the ability to simulate finite PRMs with intensities $\mathbf{1}_{\{f(t)\ge c\}}\mathrm{d}\mu(f)$ for all $t\in T$ and $c>0$.

3. Exact simulation of continuous max-id processes

The main ingredient of our algorithm is the possibility to simulate from the finite PRMs with intensities $\mathbf{1}_{\{f(t)\geq c\}}\mathrm{d}\mu$ for all $t \in T$ and $c > 0$. Based on our developments in Section 2, Algorithm 1 provides an exact simulation procedure for a continuous max-id process with vertex $\mathbf{0}$.

Algorithm 1: Exact simulation of continuous max-id process with vertex $\mathbf{0}$

Result: Unbiased sample of $(X_{t_1}, \ldots, X_{t_d})$ and approximation of the max-id process $(X_t)_{t\in T}$.

1 Set $\mu_j(\cdot) = \mu\left(\cdot \cap \left\{f \in C_{\mathbf{0}}(T) \mid f(t_j) > 0, f(t_k) = 0, \; k < j, \; k \in J_0\right\}\right)$, $j \in J_0$;

2 Set $\tilde{\mu}(\cdot) = \mu\left(\cdot \cap \left\{f \in C_{\mathbf{0}}(T) \mid f(t_j) = 0, j \in J_0\right\}\right)$;

3 **for** $j \in J_0$ **do**

4   Simulate a finite PRM N_j with intensity μ_j and set $\hat{X}_j = \max_{f\in N_j} f$;

5 **end**

6 Set $\tilde{X} = \mathbf{0}$;

7 **for** $i = 1, \ldots, d, i \notin J_0$ **do**

8   **if** $\tilde{X}_{t_i} = 0$ **then**

9   Set $\tilde{N}^+ = \emptyset$;

10   Set $c_u = \infty$ and $c_l = c$ for some $c > 0$;

11   **while** $\tilde{N}^+ = \emptyset$ **do**

12   Simulate a finite PRM $\tilde{N}^+$ with intensity $\mathbf{1}_{\{c_u > f(t_i) \geq c_l\}}\mathrm{d}\tilde{\mu}(f)$;

13   **for** $f \in \tilde{N}^+$ **do**

14   **if** $f(t_k) \geq \tilde{X}_{t_k}$ *for some* $k < i, k \notin J_0$ **then**

15   Set $\tilde{N}^+ = \tilde{N}^+ \setminus \{f\}$

16   **end**

17   **end**

18   Set $c_u = c_l$ and set $c_l = c_l/2$;

19   **end**

20   **else**

21   Simulate a finite PRM $\tilde{N}^+$ with intensity $\mathbf{1}_{\{f(t_i)\geq \tilde{x}_{t_i}\}}\mathrm{d}\tilde{\mu}(f)$;

22   **for** $f \in \tilde{N}^+$ **do**

23   **if** $f(t_k) \geq \tilde{X}_{t_k}$ *for some* $k < i, k \notin J_0$ **then**

24   Set $\tilde{N}^+ = \tilde{N}^+ \setminus \{f\}$

25   **end**

26   **end**

27   **end**

28   Set $\tilde{N}^+_{t_i} = \left\{f \in \tilde{N}^+ \mid f(t_i) \geq \tilde{f}(t_i) \text{ for all } \tilde{f} \in \tilde{N}^+\right\}$;

29   Set $\tilde{X} = \max\left\{\max_{f\in\tilde{N}^+_{t_i}} f, \tilde{X}\right\}$;

30 **end**

31 Set $\hat{X} = \max\left\{\max_{j\in J_0} \hat{X}_j, \tilde{X}\right\}$;

32 **return** $\hat{X}$

The validity of Algorithm 1 is verified in the following theorem.

Theorem 2 (*Validity of Algorithm 1*). *Let X denote a continuous max-id process with vertex $h_X = \mathbf{0}$ and exponent measure μ. Then, Algorithm 1 stops after finitely many steps and its output $\hat{X}$ approximates X from below and satisfies $\hat{X}_t \sim X_t$.*

Clearly, Algorithm 1 reduces to lines 6–30 if no margin of X_t has an atom at 0, since in such cases we have $J_0 = \emptyset$ and $\tilde{\mu} = \mu$. Moreover, it is worth mentioning that, even though $\hat{X}_t$ is max-id, the stochastic process $\hat{X} \leq X$ is generally not max-id, since N_t^+ is not a PRM on $C_{\mathbf{0}}(T)$.

Remark 2 (*Reason for Splitting μ into $\sum_{j\in J_0} \mu_j + \tilde{\mu}$*). The reason for splitting μ into the disjoint parts μ_j and $\tilde{\mu}$ is to divide the simulation of X into separate simulations of finite random measures. First, we directly simulate $\left(X_{t_j}\right)_{j\in J_0}$, i.e. those locations at which $\{X_{t_j} = 0\}$ occurs with positive probability, since a naive simulation of X_{t_j} via the respective extremal functions at t_j may result in the necessity of simulating an infinite PRM with positive probability (Remark 1). Second,

we simulate those atoms of a PRM N with intensity μ, which have not been simulated yet and possibly contribute to $X_t = \max_{f \in N} f(t)$. Since $(\tilde{X}_{t_j})_{j \in J_0} = \mathbf{0}$ by the definition of $\tilde{\mu}$, we can solely focus on the simulation of $(\tilde{X}_{t_j})_{j \notin J_0}$. This precisely requires the simulation of the extremal functions at $(t_j)_{j \notin J_0}$ of the PRM $\tilde{N}$ with intensity $\tilde{\mu}$. The key observation is that the definition of the μ_j ensures that $\tilde{X}$ does not have atoms at 0 at the locations $(t_j)_{j \notin J_0}$, which implies that the extremal point measure $\tilde{N}^+_{(t_j)_{j \notin J_0}}$ is finite by (9). Therefore, we can use (10) to obtain a sample of $\tilde{N}^+_{(t_j)_{j \notin J_0}}$ via the simulation of finite PRMs. Combining the two simulated processes by taking pointwise maxima we obtain an approximation $\hat{X}$ of X which satisfies $\hat{X}_t \sim X_t$.

Remark 3 (*Simulation Algorithm for Max-stable Processes [10]*). A max-stable process with unit Fréchet margins can be represented as $X \sim \max_{i \in \mathbb{N}} \zeta_i \psi_i$, where $N = \sum_{i \in \mathbb{N}} \delta_{(\zeta_i, \psi_i)}$ is a PRM with intensity $\mathrm{d}\mu = s^{-2} \mathrm{d}s \mathrm{d}Q$ and Q is a probability measure on $C_{\mathbf{0}}(T)$ such that $\int \psi(t) Q(\mathrm{d}\psi) = 1$ for all $t \in T$. In this case, one can show that N^+_t only contains a single function, denoted as $\hat{\psi}_t$. The regular conditional distribution of $\hat{\psi}_t$ given $X_t = z$ is obtained in [12, Proposition 4.2]. This result can be used to represent the PRM with intensity $\mathbf{1}_{\{s\psi(t) > 0\}} s^{-2} \mathrm{d}s Q(\mathrm{d}\psi)$ as a PRM with intensity $s^{-2} \mathrm{d}s \mathrm{d}Q_t$, where Q_t denotes the conditional distribution of $\hat{\psi}_t / X_t$ given X_t. Thus, one may simulate a PRM with intensity $\mathbf{1}_{\{f(t) \geq c\}} \mathrm{d}\mu$ by successively simulating points of a PRM with intensity $\mathbf{1}_{\{s \geq c\}} s^{-2} \mathrm{d}s \mathrm{d}Q_t$. With this specific procedure for the simulation of a PRM with intensity $\mathbf{1}_{\{f(t) \geq c\}} \mathrm{d}\mu$, Algorithm 1 essentially reduces to the exact simulation algorithm of continuous max-stable processes in [10].

Remark 4 (*Conditional Distribution of Max-id Process*). Similar to max-stable processes with unit Fréchet margins, [12, Proposition 4.1] provides the conditional distribution of the extremal function of a continuous max-id process X with continuous marginal distributions at a location t, given that $X_t = z$. Intuitively, the conditional distribution can be described as the regular conditional distribution of the exponent measure μ given $X_t = z$, denoted as $Q_{t,z}$, where the formal definition of a regular conditional distribution of a possibly infinite exponent measure can be found in [12, Appendix A2]. Thus, the extremal function for a single location t can be found by first drawing a random variable $Z \sim X_t$ and then drawing the extremal function according to $Q_{t,z}$. Surprisingly, not only the extremal function at a location t follows the conditional (on Z) distribution $Q_{t,z}$, but so do the subextremal functions. More formally, assume that you are given a PRM $\sum_{i \in \mathbb{N}} \delta_{Z_i}$ on $(0, \infty)$ where $Z := \max_{i \in \mathbb{N}} Z_i \sim X_t$. Then, conditioned on $(Z_i)_{i \in \mathbb{N}}$, the PRM with intensity $\mathbf{1}_{\{f(t) > 0\}} \mathrm{d}\mu$ can be represented as $\sum_{i \in \mathbb{N}} \delta_{f_{Z_i}}$, where the $f_{Z_i} \sim Q_{t,Z_i}$ are independent. [38] have recently and independently proposed an algorithm for the exact simulation of max-id processes, which is based on the just described procedure to simulate a PRM with intensity $\mathbf{1}_{\{f(t) > 0\}} \mathrm{d}\mu$. However, determining and simulating the conditional distribution $Q_{t,z}$ of an exponent measure is a challenging task and is only a sufficient but not a necessary criterion for the simulation of the PRM with intensity (10).

Remark 5 (*Simulation Algorithm for Max-id Random Vectors*). Exponent measures of non-negative max-id random vectors are often described via the geometric structure of $[0, \infty)^d$, e.g. as scale mixtures of probability distributions on unit spheres. Examples of such families of max-id random vectors are given by random vectors with reciprocal Archimedean copula [13], max-stable distributions [31] and reciprocals of exogenous shock models [34]. For these families it is generally surprisingly inconvenient to apply Algorithm 1 due to the difficulty of describing the PRM with intensity $\mathbf{1}_{\{f(i) \geq c\}} \mathrm{d}\mu$ in a simple manner. Therefore, we provide a simulation algorithm which is tailored to the specific representations of exponent measures on $[0, \infty)^d$ in Appendix A.

3.1. Complexity of Algorithm 1

The main difficulty of Algorithm 1 lies in the simulation of the atoms (functions) of the PRMs in line 12 and 21. Therefore, to analyze the complexity of Algorithm 1, we may focus on the number of functions that need to be simulated to obtain $\hat{X}$. Since the number of simulated functions during the execution of Algorithm 1 is a random variable, we characterize its complexity in terms of the expected number of simulated functions. To this purpose, we extend a result by [27,28] about the expected size of the extremal point measure of max-stable processes at locations $t = (t_i)_{1 \leq i \leq d}$ to max-id processes.

Lemma 1. *Let X denote a continuous max-id process with vertex $h_X = \mathbf{0}$ and exponent measure μ. The expected size of the extremal point measure at location $t = (t_1, \ldots, t_d)$ is given by*

$$\mathbb{E}\left[|N^+_t|\right] = \mathbb{E}\left[\int_{C_{\mathbf{0}}(T)} \mathbf{1}_{\{f(t_i) \geq X_{t_i} \text{ for some } 1 \leq i \leq d\}} \mathrm{d}\mu(f)\right].$$

To deduce the expected number of simulated functions during the execution of Algorithm 1 we additionally assume that the simulation of the atoms of a PRM with intensity $\tilde{\mu}$ can be conducted in a top-down fashion as follows:

For all $(t_i)_{i \notin J_0}$ we assume that we can consecutively simulate the atoms $\left(f_j^{(i)}\right)_{j \in \mathbb{N}}$ of a PRM

$$\tilde{N} = \sum_{j \in \mathbb{N}} \delta_{f_j^{(i)}} \text{ with intensity } \tilde{\mu} \text{ such that } f_1^{(i)}(t_i) \geq f_2^{(i)}(t_i) \geq \cdots . \tag{13}$$

Assumption (13) allows to conduct lines 7–30 of Algorithm 1 more efficiently: For a fixed $i \notin J_0$ one consecutively simulates $f_1^{(i)}, f_2^{(i)}, \ldots$ such that $f_1^{(i)}(t_i) \geq f_2^{(i)}(t_i) \geq \cdots$ and stops as soon as one has found all extremal functions at a location t_i. All extremal functions at location t_i are found as soon as one has found a $j \in \mathbb{N}$ such that $f_j^{(i)}$ is an extremal function at location t_i and $f_{j+1}^{(i)}$ is subextremal function at location t_i. In general, it is necessary to simulate the $f_j^{(i)}$ until the first subextremal function at a location t_i is found, since $\tilde{X}$ may not have continuous marginal distributions and there may exist more than one extremal function at a location $(t_i)_{i \notin J_0}$. Of course, if the distribution of $\tilde{X}_{t_i}$ is continuous, one can stop as soon as the first extremal function at location $(t_i)_{i \notin J_0}$ is found, since there can only exist one extremal function at each continuous margin of $\tilde{X}$ by [12, Proposition 2.5]. Thus, assumption (13) allows to avoid the simulation of more than one subextremal function at each location $(t_i)_{i \notin J_0}$, which may not be excluded if one conducts Algorithm 1 in its original formulation of Theorem 2.

For the remainder of this subsection we assume that Algorithm 1 is conducted according to assumption (13). Assumption (13) may be regarded as reasonable, since it is satisfied for many continuous max-id processes. For instance, it is satisfied if one assumes that $\tilde{X}$ has continuous marginal distributions and that one conducts the simulation of the PRMs in lines 12 and 21 of Algorithm 1 based on the conditional distribution of a max-id process as described in Remark 4. Moreover, the assumption may also be satisfied when simulating certain exchangeable max-id sequences, see Sections 5 and 6.

Theorem 3. *Under the assumption that a PRM with intensity $\tilde{\mu}$ may be simulated according to assumption* (13)*, the expected number of simulated functions during the execution of Algorithm 1 is given by*

$$d - |J_0| + \mu \left(\left\{ f \in C_{\mathbf{0}}(T) \mid f(t_j) > 0 \text{ for some } j \in J_0 \right\} \right) + \sum_{i \notin J_0} \mathbb{E} \left[\tilde{\mu} \left(\left\{ f \in C_{\mathbf{0}}(T) \mid f(t_i) \in \left[\tilde{X}_{t_i}, \infty \right) \right\} \right) \right].$$

Moreover, when $\tilde{X}_{(t_i)_{i \notin J_0}}$ has continuous marginal distributions, the expected number of simulated functions during the execution of Algorithm 1 is equal to

$$d - |J_0| + \mu \left(\left\{ f \in C_{\mathbf{0}}(T) \mid f(t_j) > 0 \text{ for some } j \in J_0 \right\} \right).$$

Theorem 3 may be interpreted as follows: The expected number of simulated functions is equal to the number of locations where $\mathbf{X}$ has continuous margins plus an additional term which accounts for the possibility that $|N_t^+| > 1$ is possible at locations where $\mathbf{X}$ has non-continuous margins.

It is easy to see that Theorem 3 includes the complexity characterization [10, Proposition 9] of the algorithm for the simulation of continuous max-stable processes described in Remark 3. There, the authors showed that the expected number of simulated functions in their algorithm is equal to the number of locations where the continuous max-stable process is simulated exactly. Theorem 3 shows that the same result holds when Algorithm 1 is applied to continuous max-id processes with continuous margins. Thus, when measuring simulation complexity only in terms of the expected number of simulated functions, there is no increase in simulation complexity when considering a continuous max-id processes with continuous margins instead of a continuous max-stable process. Moreover, it follows that, as a byproduct, we have shown that the expected number of simulated functions in the algorithm for exact simulation of a continuous max-id process with continuous margins and compact index set of [38] is equal to the number of locations where the max-id process is simulated exactly, since it is exactly based on the assumption that the PRMs appearing in Algorithm 1 may be simulated according to assumption (13).

4. Exact simulation of exchangeable max(min)-id sequences

When considering max-id sequences, i.e. $T = \mathbb{N}$, the assumption of continuity of the max-id process $\mathbf{X}$ is irrelevant, since $C_{\mathbf{0}}(\mathbb{N}) = [0, \infty)^{\mathbb{N}} \setminus \{\mathbf{0}\}$. Therefore, Algorithm 1 is applicable to all max-id sequences with vertex $h_{\mathbf{X}} = \mathbf{0}$, which may be satisfied for every max-id sequence after suitable transformations of the margins. However, to apply Algorithm 1, it remains to find a suitable description of the exponent measure of a max-id sequence $\mathbf{X}$ on $[0, \infty)^{\mathbb{N}} \setminus \{\mathbf{0}\}$ such that the PRM with intensity $\mathbf{1}_{\{f(n) \geq c\}}$ can be simulated. To achieve this, we focus on the results of [7], who describe the structure of exponent measures of exchangeable min-id sequences, i.e. of exchangeable sequences $\mathbf{Y} := 1/\mathbf{X}$, where $\mathbf{X}$ is max-id. To this purpose, let us recall the most important results of [7].

Theorem 4 ([7, Corollary 3.7]). $\boldsymbol{Y} \in (0, \infty]^{\mathbb{N}}$ *is an exchangeable min-id sequence if and only if*

$$\boldsymbol{Y} \sim (\inf\{t \geq 0 \mid H(t) \geq E_i\})_{i \in \mathbb{N}}, \tag{14}$$

where $(E_i)_{i \in \mathbb{N}}$ *are iid Exp(1) and* $H = (H_t)_{t \geq 0} \in [0, \infty]^{[0,\infty)}$ *is a (unique in law) nnnd id càdlàg process which satisfies* $H_0 = 0$.

We say that an exchangeable max-id sequence $\boldsymbol{X}$ corresponds to an id process H if and only if $\boldsymbol{Y} = 1/\boldsymbol{X}$ is the exchangeable min-id sequence corresponding to H. Similar to max-id sequences, the survival function of $\boldsymbol{Y}$ can be expressed in terms of an exponent measure $\bar{\mu}$. It can be related to the exponent measure of $\boldsymbol{X}$ noting that $\mathbb{P}(\boldsymbol{Y} > \boldsymbol{x}) = \mathbb{P}\left(\boldsymbol{X} < \frac{1}{\boldsymbol{x}}\right) = \exp\left(-\mu\left(\left(-\infty, \frac{1}{\boldsymbol{x}}\right)^{\complement}\right)\right) = \exp\left(-\bar{\mu}\left((\boldsymbol{x}, \infty]^{\complement}\right)\right)$, where $\bar{\mu}(A) := \mu\left(\{\boldsymbol{x} \in [0, \infty)^{\mathbb{N}} \mid 1/\boldsymbol{x} \in A\}\right)$ is called the exponent measure of the exchangeable min-id sequence $\boldsymbol{Y}$. From this relation it is easy to see that a PRM $\bar{N} = \sum_{i=1}^{\infty} \delta_{\boldsymbol{f}_i}$ with intensity $\bar{\mu}$ can be transformed to a PRM $N = \sum_{i=1}^{\infty} \delta_{1/\boldsymbol{f}_i}$ with intensity μ. Thus, we can generate atoms of N by taking reciprocals of atoms of $\bar{N}$. In the following, we will show how the correspondence (14) can be used to generate atoms from $\bar{N}$ (and thus from N).

Let us recall several facts about id processes. It is well-known that id random vectors on $[0, \infty]^d$ are in one-to-one correspondence with a pair (υ, b), where υ is a (Lévy)measure on $[0, \infty]^d \setminus \{\boldsymbol{0}\}$ satisfying certain integrability conditions and $b \in [0, \infty]^d$ is a deterministic (drift)vector. [32] has elegantly extended this characterization to id processes and [7] have used these results to prove that the Laplace-transform of an nnnd id càdlàg process which satisfies $H_0 = 0$ is given by

$$\mathbb{E}\left[\exp\left(-\sum_{i=1}^d a_i H_{t_i}\right)\right] = \exp\left(-\sum_{i=1}^d a_i b(t_i) - \int_{\mathbf{M}} 1 - \exp\left(-\sum_{i=1}^d a_i g(t_i)\right) \mathrm{d}\nu(g)\right), \quad \boldsymbol{a}, \boldsymbol{t} \in [0, \infty)^d, \; d \in \mathbb{N}, \tag{15}$$

where ν is a unique (Lévy)measure on the path space

$$\mathbf{M} := \left\{g : [0, \infty) \to [0, \infty] \mid g(0) = 0, g \text{ nnnd and càdlàg}, \; g \neq \boldsymbol{0}\right\},$$

which satisfies $\int_{\mathbf{M}} \min\{1, g(t)\} \mathrm{d}\nu(g) < \infty$ for all $t \geq 0$ and $b \in \mathbf{M}$ is a unique deterministic (drift)function. From (15) and the formula for the Laplace transform of a PRM, see e.g. [31, Section 3], one can deduce that

$$H \sim b + \hat{H}$$

may be decomposed into a deterministic drift b and a "completely random" process $\left(\hat{H}_t\right)_{t \geq 0} \sim \left(\int_{\mathbf{M}} g(t) \mathrm{d}N_H(g)\right)_{t \geq 0}$, where $N_H := \sum_{i \in \mathbb{N}} \delta_{g_i}$ is a PRM on $\mathbf{M}$ with intensity measure ν. Combining (14) and (15) we obtain that

$$\mathbb{P}(\boldsymbol{Y} > \boldsymbol{x}) = \mathbb{E}\left[\exp\left(-\sum_{i=1}^{\infty} H_{x_i}\right)\right] = \exp\left(-\sum_{i=1}^{\infty} b(x_i) - \int_{\mathbf{M}} 1 - \exp\left(-\sum_{i=1}^{\infty} g(x_i)\right) \mathrm{d}\nu(g)\right), \quad \boldsymbol{x} \in [0, \infty)^{\mathbb{N}}.$$

This shows that the exponent measure $\bar{\mu}$ of the exchangeable min-id sequence $\boldsymbol{Y}$ is given by

$$\bar{\mu}(A) = \bar{\mu}_b(A) + \int_{\mathbf{M}} \otimes_{i=1}^{\infty} \left(1 - \exp\left(-g(\cdot)\right)\right)(A) \mathrm{d}\nu(g),$$

where

- $\bar{\mu}_b$ denotes the exponent measure of an iid sequence with stochastic representation $\left(\inf\left\{t \geq 0 \mid b(t) \geq E_i^{(1)}\right\}\right)_{i \in \mathbb{N}}$ and marginal distribution function $1 - \exp(-b(\cdot))$,
- $\otimes_{i=1}^{\infty}\left(1 - \exp\left(-g(\cdot)\right)\right)$ denotes the distribution of an iid sequence with marginal distribution function $1 - \exp(-g(\cdot))$ and $\int_{\mathbf{M}} \otimes_{i=1}^{\infty}\left(1 - \exp\left(-g(\cdot)\right)\right)(A) \mathrm{d}\nu(g)$ denotes the exponent measure of an exchangeable max-id sequence with stochastic representation $\left(\inf\left\{t \geq 0 \mid \hat{H}(t) \geq E_i^{(2)}\right\}\right)_{i \in \mathbb{N}}$

and $\left(E_i^{(j)}\right)_{i \in \mathbb{N}}, j \in \{1, 2\}$ denote independent sequences of iid unit Exponential distributed random variables. In other words, $\bar{\mu}$ is the sum of the exponent measure of an iid sequence and a (possibly infinite) mixture of iid sequences. Since addition of two exponent measures stochastically corresponds to applying component-wise minima to two independent min-id sequences, we get

$$\boldsymbol{Y} \sim \min\left\{\left(\inf\left\{t \geq 0 \mid b(t) \geq E_i^{(1)}\right\}\right)_{i \in \mathbb{N}}; \left(\inf\left\{t \geq 0 \mid \hat{H}(t) \geq E_i^{(2)}\right\}\right)_{i \in \mathbb{N}}\right\},$$

which shows that the only difficulty in the simulation of $\boldsymbol{Y}$ is the simulation of the sequence $\left(\inf\left\{t \geq 0 \mid \hat{H}(t) \geq E_i^{(2)}\right\}\right)_{i \in \mathbb{N}}$. Hence, for our analysis, we can ignore the presence of $\bar{\mu}_b$, i.e. assume that $b = 0$, and focus on the simulation of the min-id sequence $\left(\inf\left\{t \geq 0 \mid \hat{H}(t) \geq E_i^{(2)}\right\}\right)_{i \in \mathbb{N}}$ with exponent measure of the form

$$\bar{\mu}(A) = \int_{\mathbf{M}} \otimes_{i=1}^{\infty} \left(1 - \exp(-g(\cdot))\right)(A) \mathrm{d}\nu(g). \tag{16}$$

Representation (16) implies that the atoms of the PRM $\bar{N}$ with intensity $\bar{\mu}$ can be generated as follows:

(i) Generate a PRM $N_H = \sum_{i \in \mathbb{N}} \delta_{g_i}$ on **M** with intensity measure ν;
(ii) For each g_i, draw an iid sequence $\bar{\boldsymbol{f}}_i$ with distribution function $1 - \exp(-g_i(\cdot))$;
(iii) Set $\bar{N} = \sum_{i \in \mathbb{N}} \delta_{\bar{\boldsymbol{f}}_i}$.

As mentioned previously, a PRM N with intensity μ is then obtained by taking the reciprocal of each atom of $\bar{N}$, i.e. by defining $N := \sum_{i \in \mathbb{N}} \delta_{1/\bar{\boldsymbol{f}}_i}$. Therefore, a PRM with intensity $\mathbf{1}_{\{f(n) \geq c\}} d\mu$ can be generated by the simulation of a finite PRM with intensity

$$\bar{\mu}\left(\cdot \cap \{f(n) \leq 1/c\}\right) = \int_{\mathbf{M}} \otimes_{i=1}^{\infty} \left(1 - \exp\left(-g(\cdot)\right)\right)\left(\cdot \cap \{f(n) \leq 1/c\}\right) d\nu(g)$$

$$= \int \otimes_{i=1}^{\infty} \left(1 - \exp\left(-g(\cdot)\right)\right)\left(\{\boldsymbol{f} \in \ \cdot \ \} \mid f(n) \leq 1/c\right)\left(1 - \exp\left(-g\left(1/c\right)\right)\right) d\nu(g). \tag{17}$$

It is important to observe that $(1 - \exp(-g(1/c)))\, d\nu(g)$ defines an exponent measure with total finite mass

$$C_c := -\log\left(\mathbb{E}\left[\exp\left(-H_{1/c}\right)\right]\right) = -\log\left(\mathbb{P}\left(Y_1 > 1/c\right)\right) = -\log\left(\mathbb{P}\left(X_1 < c\right)\right). \tag{18}$$

Thus, to simulate from a PRM with intensity $\mathbf{1}_{\{f(n) \geq c\}} d\mu$, it is sufficient to be able to simulate from the probability measure $P_c := C_c^{-1}\left(1 - \exp(-g(1/c))\right) d\nu(g)$ on **M**. Since the measure ν can be chosen rather arbitrarily, it is hopeless to expect a general recipe for the simulation of P_c. However, there are many families of stochastic processes for which ν can be conveniently described such that simulation from P_c becomes feasible. One of these families is the class of self-similar additive processes [17], [33, Section 3], which will be investigated in the next section.

5. Exchangeable Sato-frailty sequences

Choosing $(H_t)_{t \geq 0}$ in (14) as a non-negative and non-decreasing additive process, also called additive subordinator, gives rise to the class of so-called exchangeable exogenous shock models [20,34]. Exchangeable exogenous shock models are characterized by the property that every d-dimensional margin $\boldsymbol{Y}_d$ of $\boldsymbol{Y}$ can be stochastically represented as the minimum of independent random shocks, each of them affecting a certain subset of components of $\boldsymbol{Y}_d$. More formally, every d-dimensional margin of the exchangeable exogenous shock model $\boldsymbol{Y}$ can be represented as

$$\boldsymbol{Y}_d \sim \left(\min\{E_I \mid I \subset \{1, \ldots, d\}, \ i \in I\}\right)_{1 \leq i \leq d}, \tag{19}$$

where the shocks $(E_I)_{I \subset \{1, \ldots, d\}}$ denote independent non-negative random variables with continuous distribution function and $E_{I_1} \sim E_{I_2}$ if $|I_1| = |I_2|$. Moreover, the distribution of the E_I is uniquely linked to the Laplace transform of the associated additive subordinator H. In principle, the results of [20,34] could be used to simulate the d-dimensional margins of an exchangeable exogenous shock model. However, even if the Laplace transform of the associated additive subordinator is known analytically, it is numerically challenging to compute the distribution of the individual shocks E_I and 2^d random variables have to be simulated to determine $\boldsymbol{Y}_d$. Thus, if d is large, it is practically infeasible to simulate an exchangeable exogenous shock model via the representation (19). Alternatively, one could use the representation (14) to generate a sample of $\boldsymbol{X}$, which circumvents the curse of dimensionality. Unfortunately, the simulation of the additive subordinator H is usually infeasible or only possible approximatively. Therefore, an exact and efficient simulation of high dimensional exchangeable exogenous shock models has remained an open problem to date.

A subclass of exchangeable exogenous shock models has been investigated in [21] by restricting $(H_t)_{t \geq 0}$ to the class of self-similar subordinators (aka Sato subordinators), meaning that H is an additive subordinator and that there exists some index $\gamma > 0$ such that for all $a \geq 0$ $(H_{at})_{t \geq 0} \sim (a^{\gamma} H_t)_{t \geq 0}$. The exchangeable sequences associated to self-similar subordinators are called exchangeable Sato-frailty sequences. [33, Section 3] shows that every self-similar additive process H with index γ is uniquely associated to its distribution at unit time. The law of H_1 belongs to the class of self-decomposable distributions, meaning that for every $c \in (0, 1)$ there exists a random variable $H^{(c)}$ independent of H_1 such that $H_1 \sim cH_1 + H^{(c)}$. Self-decomposable laws constitute a broad subclass of infinitely divisible distributions, e.g. containing the (inverse-)Gaussian, Laplace, (tempered-)stable, Fréchet, Pareto, Exponential and (inverse-)Gamma distribution as well as several laws appearing in financial modeling as the CGMY, Normal Inverse Gaussian and Meixner distribution [3,8,36] to provide some examples. Furthermore, every self-decomposable distribution can be obtained as the law of a unique self-similar additive process with index γ at unit time. Thus, there is a one-to-one correspondence of self-similar additive subordinators with index γ, non-negative self-decomposable distributions and the class of exchangeable Sato-frailty sequences.

Moreover, [33, Proposition 16.5] shows that the index γ of a self-similar process can be changed to an arbitrary index $\tilde{\gamma}$ via the simple time change $t \mapsto t^{\tilde{\gamma}/\gamma}$. Combined with [7, Corollary 3.2], which shows that a time-change of the self-similar subordinator corresponds to the marginal transformation $a \mapsto a^{\gamma/\tilde{\gamma}}$ of the associated exchangeable Sato-frailty sequence, we can w.l.o.g. assume that $\gamma = 1$ to simplify further derivations.

The key quantity of our simulation algorithm will be the univariate Lévy measure υ of the self-decomposable distribution of H_1. It allows us to derive a convenient representation of the path Lévy measure of the associated self-similar subordinator, which then translates into a simple representation of the exponent measure of the associated exchangeable Sato-frailty sequence via (16). To this purpose, we recall several characterizations of self-decomposable laws, which are provided in [33, Section 3]. First, [33, Theorem 15.10] shows that the Lévy measure υ of a non-negative self-decomposable distribution is absolutely continuous w.r.t. the Lebesgue measure with density of the form $\upsilon(\mathrm{d}a) = k(a)a^{-1}\mathbf{1}_{\{a>0\}}\mathrm{d}a$, where k denotes some non-increasing right-continuous function such that $\int_0^\infty \min\{a, 1\}k(a)a^{-1}\mathrm{d}a < \infty$.

Noting that k defines a measure ρ_k on $(0, \infty)$ by $\rho_k((a, \infty)) := k(a)$, $a > 0$, we can rewrite υ as $\upsilon(\mathrm{d}a) = \rho_k((a, \infty))a^{-1}\mathrm{d}a$. It turns out that ρ_k defines the Lévy measure of another non-negative id distribution [33, Theorem 17.5]. Thus, there exists a non-negative and non-decreasing Lévy process, also called Lévy subordinator, $\left(L_t^{(k)}\right)_{t\geq 0}$ with univariate Lévy measure ρ_k and, according to [33, Equation (17.2)], the self-decomposable distribution with Lévy measure υ can be recovered from $L^{(k)}$ as the distribution of the infinitely divisible random variable $\int_0^\infty \exp(-s)\mathrm{d}L_s^{(k)}$. Due to this representation $L^{(k)}$ is called the Background Driving Lévy process (BDLP) of the self-decomposable distribution with Lévy measure υ. [17] shows that not only the self-decomposable distribution associated with Lévy measure υ, but also the associated self-similar subordinator H can be recovered from (the law of) $L^{(k)}$ by

$$(H_t)_{t\geq 0} := \left(\int_{-\log(\min\{t;1\})}^\infty \exp(-s)\mathrm{d}L_s^{(k,1)} + \int_0^{\log(\max\{1,t\})} \exp(s)\mathrm{d}L_s^{(k,2)}\right)_{t\geq 0}, \tag{20}$$

where $L_{i=1,2}^{(k,i)}$ denote two iid copies of $L^{(k)}$. This particular representation of the self-similar subordinator H allows us to derive a representation of its associated path Lévy measure in terms of the Lévy measure of the BDLP.

Lemma 2 (*Lévy Measure of Self-similar Subordinator Via Lévy Measure of BDLP*)**.** *Let $(H_t)_{t\geq 0}$ denote a self-similar subordinator with index 1 and let $L^{(k)}$ denote the BDLP associated to H_1. The Lévy measure of H can be expressed as*

$$\nu(A) = \int_{(0,\infty)^2} \mathbf{1}_{\left\{as\mathbf{1}_{\{\cdot\geq s\}}\in A\right\}} s^{-1}\mathrm{d}s \otimes \rho_k(\mathrm{d}a); \quad A \in \mathcal{B}(\mathbf{M}), \tag{21}$$

where ρ_k denotes the Lévy measure of $L^{(k)}$.

Assuming that $k(\cdot) = \rho_k((\cdot, \infty))$ is differentiable we obtain the following corollary.

Corollary 1 (*Lévy Measure of Self-similar Subordinator Via Density*)**.** *Let $(H_t)_{t\geq 0}$ denote a self-similar subordinator with index 1 associated to the self decomposable distribution with Lévy measure $\mathrm{d}\upsilon = k(a)a^{-1}\mathrm{d}a$. If k is differentiable, then the path Lévy measure ν of $(H_t)_{t\geq 0}$ is given by*

$$\nu(A) = -\int_{(0,\infty)^2} \mathbf{1}_{\left\{a\mathbf{1}_{\{\cdot\geq s\}}\in A\right\}} k'\left(as^{-1}\right)s^{-2}\mathrm{d}s\mathrm{d}a; \quad A \in \mathcal{B}(\mathbf{M}).$$

Having determined the Lévy measure of a self-similar subordinator we can express the exponent measure of the associated exchangeable Sato-frailty sequence by (16) as

$$\bar{\mu}(A) = -\int_0^\infty \int_0^\infty \otimes_{i=1}^\infty \left(1 - \exp\left(-a\right)\mathbf{1}_{\{\cdot\geq s\}}\right)\left(\left\{(f(i))_{i\in\mathbb{N}}\in A\right\}\right) k'\left(as^{-1}\right)s^{-2}\mathrm{d}a\mathrm{d}s,$$

assuming that $k(\cdot) = \rho_k((\cdot, \infty))$ is differentiable. Thus, to apply Algorithm 1, we need to simulate a PRM with intensity (17) expressible as

$$-\int_0^\infty \int_0^{1/c} \otimes_{i=1}^\infty \left(1 - \exp\left(-a\mathbf{1}_{\{\cdot\geq s\}}\right)\right)\left(\left\{f\in\ \cdot\ \right\}\ \middle|\ f(n)\leq\frac{1}{c}\right) k'\left(as^{-1}\right)s^{-2}\left(1 - \exp(-a)\right)\mathrm{d}s\mathrm{d}a.$$

The only difficulty in the simulation of this PRM is the simulation of the random vector $(A^{(c)}, S^{(c)})$ with joint distribution $-\mathbf{1}_{\{s\in(0,1/c)\}}\mathbf{1}_{\{0<a\}}C_c^{-1}k'\left(as^{-1}\right)s^{-2}\left(1 - \exp(-a)\right)\mathrm{d}a\mathrm{d}s$, where the normalizing constant C_c was defined in (18). However, it is easy to see that the marginal density of $A^{(c)}$ is given by

$$g_{A^{(c)}}(a) = \mathbf{1}_{\{a>0\}}C_c^{-1}a^{-1}k(ac)\left(1 - \exp(-a)\right) \tag{22}$$

and that the conditional density of $S^{(c)}$ given $A^{(c)}$ is given by

$$g_{S^{(c)}|A^{(c)}}(s) = -\mathbf{1}_{\{s\in(0,1/c)\}}k'\left(A^{(c)}s^{-1}\right)s^{-2}\left(\left(A^{(c)}\right)^{-1}k\left(A^{(c)}c\right)\right)^{-1}. \tag{23}$$

Thus, a sample of $(A^{(c)}, S^{(c)})$ can be generated by first sampling a random variable $A^{(c)}$ with density $g_{A^{(c)}}$ and then, given $A^{(c)}$, sampling a random variable $S^{(c)}$ according to the conditional density $g_{S^{(c)}|A^{(c)}}$. Altogether, this implies that we can sample from a PRM with intensity (17) by the following procedure:

(i) Draw a random variable $\Theta \sim Poi(C_c)$;

(ii) For $1 \le i \le \Theta$ draw independent random variables $A_i^{(c)}$ according to the density $g_{A^{(c)}}$ and, conditioned on $A_i^{(c)}$, draw random variables $S_i^{(c)}$ according to the density $g_{S^{(c)}|A_i^{(c)}}$;

(iii) For each pair $(A_i^{(c)}, S_i^{(c)})$ draw an iid sequence $\boldsymbol{f}^{\left(A_i^{(c)}, S_i^{(c)}\right)} := \left(f_j^{\left(A_i^{(c)}, S_i^{(c)}\right)}\right)_{j \in \mathbb{N}}$ with marginal distribution $\left(1 - \exp\left(-A_i^{(c)}\right)\right)\delta_{S^{(c)}} + \exp\left(-A_i^{(c)}\right)\delta_\infty$ and set $f_n^{\left(A_i^{(c)}, S_i^{(c)}\right)} = S_i^{(c)}$;

(iv) The PRM with intensity (17) is given by $\sum_{i=1}^{\Theta} \delta_{1/\boldsymbol{f}^{\left(A_i^{(c)}, S_i^{(c)}\right)}}$.

Therefore, Algorithm 1 can be employed to generate exchangeable Sato-frailty sequences if the associated function k is differentiable. When k is not differentiable, one can still obtain a representation of $\bar{\mu}$ in terms of ρ_k via Lemma 2. However, the simulation procedure of a PRM with intensity (17) slightly changes and requires the simulation of a random vector with density $\mathbf{1}_{\{s \in (0,1/c)\}}\mathbf{1}_{\{0<a\}}C_c^{-1}s^{-1}(1 - \exp(-as))\mathrm{d}s\mathrm{d}\rho_k(a)$, which cannot be conducted without assuming further regularity properties of ρ_k.

Remark 6 (*Sampling of the Densities $g_{A^{(c)}}$ and $g_{S^{(c)}|A^{(c)}}$*). If the function k is known analytically, one can use rejection sampling to obtain (exact) samples from random variables with density $g_{A^{(c)}}$, see e.g. [24, p. 235 ff.] for more details on rejection sampling. To simulate a random variable with density $g_{S^{(c)}|A^{(c)}}$ one could also use rejection sampling if k' is known analytically, but one should notice that its associated distribution function is given by $G_{S^{(c)}|A^{(c)}}(s) = k(A^{(c)}s^{-1})A^{(c)}k\left(A^{(c)}c\right)^{-1}\mathbf{1}_{\{0<s<1/c\}}$. Therefore, rejection sampling and the (numerical) inverse transform sampling method may be used to sample random variables with density $g_{A^{(c)}}$ and conditional density $g_{S^{(c)}|A^{(c)}}$.

Remark 7 (*Extension to Exchangeable Exogenous Shock Models*). [32] shows that an id process H is additive if and only if its path Lévy measure is concentrated on one-time jump functions of the form $a\mathbf{1}_{\{\cdot \ge s\}}$. In many cases, the univariate Lévy measure υ_t of the real-valued random variable H_t is absolutely continuous w.r.t. to the Lebesgue measure, meaning that $\upsilon_t(\mathrm{d}a) = k(a,t)\mathrm{d}a$. If $k(a,\cdot)$ is differentiable on $(0,\infty)$ for almost all a one can obtain a similar expression of the path Lévy measure of an additive process as in Corollary 1. It can be easily checked that the image measure of the map $\left((0,\infty) \times (-\infty,\infty), k'(a,s)\mathrm{d}a\mathrm{d}s\right) \to \mathbf{M}; \ (s,a) \mapsto a\mathbf{1}_{\{\cdot \ge s\}}$ satisfies the conditions of [32, Theorem 2.8] and thus defines a valid Lévy measure of a driftless additive process $\tilde{H}$. We obtain that

$$\mathbb{E}\left[\exp\left(iz\left(b(t) + \tilde{H}_t\right)\right)\right] = \exp\left(izb(t) + \int_{\mathbb{R}\setminus\{0\}}\int_0^\infty \left(\exp\left(iza\mathbf{1}_{\{t\ge s\}}\right) - 1 - iza\mathbf{1}_{\{t\ge s\}}\mathbf{1}_{\{a\le 1\}}\right)k'(a,s)\mathrm{d}s\mathrm{d}a\right)$$

$$= \exp\left(izb(t) + \int_{\mathbb{R}\setminus\{0\}}\int_0^t \left(\exp\left(iza\right) - 1 - iza\mathbf{1}_{\{a\le 1\}}\right)k'(a,s)\mathrm{d}s\mathrm{d}a\right)$$

$$= \exp\left(izb(t) + \int_{\mathbb{R}\setminus\{0\}} \left(\exp\left(iza\right) - 1 - iza\mathbf{1}_{\{a\le 1\}}\right)k(a,t)\mathrm{d}a\right) = \mathbb{E}\left[\exp\left(izH_t\right)\right],$$

since $\lim_{s\to 0}k(a,s) = 0$ for almost all $a \ne 0$ by the stochastic continuity of additive processes. Thus, $b + \tilde{H}$ and H are identical in distribution, given that additive processes are uniquely determined by their marginal distributions. Therefore, the path Lévy measure of H is given by

$$\nu(A) = \int_{\mathbb{R}\setminus\{0\}}\int_0^\infty \mathbf{1}_{\left\{a\mathbf{1}_{\{\cdot \ge s\}} \in A\right\}}k'(a,s)\mathrm{d}a\mathrm{d}s; \ A \in \mathcal{B}\left(D\left([0,\infty)\right)\right),$$

where $D\left([0,\infty)\right)$ denotes space of real-valued càdlàg functions. Similar to Sato-frailty sequences, this allows to sample the exchangeable sequence associated to an additive subordinator with univariate Lévy measure $\upsilon_t(\mathrm{d}a) = \mathbf{1}_{\{a>0\}}k(a,t)\mathrm{d}a$ by repeatedly drawing random vectors $(A^{(c)}, S^{(c)})$ and conditionally iid sequences.

Remark 8 (*Extension to Exchangeable Max-stable Sequences*). Stochastic processes H which satisfy $(H_{nt})_{t\ge 0} \sim \left(\sum_{i=1}^n H_t^{(i)}\right)_{t\ge 0}$ for all $n \in N$ and iid copies $\left(H^{(i)}\right)_{i\in\mathbb{N}}$ of H are called strongly infinitely divisible w.r.t. time (strong-idt). [23] has shown that the max-id sequence $\boldsymbol{X}$ corresponding to a strong-idt process in (14) is max-stable, meaning that its marginal distributions can be obtained as a limit distribution of scaled maxima of iid random vectors. The general form of the exponent measure of an exchangeable max-stable sequence $\boldsymbol{X}$ has been derived in [19]. However, it still involves the law of a stochastic process and does not directly translate into a simple simulation procedure for Algorithm 1. [4,25] have investigated subfamilies of strong-idt processes with the particular representation $\left(H_f(t)\right)_{t\ge 0} = \left(\int_0^\infty f(s/t)\mathrm{d}L_s\right)_{t\ge 0}$, where $(L_t)_{t\ge 0}$ denotes a Lévy subordinator and f denotes a non-negative, non-increasing left-continuous function. [25] provides exact simulation algorithms for the d-dimensional margins of the corresponding exchangeable max-stable sequence in the particular case $f(s) = \lim_{u\nearrow s} -\log(F(u))$ for some fixed distribution function F, whereas the models in [4] could only

be simulated when L is a compound Poisson process. Lemma 3 from Appendix B yields a rather simple representation of the exponent measure of the exchangeable max-stable sequence associated to H_f in terms of the path Lévy measure of L, which can be translated into a representation of the exponent measure of X as a mixture of iid sequences in terms of the law of a random vector $(A^{(c)}, S^{(c)})$. Thus, similar to Sato-frailty sequences, the examples from [4,25] can essentially be simulated by repeated simulations of a random vector $(A^{(c)}, S^{(c)})$ and conditionally iid sequences.

6. Illustration of the simulation algorithm

In this section we exemplarily demonstrate how Algorithm 1 can be used to simulate an exchangeable Sato-frailty sequence in practice. We rather aim at providing a proof-of-concept like exposition than to fine-tune the presented example to its most efficient simulation procedure. We chose the Inverse Gaussian (IG) distribution as our guiding example. The IG distribution is known to be self-decomposable [15] and its Lévy measure is given by

$$\upsilon(\mathrm{d}a) = \mathbf{1}_{\{a>0\}} \frac{\delta}{\sqrt{2\pi}} a^{-\frac{3}{2}} \exp\left(-\frac{\gamma^2 a}{2}\right) \mathrm{d}a, \text{ where } \delta, \gamma > 0.$$

Therefore, the Lévy measure of the associated self-similar subordinator is characterized by the function

$$k(a) = \frac{\delta}{\sqrt{2\pi}} a^{-1/2} \exp\left(-\frac{\gamma^2 a}{2}\right), \ a > 0.$$

One should note that a simulation of the associated exchangeable Sato-frailty sequence X via its stochastic representation (14) would either require the simulation of the whole path of the infinitely active BDLPs in (20) or the direct simulation of the increments of the associated self-similar subordinator. However, the simulation of the whole path of the BDLPs cannot be practically achieved nor can the increments of the associated self-similar subordinator be efficiently simulated, since their law cannot be easily characterized. Thus, Algorithm 1 can be seen as the natural choice regarding the simulation of X. It is quite easy to see that the associated random vector $(A^{(c)}, S^{(c)})$ can be simulated by rejection sampling for the random variable $A^{(c)}$ with density (22) and inverse transform sampling for the random variable $S^{(c)}|A^{(c)}$ with conditional density (23). We have simulated the corresponding sequence X for various values of (δ, γ) and report our results in terms of scatterplots of the associated copula $C(u_1, \ldots, u_d) = \mathbb{P}(F_1(X_1) \le u_1, \ldots, F_d(X_d) \le u_d)$, since copulas do not depend on the marginal distribution of X. In particular, the associated copula is independent of the index of self-similarity of the associated self-similar subordinator. The copula corresponding to X has been analytically derived in [21] and is given by

$$C(u_1, \ldots, u_d) = \prod_{i=1}^{d} \exp\left(\delta\gamma\left(\sqrt{1 + i\left(\frac{\log(u_{[i]})}{\delta\gamma} + 1\right)^2} - i - \sqrt{1 + (i-1)\left(\frac{\log(u_{[i]})}{\delta\gamma} + 1\right)^2} - (i-1)\right)\right),$$

where $u_{[i]}$ is defined as the ith order statistic of $(u_1, \ldots, u_d)$. Thus, the associated copula only depends on $\delta\gamma$. Fig. 2 provides the empirical copula plots for dimensions $d \in \{2, 3\}$ and $\delta\gamma \in \{1/10, 2, 10\}$ and shows that the margins of X become less dependent with increasing $\delta\gamma$.

To empirically verify the complexity characterization of Algorithm 1 given in Theorem 3, Fig. 3 shows the average and standard deviation of the number of simulated sequences to produce one sample of the exchangeable Sato-frailty sequences associated to the IG distributions with $\delta\gamma \in \{1/10, 2, 10\}$ and various dimensions $d \in \{1, 5, 10, 25, 50, 100, 250, 500, 1000, 2500, 5000, 10000\}$ over 500 repetitions. Note that we applied Algorithm 1 in accordance with assumption (13) as follows: After simulating the random vectors $\left(\left(A_i^{(c)}, S_i^{(c)}\right)\right)_{1 \le i \le \Theta}$ where $\Theta \sim Poi(C_c)$ via the (conditional) densities (22) and (23), we can simulate the sequences associated to the $\left(\left(A_i^{(c)}, S_i^{(c)}\right)\right)_{1 \le i \le \Theta}$ in increasing order of the $S_i^{(c)}$. It is easy to see that this procedure allows to simulate a PRM with intensity μ according to assumption (13). Therefore, the expected number of simulated sequences is equal to d, since the one-dimensional marginal distributions of exchangeable Sato-frailty sequences are continuous. Fig. 3 empirically verifies this result, showing that the average number of simulated sequences over 500 repetitions is always close to d, independently of $\delta\gamma$ and d. Interestingly, the standard deviation of the number of simulated sequences seems to depend on $\delta\gamma$. Thus, the example shows that, even though the expected number of simulated sequences (or functions) in Algorithm 1 is always equal to d when the margins of X follow a continuous distribution, its standard deviation may depend on properties of the associated continuous max-id process.

7. Discussion

We have developed an algorithm for the exact simulation of continuous max-id processes, based on their representation as a pointwise maximum over functions in a PRM. Our algorithm is solely based on the ability to simulate PRMs with finite intensity measures, which facilitates its wide applicability. The complexity of the algorithm has been characterized in terms of the expected number functions that need to be simulated to obtain a sample of the continuous max-id process.

Exemplarily, we have derived the exponent measure of an exchangeable Sato-frailty sequence and demonstrated the applicability of our algorithm theoretically and in practice, thereby providing the first exact simulation algorithm
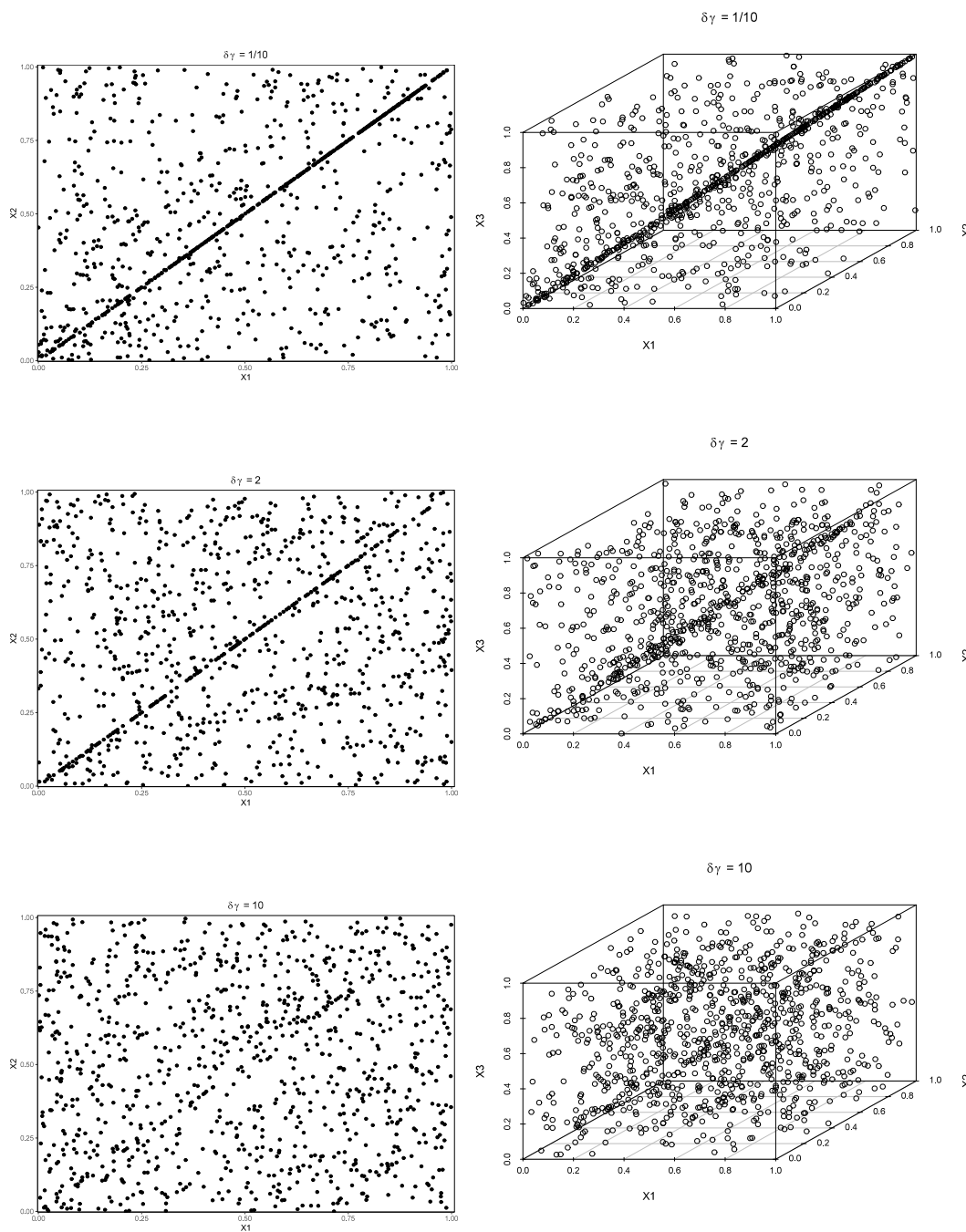
Fig. 2. Empirical copula plots of the two- and three-dimensional margins of 1000 samples of the exchangeable Sato-frailty sequences associated to the IG distribution with parameters $\delta\gamma \in \{1/10, 2, 10\}$ (top, middle, bottom).

for high dimensional samples of this family. We have sketched how the simulation algorithm for exchangeable Sato-frailty sequences can be generalized to certain families of exogenous shock models and max-stable sequences without increasing its practical complexity. This enables the possibility to consider the construction principle of exchangeable min-id sequences in (4) by means of its desired analytical properties, without the need of having a suitable simulation algorithm for the associated id process at hand. Appendix A discusses an alternative simulation algorithm for max-id random vectors, thereby accounting for the natural geometric descriptions of many known families of finite dimensional exponent measures. An application of our algorithm to a continuous max-id process with uncountable index set is left for future research. We think that the proposed simulation algorithm may be extended to upper semicontinuous max-id processes with obvious modifications, however the technical details need to be carefully worked out and are also left for future research.
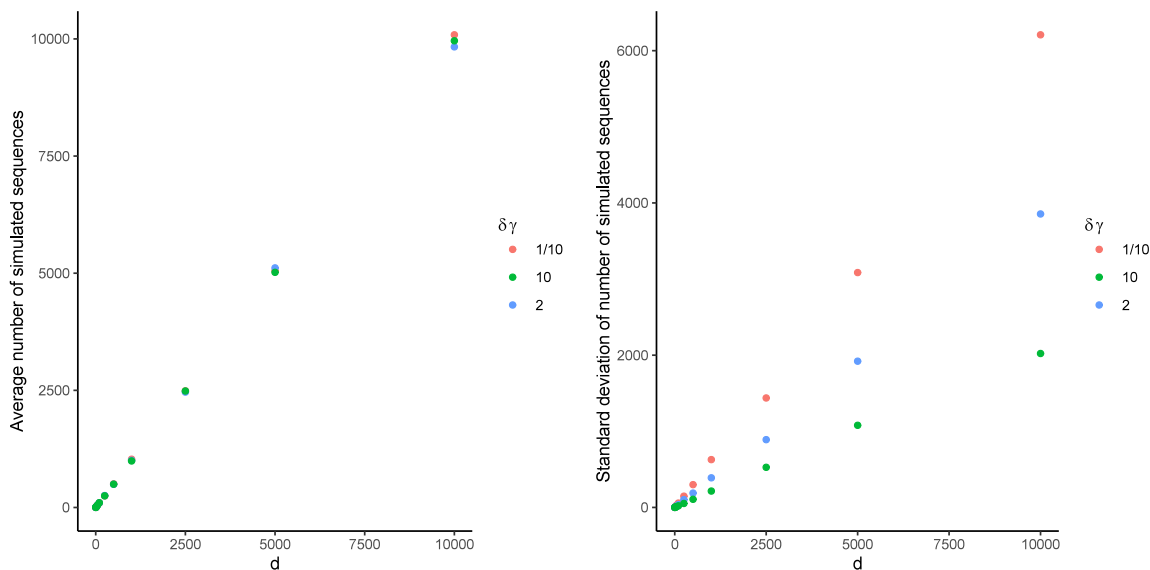
Fig. 3. The average (left) and standard deviation (right) of the number of simulated sequences to obtain one sample of the exchangeable Sato-frailty sequences associated to the IG distribution for various values of $\delta\gamma$ and dimensions d over 500 repetitions.

CRediT authorship contribution statement

Acknowledgments

Appendix A. Exact simulation of max-id random vectors

This section is devoted to the exact simulation of a max-id random vector $X \in [0, \infty)^d$. Since X can be viewed as a continuous max-id process on $T = \{1, \ldots, d\}$ Algorithm 1 is, in principle, applicable to every max-id random vector. However, the exponent measure of a max-id random vector is often more easily described by exploiting the specific geometric structure of $[0, \infty)^d$. For example, a common representation of an exponent measure of a max-id random vector is the scale mixture of a probability distribution on the non-negative unit sphere of some norm on $\mathbb{R}^d$. Two famous representatives of this class of exponent measures are the exponent measures of max-stable random vectors with unit Fréchet margins [31, Chapter 5] and random vectors with reciprocal Archimedean copula [13,18], see Example 2. In both cases, a simulation of X via Algorithm 1 would require to deviate from the natural description of the exponent measure to simulate a PRM with intensity $\mathbf{1}_{\{f(n) \geq c\}} d\mu(f)$. Thus, there is a need to adapt Algorithm 1 to exploit the natural structure of many exponent measures of max-id random vectors. Again, to simplify the theoretical developments, we can w.l.o.g. assume that $h_X = \mathbf{0}$.

Our goal is to generalize the algorithms of [10,18,35] to max-id random vectors. Similar to Algorithm 1 we will only simulate those atoms of the PRM $N = \sum_{i \in \mathbb{N}} \delta_{x_i}$ with intensity μ which may be relevant to determine $\max_{i \in \mathbb{N}} x_i = X$. We start by dividing $[0, \infty)^d$ into disjoint "slices" S_n of finite μ-measure. Then, assuming that we can simulate finite PRMs N_n with intensities $\mu (\cdot \cap S_n)$, we iteratively simulate the N_n until a stopping criterion is reached. To obtain a valid stopping criterion we need to assume that the slices S_n eventually approach $\mathbf{0}$, which is mathematically described as eventually residing in an open ball around $\mathbf{0}$. This will force the algorithm to stop after finitely many steps, since atoms of the PRM N in a neighborhood of $\mathbf{0}$ eventually cannot contribute to the maximum of the already simulated points.

Example 1. Assume that the atoms of the PRM N are given by the points in Fig. A.4. A possible execution of our algorithm could be described as follows: In the first step, all atoms of the PRM above the blue line are simulated, which corresponds
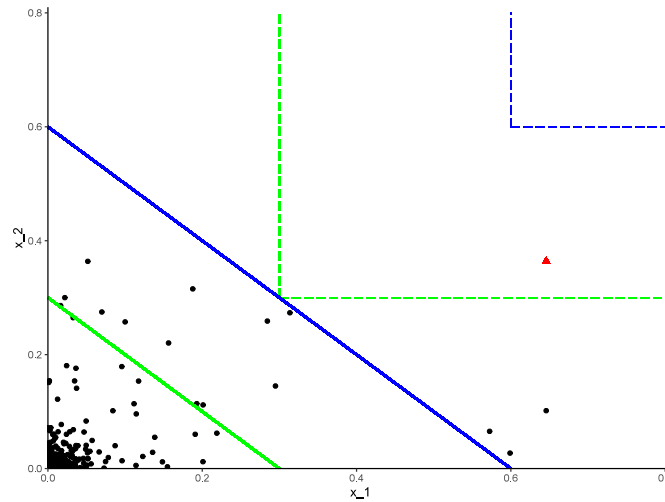
Fig. A.4. Illustration of Algorithm 2.

to $\mathcal{S}_1 = \{\boldsymbol{x} \mid x_1 + x_2 \geq 0.6, x_1, x_2 \geq 0\}$. Since the pointwise maximum of atoms of N above the solid-blue line is not above the dashed-blue line $\{\boldsymbol{x} \mid \min_{i=1,2} x_i = 0.6\}$ there are possibly some atoms of N which can contribute to the pointwise maximum of N and which have not yet been simulated. Therefore, in a second step, we simulate all points between the solid-blue and the solid-green line, which corresponds to $\mathcal{S}_2 = \{\boldsymbol{x} \mid 0.3 \leq x_1 + x_2 < 0.6, x_1, x_2 \geq 0\}$. The red triangle denotes the pointwise maximum of the simulated points above the solid-green line. Since it is above the dashed-green line $\{\boldsymbol{x} \mid \min_{i=1,2} x_i = 0.3\}$ it is the maximum of the PRM N and the algorithms stops.

Let us formalize the proposed algorithm. Let $B_r^\infty(\boldsymbol{x}) := \{\boldsymbol{y} \in \mathbb{R}^d \mid \max_{1 \leq i \leq d} |y_i - x_i| < r\}$ denote the open ball of radius r around $\boldsymbol{x}$ w.r.t. the supremum norm. We assume that we can simulate from finite PRMs N_n with intensities $\mu(\cdot \cap \mathcal{S}_n)$, where $(\mathcal{S}_n)_{n \in \mathbb{N}}$ is a sequence of disjoint sets which satisfy

  (i) $\mu(\mathcal{S}_n) < \infty$ for all $n \in \mathbb{N}$,
  (ii) $\bigcup_{n=1}^\infty \mathcal{S}_n = [0, \infty)^d \setminus \{\boldsymbol{0}\}$,
  (iii) for all $r > 0$ there exists $m(r) \in \mathbb{N}$ such that $\bigcup_{n \geq m} \mathcal{S}_n \subset B_r^\infty(\boldsymbol{0})$.

Under these conditions on $\mathcal{S}_n$ we can propose the following algorithm for the exact simulation of max-id random vectors with exponent measure μ.

Algorithm 2: Exact simulation of max-id random vector with vertex $\boldsymbol{0}$

    Result: Unbiased sample of $(X_1, \ldots, X_d)$ with exponent measure μ.

1 Set $\mu_j(\cdot) = \mu\left(\cdot \cap \{\boldsymbol{x} \in [0, \infty)^d \setminus \{\boldsymbol{0}\} \mid x_j > 0, \ x_k = 0, \ k < j, k \in J_0\}\right)$, $j \in J_0$;

2 Set $\tilde{\mu}(\cdot) = \mu\left(\cdot \cap \{\boldsymbol{x} \in [0, \infty)^d \setminus \{\boldsymbol{0}\} \mid x_j = 0, j \in J_0\}\right)$;

3 for $j \in J_0$ **do**

4   Simulate a finite PRM N_j with intensity μ_j and set $\hat{\boldsymbol{X}}_j = \max_{\boldsymbol{x} \in N_j} \boldsymbol{x}$;

5 end

6 Set $\tilde{X} = \boldsymbol{0}$ and $n = 1$;

7 while *there is no* $r > 0$ *such that* $\bigcup_{m \geq n} \mathcal{S}_m \subset B_r^\infty(\boldsymbol{0})$ ***and*** $\min_{i \notin J_0} \tilde{X}_i \geq r$ **do**

8   Simulate the finite PRM $\tilde{N}_n$ with intensity $\tilde{\mu}(\cdot \cap \mathcal{S}_n)$;

9   Set $\tilde{X} = \max\left\{\max_{\boldsymbol{x}^{(n)} \in \tilde{N}_n} \boldsymbol{x}^{(n)}, \tilde{X}\right\}$;

10   Set $n = n + 1$;

11 end

12 Set $\hat{\boldsymbol{X}} = \max\left\{\max_{j \in J_0} \hat{\boldsymbol{X}}_j, \tilde{X}\right\}$;

13 return $\hat{\boldsymbol{X}}$

Theorem 5 (*Validity of Algorithm 2*)**.** *Let* $\boldsymbol{X} \in [0, \infty)^d$ *denote a max-id random vector with exponent measure* μ. *Then, Algorithm 2 stops after finitely many steps and its output* $\hat{\boldsymbol{X}}$ *satisfies* $\hat{\boldsymbol{X}} \sim \boldsymbol{X}$.

The use of Algorithm 2 is illustrated by the following example in which we provide an exact simulation algorithm for a large family of max-id distributions. As a byproduct, the simulation algorithm for max-stable random vectors [10, Algorithm 2] and the simulation algorithm for random vectors with reciprocal Archimedean copula [18, Algorithm 1] are unified in a common simulation scheme.

Example 2 (*Common Simulation Scheme for Max-stable Random Vectors with Unit Fréchet Margins and Random Vectors with Reciprocal Archimedean Copula*). Let $S_{\|\cdot\|} := \{\boldsymbol{x} \in [0,\infty)^d \mid \|\boldsymbol{x}\| = 1\}$ denote the non-negative part of the unit sphere of some norm $\|\cdot\|$ on $\mathbb{R}^d$. Consider an exponent measure of the form

$$\mu(A) = \mu_1 \otimes \mu_2 \left(\{(m_1, \boldsymbol{m}_2) \in (0,\infty) \times S_{\|\cdot\|} \mid m_1 \boldsymbol{m}_2 \in A\} \right), A \in \mathcal{B}\left([0,\infty)^d\right), \tag{A.1}$$

where μ_2 is a probability measure on $S_{\|\cdot\|}$ and μ_1 is a measure on $(0,\infty)$ which satisfies $\mu_1([r,\infty)) < \infty$ for all $r > 0$. Setting $\mu_1 = s^{-2}\mathrm{d}s$ and $\|\cdot\| = \|\cdot\|_1$ yields the family of max-stable distributions with unit Fréchet margins [31, Section 5], whereas setting $\|\cdot\| = \|\cdot\|_1$ and μ_2 to the uniform distribution on $S_{\|\cdot\|_1}$ yields the family of distributions with reciprocal Archimedean copula and marginal distribution function $\exp\left(-\mu_1((\cdot,\infty))\right)$ [18].

Let $(E_j)_{j\in\mathbb{N}}$ denote a sequence of iid unit Exponential random variables and let $(\boldsymbol{Y}_i)_{i\in\mathbb{N}}$ denote a sequence of iid random vectors with distribution μ_2, which is independent of $(E_j)_{j\in\mathbb{N}}$. It is well known that the standard Poisson point process on $[0,\infty)$ with unit intensity can be represented as $\sum_{i=1}^\infty \delta_{\sum_{j=1}^i E_j}$. Denoting $\mu_1^\leftarrow(t) := \sup\{s \in (0,\infty) \mid \mu_1([s,\infty)) \geq t\}$ it is easy to see that [31, Proposition 3.7] implies that $\sum_{i=1}^\infty \delta_{\mu_1^\leftarrow\left(\sum_{j=1}^i E_j\right)}$ is a PRM with intensity μ_1. Moreover, [31, Proposition 3.8] implies that

$$N = \sum_{i=1}^\infty \delta_{\left(\mu_1^\leftarrow\left(\sum_{j=1}^i E_j\right), \boldsymbol{Y}_i\right)}$$

is a PRM with intensity $\mu_1 \otimes \mu_2$. Therefore, simulating $N\left(\cdot \cap \left([r,\infty) \times S_{\|\cdot\|}\right)\right)$ is achieved by iteratively simulating the iid random vectors $(E_i, \boldsymbol{Y}_i)$ until $\mu_1^\leftarrow\left(\sum_{j=1}^i E_j\right) < r$. Note that this only requires the simulation of finitely many random vectors since $\mu_1([r,\infty)) < \infty$. Choosing $\mathcal{S}_n = \left[\frac{1}{n}, \frac{1}{n-1}\right) \times S_{\|\cdot\|}$ one can easily check that the $\mathcal{S}_n$ satisfy all the required constraints. Therefore, Algorithm 2 can be applied to exponent measures of the form (A.1). The stopping criterion of Algorithm 2 depends on the chosen norm $\|\cdot\|$, but if $\|\cdot\| = \|\cdot\|_p$ for some $p \geq 1$, it is easy to see that the algorithm stops at least as soon as $\mu_1^\leftarrow\left(\sum_{j=1}^i E_j\right) < \min_{i\notin J_0} \tilde{X}_i$.

Remark 9. Example 2 is easily extended to exponent measures of the form

$$\mu(A) = \int_0^\infty \int_S \mathbf{1}_{\{m_1\boldsymbol{m}_2\in A\}} K(m_1, \mathrm{d}\boldsymbol{m}_2)\mathrm{d}\mu_1(m_1),$$

where S denotes a bounded subset of $[0,\infty)^d \setminus \{0\}$ and $K(\cdot,\cdot)$ is a Markov kernel which satisfies $K(m_1, S) = 1$ μ_1-almost surely. For example, such ideas are used in [16, Section 3.3] to construct the finite dimensional distributions of a spatial max-id process.

Appendix B. Proofs and technical lemmas

Lemma 3 (*Lévy Measure of Stochastic Integral w.r.t. id Process*). *Let $f : [0,\infty) \times [0,\infty) \to [0,\infty)$ denote a measurable function such that $f(s,\cdot)$ is non-decreasing and right-continuous for all $s \in [0,\infty)$. Let $(H_t)_{t\geq 0}$ denote a non-negative càdlàg id process of bounded variation on compact sets with Lévy measure ν and drift b. Moreover, assume that $0 \leq H_f(t)(\omega) := \int_0^\infty f(s,t)\mathrm{d}H_s(\omega) < \infty$ for all $t \geq 0$ and $\omega \in \Omega$ and that the conditions of [30, Theorem 2.7] are satisfied. Then*

$$\left(H_f(t)\right)_{t\geq 0} = \left(\int_0^\infty f(s,t)\mathrm{d}H_s\right)_{t\geq 0}$$

defines an nnnd càdlàg id process with Lévy measure

$$\nu_f(A) = \nu\left(\left\{x \in D([0,\infty)) \ \middle| \ \int_0^\infty f(s,\cdot)\mathrm{d}x(s) \in A \text{ and } \int_0^\infty f(s,\cdot)\mathrm{d}x(s) \neq \boldsymbol{0}\right\}\right),$$

$A \in \mathcal{B}(D([0,\infty)))$ and drift $b_f(t) := \int_0^\infty f(s,t)b(\mathrm{d}s)$.

Proof of Lemma 3. Well-definedness follows from the conditions of [30, Theorem 2.7]. Infinite divisibility is obvious. The càdlàg property of H_f follows from $H_f(t) < \infty$ for all $t > 0$ and the non-decreasingness and right-continuity of $f(s,\cdot)$.

Since ν is σ-finite [32, Proposition 2.10] implies that there exists a PRM $N = \sum_{i \in \mathbb{N}} \delta_{x_i}$ with intensity ν, such that

$$(H_t)_{t \geq 0} \sim \left(\int x(t) N(\mathrm{d}x) + b(t) \right)_{t \geq 0} = \left(\sum_{i \in \mathbb{N}} x_i(t) + b(t) \right)_{t \geq 0}.$$

Moreover, N, b and ν can be chosen to be concentrated on the space of non-negative càdlàg functions of bounded variation on compact sets [32, Theorem 3.4], denoted as BV_r^+. Therefore,

$$
\begin{aligned}
\left(H_f(t) \right)_{t \geq 0} &\sim \left(\int_0^\infty f(s,t) \left(\lim_{n \to \infty} \sum_{i=1}^n x_i \right)(\mathrm{d}s) + \int_0^\infty f(s,t) b(\mathrm{d}s) \right)_{t \geq 0} \\
&= \left(\lim_{n \to \infty} \int_0^\infty f(s,t) \left(\sum_{i=1}^n x_i \right)(\mathrm{d}s) + \int_0^\infty f(s,t) b(\mathrm{d}s) \right)_{t \geq 0} \\
&= \left(\sum_{\substack{i \in \mathbb{N} \\ \int_0^\infty f(s,\cdot) x_i(\mathrm{d}s) \neq \mathbf{0}}} \int_0^\infty f(s,t) x_i(\mathrm{d}s) + \int_0^\infty f(s,t) b(\mathrm{d}s) \right)_{t \geq 0} = \left(\int_{BV_r^+} \tilde{x}(t) d\tilde{N}(\tilde{x}) + \int_0^\infty f(s,t) b(\mathrm{d}s) \right)_{t \geq 0},
\end{aligned}
$$

where

$$\tilde{N} := \sum_{\substack{i \in \mathbb{N} \\ \int_0^\infty f(s,\cdot) x_i(\mathrm{d}s) \neq \mathbf{0}}} \delta_{\int_0^\infty f(s,\cdot) x_i(\mathrm{d}s)}$$

denotes a PRM on $D([0,\infty))$ with intensity ν_f, since the map $x \mapsto \int_0^\infty f(s,\cdot) x(\mathrm{d}s)$ is measurable in $D([0,\infty))$ equipped with the sigma-algebra generated by the finite dimensional projections. It is easy to see that ν_f satisfies $\nu_f(\mathbf{0}) = 0$ and $\int_{BV_r^+} x(t) \mathrm{d}\nu_f(x) < \infty$ by the conditions of [30, Theorem 2.7]. Thus, ν_f is a Lévy measure and the lemma is proven. $\quad\square$

Proof of Theorem 2. By (11), μ_j is a finite measure for each $j \in J_0$. Thus, $\hat{X}_j$ is obtained by the simulation of a finite PRM N_j with intensity μ_j. Therefore, Algorithm 1 stops after finitely many steps if and only if the for-loop from lines 7–30 stops after finitely many steps. Thus, consider the setting of line 7 and let $\tilde{N}$ denote a PRM with intensity $\tilde{\mu}$ defined in (12). By the definition of $\tilde{\mu}$ we obtain that the associated max-id process $\tilde{X}$ satisfies $\mathbb{P}(\tilde{X}_{t_j} = 0) = 1$ for all $j \in J_0$ and $\mathbb{P}\left(\tilde{X}_{t_i} > 0 \right) = 1$ for all $i \notin J_0$. Thus, if $\tilde{X}_{t_i} = 0$ and $i \notin J_0$, there is almost surely some $c > 0$ such that $\tilde{N}_{t_i}^+ \subset \tilde{N}\left(\cdot \cap \{f \in C_0(T) \mid f(t_i) \geq c\} \right)$. Moreover, if $\tilde{X}_{t_i} > 0$, we get that $\tilde{N}_{t_i}^+ \subset \tilde{N}(\cdot \cap \{f \in C_0(T) \mid f(t_i) \geq \tilde{X}_{t_i}\})$ almost surely. Thus, the simulation of $\tilde{X}$ only requires the simulation of PRMs with finite intensity measures and stops after finitely many steps. It remains to prove that $X_t \sim \hat{X}_t$ and that $\hat{X}$ approximates X from below. Observe that $\hat{X} = \max\{\max_{j \in J_0} \hat{X}_j, \tilde{X}\}$ in line 31 is the maximum of two independent stochastic processes. The first process $\max_{j \in J_0} \hat{X}_j$ is an exact simulation of the sample path of a continuous max-id process with exponent measure $\mu\left(\cdot \cap \left(\bigcup_{j \in J_0} \{f \in C_0(T) \mid f(t_j) > 0\} \right) \right)$. The second process $\tilde{X}$ is an exact simulation of $\max_{f \in \tilde{N}_{(t_i)_{i \notin J_0}}^+} f$. Thus, $\hat{X}_t = \max\{\max_{j \in J_0} \hat{X}_j(t); \tilde{X}_t\}$ is an exact simulation of a max-id random vector with exponent measure

$$
\begin{aligned}
\sum_{j \in J_0} \mu_j &\left(\{(f(t_1), \ldots, f(t_d)) \in \cdot\} \right) + \tilde{\mu}\left(\{(f(t_1), \ldots, f(t_d)) \in \cdot\} \right) \\
&= \mu\left(\{(f(t_1), \ldots, f(t_d)) \in \cdot\} \cap \left(\bigcup_{j \in J_0} \{f(t_j) > 0\} \right) \right) + \mu\left(\{(f(t_1), \ldots, f(t_d)) \in \cdot\} \cap \{f(t_j) = 0 \ \forall j \in J_0\} \right) \\
&= \mu\left(\{(f(t_1), \ldots, f(t_d)) \in \cdot\} \right),
\end{aligned}
$$

which is the exponent measure of X_t and shows that $\hat{X}_t \sim X_t$. Furthermore, it is easy to see that

$$\hat{X} \leq \max_{\substack{f \in \tilde{N} \text{ or } f \in N_j \\ \text{for some } j \in J_0}} f \sim X,$$

which shows that $\hat{X}$ approximates X from below. $\quad\square$

Proof of Lemma 1. Recall that [12, Appendix A.3] verifies that N_t^+ and N_t^- are well-defined point measures. Thus, $|N_t^+|$ is an $\mathbb{N}_0 \cup \{\infty\}$-valued random variable and $\mathbb{E}\left[|N_t^+| \right]$ is well defined. Following the ideas of [27,28], consider some $a > 0$

and the set $A_a = \{f \in C_\mathbf{0}(T) \mid f(t_i) \geq a \text{ for some } 1 \leq i \leq d\}$. Then $\mu(A_a) < \infty$, $|N(A_a)| \sim Poi(\mu(A_a))$ and

$$\mathbb{E}\left[|N_\mathbf{t}^+(A_a)|\right] = \mathbb{E}\left[|N(A_a)|\right] - \mathbb{E}\left[|N_\mathbf{t}^-(A_a)|\right] = \int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq a \text{ for some } 1 \leq i \leq d\}} d\mu(f) - \mathbb{E}\left[\mathbb{E}\left[|N_\mathbf{t}^-(A_a)| \,\Big|\, N_\mathbf{t}^+\right]\right]$$

$$= \int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq a \text{ for some } 1 \leq i \leq d\}} d\mu(f) - \mathbb{E}\left[\int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq a \text{ for some } 1 \leq i \leq d\}} \mathbf{1}_{\{f(t_i) < X_{t_i} \text{ for all } 1 \leq i \leq d\}} d\mu(f)\right]$$

$$= \mathbb{E}\left[\int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq a \text{ for some } 1 \leq i \leq d\}} \mathbf{1}_{\{f(t_i) \geq X_{t_i} \text{ for some } 1 \leq i \leq d\}} d\mu(f)\right],$$

where we used that, conditioned on $N_\mathbf{t}^+$, $N_\mathbf{t}^-$ is distributed as a PRM with intensity $\mathbf{1}_{\{f(t_i) < \max_{\tilde{f} \in N_\mathbf{t}^+} \tilde{f}(t_i)\}} d\mu(f)$. We conclude by considering three cases:

(i) Assume that $\mathbb{P}(X_{t_j} > 0) = 1$ for all $1 \leq i \leq d$. When $a \searrow 0$ the monotone convergence theorem implies that

$$\mathbb{E}\left[|N_\mathbf{t}^+|\right] = \mathbb{E}\left[\int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq X_{t_i} \text{ for some } 1 \leq i \leq d\}} d\mu(f)\right].$$

(ii) If $\mathbb{P}(X_{t_j} = 0) > 0$ for some $1 \leq i \leq d$ and μ is a finite measure then one may set $a = 0$, which immediately implies

$$\mathbb{E}\left[|N_\mathbf{t}^+|\right] = \mathbb{E}\left[\int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq X_{t_i} \text{ for some } 1 \leq i \leq d\}} d\mu(f)\right].$$

(iii) If $\mathbb{P}(X_{t_j} = 0) > 0$ for some $1 \leq i \leq d$ and μ is an infinite measure then

$$\mathbb{E}\left[|N_\mathbf{t}^+|\right] = \infty = \mathbb{E}\left[\int_{C_\mathbf{0}(T)} \mathbf{1}_{\{f(t_i) \geq X_{t_i} \text{ for some } 1 \leq i \leq d\}} d\mu(f)\right],$$

since $\mathbb{P}(|N_\mathbf{t}^+| = \infty) \geq \mathbb{P}(X_{t_i} = 0 \text{ for some } 1 \leq i \leq d) > 0$. $\quad\square$

Proof of Theorem 3. Obviously, the expected number of simulated functions (atoms) of the PRMs with intensities $(\mu_j)_{j \in J_0}$ is

$$\sum_{j \in J_0} \mu_j(C_\mathbf{0}(T)) = \sum_{j \in J_0} \mu\left(\cdot \cap \left\{f \in C_\mathbf{0}(T) \mid f(t_j) > 0, f(t_k) = 0, \ k < j, \ k \in J_0\right\}\right)$$

$$= \mu\left(\left\{f \in C_\mathbf{0}(T) \mid f(t_j) > 0 \text{ for some } j \in J_0\right\}\right).$$

Thus, the expected number of functions that need to be simulated to obtain $\left(\hat{X}_j\right)_{j \in J_0}$ is equal to $\mu\left(\left\{f \in C_\mathbf{0}(T) \mid f(t_j) > 0\right.\right.$ for some $j \in J_0 \Big\}\Big)$.

It remains to compute the expected number of simulated functions to obtain $\tilde{X}$. At each location $(t_i)_{i \notin J_0}$, according to Algorithm 1 and assumption (13), we can consecutively simulate the atoms $f_1^{(i)}, f_2^{(i)}, \ldots$ of a PRM $\tilde{N}^{(i)}$ with intensity $\tilde{\mu}$ such that $f_1^{(i)}(t_i) \geq f_2^{(i)}(t_i) \geq \cdots$ until the first subextremal function is found. Since all simulated atoms which satisfy $f_j^{(i)}(t_k) \geq \tilde{X}_{t_k}$ for some $k < i, k \notin J_0$, are rejected we obtain that the number of functions that need to be simulated to obtain $\tilde{X}$ is

$$\left|\tilde{N}_{(t_i)_{i \notin J_0}}^+\right| + \sum_{i \notin J_0}\left(\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq X_{t_{i_k}} \text{ for some } k \notin J_0, k < i; \ f_j^{(i)}(t_i) \geq X_{t_i}\right\}\right| + 1\right).$$

Note that the number of rejected functions is increased by 1, since we have to simulate until the first subextremal function at each location t_i is obtained. Thus, the expected number of functions that need to be simulated to obtain $\tilde{X}$ is given by

$$\mathbb{E}\left[\left|\tilde{N}_{(t_i)_{i \notin J_0}}^+\right|\right] + d - |J_0| + \sum_{i \notin J_0} \mathbb{E}\left[\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; \ f_j^{(i)}(t_i) \geq \tilde{X}_{t_i}\right\}\right|\right].$$

The expectation of the first term is provided by Lemma 1. To determine the remaining expectation we calculate

$$\mathbb{E}\left[\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; \ f_j^{(i)}(t_i) \geq \tilde{X}_{t_i}\right\}\right|\right] =$$

$$\mathbb{E}\left[\mathbb{E}\left[\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; \ f_j^{(i)}(t_i) \geq \tilde{X}_{t_i}\right\}\right| \,\Big|\, \tilde{N}_{(t_k)_{k \notin J_0, k < i}}^+,\right.\right.$$

$$\left.\left.\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) < \tilde{X}_{t_k} \text{ for all } k \notin J_0, k < i\right\}\right]\right].$$

Note that $\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) < \tilde{X}_{t_k} \text{ for all } k \notin J_0, k < i\right\}$ and $\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i\right\}$ are disjoint measurable sets and therefore, conditioned on $(X_{t_k})_{k \notin J_0, k < i}$, the restrictions of the PRM $\tilde{N}^{(i)}$ on each of the two sets are independent PRMs with intensities $\mathbf{1}_{\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) < \tilde{X}_{t_k} \text{ for all } k \notin J_0, k < i\right\}}$ and $\mathbf{1}_{\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i\right\}}$. Moreover, since $\tilde{N}^+_{(t_k)_{k \notin J_0, k < i}}$ and $\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) < \tilde{X}_{t_k} \text{ for all } k \notin J_0, k < i\right\}$ determine $\left(\tilde{X}_{t_k}\right)_{k \notin J_0, k < i}$ and $\tilde{X}_{t_i}$ we get

$$\mathbb{E}\left[\mathbb{E}\left[\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; f_j^{(i)}(t_i) \geq \tilde{X}_{t_i}\right\}\right| \;\middle|\; \tilde{N}^+_{(t_k)_{k \notin J_0, k < i}}, \right.\right.$$
$$\left.\left. \left\{f_j^{(i)} \mid f_j^{(i)}(t_k) < \tilde{X}_{t_k} \text{ for all } k \notin J_0, k < i\right\}\right]\right]$$
$$= \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; f(t_i) \geq \tilde{X}_{t_i}\}} \mathrm{d}\tilde{\mu}(f)\right]$$
$$= \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i\}} \mathrm{d}\tilde{\mu}(f)\right] - \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; f(t_i) < \tilde{X}_{t_i}\}} \mathrm{d}\tilde{\mu}(f)\right].$$

Now, Lemma 1 implies

$$\mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i\}} \mathrm{d}\tilde{\mu}(f)\right] - \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; f(t_i) < \tilde{X}_{t_i}\}} \mathrm{d}\tilde{\mu}(f)\right]$$
$$= \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k < i}\right|\right] - \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k \leq i\}} \left(1 - \mathbf{1}_{\{f(t_i) \geq \tilde{X}_{t_i}\}}\right) \mathrm{d}\tilde{\mu}(f)\right]$$
$$= \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k < i}\right|\right] - \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k \leq i}\right|\right] + \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_i) \geq \tilde{X}(t_i), f(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k \leq i\}} \mathrm{d}\tilde{\mu}(f)\right]$$
$$= \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k < i}\right|\right] - \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k \leq i}\right|\right] + \mathbb{E}\left[\int_{C_0(T)} \mathbf{1}_{\{f(t_i) \geq \tilde{X}(t_i)\}} \mathrm{d}\tilde{\mu}(f)\right]$$
$$= \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k < i}\right|\right] - \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k \leq i}\right|\right] + \mathbb{E}\left[\tilde{\mu}\left(\left\{f \in C_0(T) \mid f(t_i) \in [\tilde{X}_{t_i}, \infty)\right\}\right)\right].$$

Thus,

$$\mathbb{E}\left[\left|\tilde{N}^+_{(t_i)_{i \notin J_0}}\right|\right] + d - |J_0| + \sum_{i \notin J_0} \mathbb{E}\left[\left|\left\{f_j^{(i)} \mid f_j^{(i)}(t_k) \geq \tilde{X}_{t_k} \text{ for some } k \notin J_0, k < i; f_j^{(i)}(t_i) \geq \tilde{X}_{t_k}\right\}\right|\right]$$
$$= \mathbb{E}\left[\left|\tilde{N}^+_{(t_i)_{i \notin J_0}}\right|\right] + d - |J_0| + \sum_{i \notin J_0} \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k < i}\right|\right] - \mathbb{E}\left[\left|\tilde{N}^+_{(t_k)_{k \notin J_0}, k \leq i}\right|\right] + \mathbb{E}\left[\tilde{\mu}\left(\left\{f \in C_0(T) \mid f(t_i) \in [\tilde{X}_{t_i}, \infty)\right\}\right)\right]$$
$$= d - |J_0| + \sum_{i \notin J_0} \mathbb{E}\left[\tilde{\mu}\left(\left\{f \in C_0(T) \mid f(t_i) \in [\tilde{X}_{t_i}, \infty)\right\}\right)\right]. \tag{B.1}$$

If $\tilde{X}_{(t_i)_{i \notin J_0}}$ has continuous marginal distribution we can stop as soon as we found an extremal function at each location $(t_i)_{i \notin J_0}$. Therefore, the term $d - |J_0|$, which comes from the simulation of the first subextremal function, may be omitted from (B.1) and we get

$$\sum_{i \notin J_0} \mathbb{E}\left[\tilde{\mu}\left(\left\{f \in C_0(T) \mid f(t_i) \in [\tilde{X}_{t_i}, \infty)\right\}\right)\right] = \sum_{i \notin J_0} \mathbb{E}\left[\tilde{\mu}\left(\left\{f \in C_0(T) \mid f(t_i) \in (\tilde{X}_{t_i}, \infty)\right\}\right)\right]$$
$$= \sum_{i \notin J_0} \mathbb{E}\left[-\log\left(F_{t_i}\left(\tilde{X}_{t_i}\right)\right)\right]$$
$$= d - |J_0|,$$

where $F_t(x) := \mathbb{P}(\tilde{X}_t \leq x)$ denotes the marginal distribution function of $\tilde{X}_t$ and it is well known that $-\log\left(F_{t_i}(\tilde{X}_{t_i})\right) \sim \text{Exp}(1)$, since $F_{t_i}(\tilde{X}_{t_i})$ is uniformly distributed on $[0, 1]$ when $\tilde{X}_t$ follows a continuous distribution. Combining the results above we obtain the claimed complexity of Algorithm 1. $\square$

Proof of Lemma 2. Note that the Lévy measure ν_k of $L^{(k)}$ on $\mathbf{M}$ is given by the image measure of the map

$$\left(((0, \infty), \lambda_{0, \infty}) \times ((0, \infty), \rho_k)\right) \to \mathbf{M}; \; (s, a) \mapsto a\mathbf{1}_{\{\cdot \geq s\}},$$

where $\lambda_{0, \infty}$ denotes the Lebesgue measure on $(0, \infty)$ and ρ_k denotes the univariate Lévy measure of $L_1^{(k)}$. To derive the path Lévy measure of the self similar subordinator $(H_t)_{t \geq 0}$, we first need to derive the path Lévy measures of the two

independent id processes $\hat{L}_t^{(1)} := \int_{-\log(\min\{t; 1\})}^{\infty} \exp(-s) dL_s^{(k,1)}$ and $\hat{L}_t^{(2)} := \int_0^{\log(\max\{1, t\})} \exp(s) dL_s^{(k,2)}$. Note that for $t \in [0, 1]$

$$\int_{-\log(t)}^{\infty} \exp(-y) \left(a\mathbf{1}_{\{\cdot \geq s\}}\right) (dy) = a \exp(-s)\mathbf{1}_{\{s \geq -\log(t)\}}$$

and for $t > 1$

$$\int_0^{\log(t)} \exp(y) \left(a\mathbf{1}_{\{\cdot \geq s\}}\right) (dy) = a \exp(s)\mathbf{1}_{\{s \leq \log(t)\}}.$$

An application of Lemma 3 shows that the Lévy measure of $\hat{L}^{(1)}$ is given by

$$\nu_1(A) = \lambda_{0,\infty} \otimes \rho_k \left(\left\{ (s, a) \;\middle|\; a \exp(-s)\mathbf{1}_{\{s \geq -\log(\min\{\cdot, 1\})\}} \in A \right\} \right); \quad A \in \mathcal{B}(\mathbf{M})$$

and that the Lévy measure of $\hat{L}^{(2)}$ is given by

$$\nu_2(A) = \lambda_{0,\infty} \otimes \rho_k \left(\left\{ (s, a) \;\middle|\; a \exp(s)\mathbf{1}_{\{s \leq \log(\max\{\cdot, 1\})\}} \in A \right\} \right); \quad A \in \mathcal{B}(\mathbf{M}).$$

This implies that the path Lévy measure of the self-similar subordinator H is given by $\nu = \nu_1 + \nu_2$, since $\hat{L}^{(1)}$ and $\hat{L}^{(2)}$ are independent. It remains to verify (21). To this purpose we simply verify that the Laplace transform of H coincides with the Laplace transform of an id process with path Lévy measure (21), since a path Lévy measure is unique.

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(\sum_{i=1}^d z_i H(t_i) \right) \right] &= \exp\left(\int_0^{\infty} \int_0^{\infty} \left(1 - \exp\left(-\sum_{i=1}^d z_i a \exp(-s)\mathbf{1}_{\{s \geq -\log(\min\{t_i, 1\})\}} \right) \right) ds \rho_k(da) \right) \\
&\quad + \exp\left(\int_0^{\infty} \int_0^{\infty} \left(1 - \exp\left(-\sum_{i=1}^d z_i a \exp(s)\mathbf{1}_{\{s \leq \log(\max\{t_i, 1\})\}} \right) \right) ds \rho_k(da) \right) \\
&= \exp\left(\int_0^{\infty} \int_0^1 \left(1 - \exp\left(-\sum_{i=1}^d z_i a s \mathbf{1}_{\{-\log(s) \geq -\log(\min\{t_i, 1\})\}} \right) \right) s^{-1} ds \rho_k(da) \right) \\
&\quad + \exp\left(\int_0^{\infty} \int_1^{\infty} \left(1 - \exp\left(-\sum_{i=1}^d z_i a s \mathbf{1}_{\{\log(s) \leq \log(\max\{t_i, 1\})\}} \right) \right) s^{-1} ds \rho_k(da) \right) \\
&= \exp\left(\int_0^{\infty} \int_0^{\infty} \left(1 - \exp\left(-\sum_{i=1}^d z_i a s \mathbf{1}_{\{t_i \geq s\}} \right) \right) s^{-1} ds \rho_k(da) \right). \quad \square
\end{aligned}
$$

Remark 10 (*Path Lévy Measure of General Self-Similar Processes*)**.** The path Lévy measure representation in (21) is not only valid for nnnd self-similar processes but also valid for general self-similar processes, where ρ_k denotes the Lévy measure of the BDLP of H_1. Moreover, since a self-similar process with index $\gamma > 0$ corresponds to a time change of a self-similar process with index 1, the path Lévy measure $\nu^{(\gamma)}$ of a self-similar process with index γ is simply obtained by applying the same "time change" to the Lévy measure of the self-similar process with index 1, i.e. by the image measure of $(\mathbf{M}, \nu) \to \mathbf{M}$, $(f(t))_{t \geq 0} \mapsto (f(t^{\gamma}))_{t \geq 0}$.

Proof of Theorem 5. The μ_j are finite intensity measures by their definition in (11). Therefore, Algorithm 2 stops after finitely many steps if and only if the while-loop from lines 7–11 stops after finitely many steps. It is obvious that the simulation of each PRM $\tilde{N}_n$ only requires the simulation of finitely many points. Thus, we need to check that the condition $C := \{$there is no $r > 0$ such that $\cup_{m \geq n} \mathcal{S}_m \subset B_r^{\infty} (\mathbf{0})$ **and** $\min_{i \notin J_0} \tilde{X}_i \geq r \}$ is violated after finitely many steps. Let $\tilde{N}$ denote the PRM with intensity $\tilde{\mu}$. It is easy to see that condition C is eventually violated after finitely many steps if and only if $\min_{i \notin J_0, \mathbf{x} \in \tilde{N}} x_i > 0$ almost surely. By the construction of $\tilde{\mu}$ we have $\mathbb{P}(\tilde{X}_i = 0) = 0$ for all $i \notin J_0$, which implies that $\min_{i \notin J_0, \mathbf{x} \in \tilde{N}} x_i > 0$ almost surely and the algorithm stops after finitely many steps.

It remains to prove that $\hat{X} \sim X$. Clearly, if condition C is violated for some $n \in \mathbb{N}$ and $r > 0$, then all points of the PRM $\tilde{N}$ in $\cup_{m < n} \mathcal{S}_m$ have already been simulated and $\tilde{X} = \max_{\mathbf{x} \in \tilde{N}(\cdot \cap (\cup_{m < n} \mathcal{S}_m))} \mathbf{x}$. A point $\mathbf{x} \in \tilde{N} \left(\cdot \cap (\cup_{m \geq n} \mathcal{S}_m) \right)$ can only increase a non-zero component $(\tilde{X}_j)_{j \notin J_0}$ of $\tilde{X}$ if $\max_{i \notin J_0} x_i \geq r$. However, since $\cup_{m \geq n} \mathcal{S}_m \subset B_r^{\infty}$, we actually have that $\tilde{X} = \max_{\mathbf{x} \in \tilde{N}} \mathbf{x}$. Thus, $\tilde{X}$ is max-id with exponent measure $\tilde{\mu}$. Combining this with the fact that the μ_j and $\tilde{\mu}$ are supported on disjoint sets, we obtain that $\hat{X}$ is max-id with exponent measure $\sum_{j \in J_0} \mu_j + \tilde{\mu} = \mu$, which proves the claim. $\square$

References

[1] S. Asmussen, J. Rosiński, Approximations of small jumps of Lévy processes with a view towards simulation, J. Appl. Probab. 38 (2) (2001) 482–493.

[2] A.A. Balkema, L. de Haan, R.L. Karandikar, Asymptotic distribution of the maximum of n independent stochastic processes, J. Appl. Probab. 30 (1) (1993) 66–81.

[3] O.E. Barndorff-Nielsen, Normal inverse Gaussian distributions and stochastic volatility modelling, Scand. J. Stat. 24 (1) (1997) 1–13.

[4] G. Bernhart, J.-F. Mai, M. Scherer, On the construction of low-parametric families of min-stable multivariate exponential distributions in large dimensions, Depend. Model. 3 (1) (2015) 29–46.

[5] L. Bondesson, On simulation from infinitely divisible distributions, Adv. Appl. Probab. 14 (4) (1982) 855–869.

[6] G.P. Bopp, B.A. Shaby, R. Huser, A hierarchical max-infinitely divisible spatial model for extreme precipitation, J. Amer. Statist. Assoc. 116 (533) (2021) 93–106.

[7] F. Brück, J.-F. Mai, M. Scherer, Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes, Extremes (2022) http://dx.doi.org/10.1007/s10687-022-00450-w.

[8] P. Carr, H. Geman, D.B. Madan, M. Yor, The fine structure of asset returns: An empirical investigation, J. Bus. 75 (2) (2002) 305–332.

[9] P. Damien, P.W. Laud, A.F.M. Smith, Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference, J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (3) (1995) 547–563.

[10] C. Dombry, S. Engelke, M. Oesting, Exact simulation of max-stable processes, Biometrika 103 (2) (2016) 303–317.

[11] C. Dombry, F. Eyi-Minko, Strong mixing properties of max-infinitely divisible random fields, Stochastic Process. Appl. 122 (11) (2012) 3790–3811.

[12] C. Dombry, F. Eyi-Minko, et al., Regular conditional distributions of continuous max-infinitely divisible random fields, Electron. J. Probab. 18 (2013) 1–21.

[13] C. Genest, J.G. Nešlehová, L.-P. Rivest, The class of multivariate max-id copulas with ℓ_1-norm symmetric exponent measure, Bernoulli 24 (2018) 3751–3790.

[14] E. Giné, M.G. Hahn, P. Vatan, Max-infinitely divisible and max-stable sample continuous processes, Probab. Theory Related Fields 87 (1990) 139–165.

[15] C. Halgreen, Self–decomposability of the generalized inverse Gaussian and hyperbolic distributions, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 47 (1) (1979) 13–17.

[16] R. Huser, T. Opitz, E. Thibaud, Max-infinitely divisible models and inference for spatial extremes, Scand. J. Stat. 48 (1) (2021) 321–348.

[17] M. Jeanblanc, J. Pitman, M. Yor, Self-similar processes with independent increments associated with Lévy and Bessel processes, Stochastic Process. Appl. 100 (1) (2002) 223–231.

[18] J.-F. Mai, Exact simulation of reciprocal Archimedean copulas, Statist. Probab. Lett. 141 (2018) 68–73.

[19] J.-F. Mai, Canonical spectral representation for exchangeable max-stable sequences, Extremes 23 (1) (2020) 151–169.

[20] J.-F. Mai, S. Schenk, M. Scherer, Exchangeable exogenous shock models, Bernoulli 22 (2) (2016) 1278–1299.

[21] J.-F. Mai, S. Schenk, M. Scherer, Two novel characterizations of self-decomposability on the half-line, J. Theoret. Probab. 30 (1) (2017) 365–383.

[22] J.-F. Mai, M. Scherer, Lévy–frailty copulas, J. Multivariate Anal. 100 (7) (2009) 1567–1585.

[23] J.-F. Mai, M. Scherer, Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time, Extremes 17 (1) (2014) 77–95.

[24] J.-F. Mai, M. Scherer, Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications, in: Series in Quantitative Finance, vol. 6, World Scientific, 2017.

[25] J.-F. Mai, M. Scherer, Subordinators which are infinitely divisible wrt time: Construction, properties, and simulation of max-stable sequences and infinitely divisible laws, Lat. Am. J. Probab. Math. Stat. (16) (2019) 1–29.

[26] A.W. Marshall, I. Olkin, A multivariate exponential distribution, J. Amer. Statist. Assoc. 62 (317) (1967) 30–44.

[27] M. Oesting, M. Schlather, C. Zhou, On the normalized spectral representation of max-stable processes on a compact set, 2013, arXiv.

[28] M. Oesting, M. Schlather, C. Zhou, Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation, Bernoulli 24 (2) (2018) 1497–1530.

[29] S.A. Padoan, Extreme dependence models based on event magnitude, J. Multivariate Anal. 122 (2013) 1–19.

[30] B.S. Rajput, J. Rosiński, Spectral representations of infinitely divisible processes, Probab. Theory Related Fields 82 (3) (1989) 451–487.

[31] S.I. Resnick, Extreme Values, Regular Variation and Point Processes, Springer, 2013.

[32] J. Rosiński, Representations and isomorphism identities for infinitely divisible processes, Ann. Probab. 46 (6) (2018) 3229–3274.

[33] K.-I. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.

[34] M. Scherer, H. Sloot, Exogenous shock models: analytical characterization and probabilistic construction, Metrika 82 (8) (2019) 931–959.

[35] M. Schlather, Models for stationary max-stable random fields, Extremes 5 (2002) 33–44.

[36] W. Schoutens, J.L. Teugels, Lévy processes, polynomials and martingales, Commun. Statist. Stoch. Models 14 (1–2) (1998) 335–349.

[37] P. Vatan, Max-infinite divisibility and max-stability in infinite dimensions, in: Probability in Banach Spaces V, Springer, 1985, pp. 400–425.

[38] P. Zhong, R. Huser, T. Opitz, Exact simulation of max-infinitely divisible processes, Econom. Statist. (2022).

B Further Publication

B.1 A corrected Clarke test for model selection and beyond [9]

Summary

The main result of this article is that we provide an omnibus statistical test for hypotheses of the form

$$H_0: \ \mathbb{E}\left[\mathsf{C}(f_1(X, \theta_1^\star), f_2(X, \theta_2^\star), X)\right] = 0, \tag{17}$$

where C denotes a fixed, real-valued, possibly non-smooth (criterion) function and $f_1(\cdot, \theta_1^\star)$ and $f_2(\cdot, \theta_2^\star)$ denote the "optimal" representatives of some families of possibly non-smooth models $M_1 := \{f_1(\cdot, \theta_1) \mid \theta_1 \in \Theta_1\}$ and $M_2 := \{f_2(\cdot, \theta_2) \mid \theta_2 \in \Theta_2\}$ for some property of the random vector X. The main advantage of our framework is that we do not impose any smoothness assumptions on $(\mathsf{C}, f_1(\cdot, \cdot), f_2(\cdot, \cdot))$, which is usually a source of technical difficulties. Instead, we require that the function $(\theta_1, \theta_2) \mapsto \mathbb{E}\left[\mathsf{C}(f_1(X, \theta_1), f_2(X, \theta_2), X)\right]$ is differentiable at $(\theta_1^\star, \theta_2^\star)$, which may be interpreted as a smoothed version of $\mathsf{C}(f_1(\cdot, \cdot), f_2(\cdot, \cdot), \cdot)$. With an emphasis on non-smooth $(\mathsf{C}, f_1(\cdot, \cdot), f_2(\cdot, \cdot))$ and econometric applications, we provide several examples that illustrate the wide applicability of our framework.

We further prove that the test statistic

$$T_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{C}(f_1(X_i, \hat{\theta}_{1,n}), f_2(X_i, \hat{\theta}_{2,n}), X_i)$$

asymptotically follows a normal distribution under H_0, where $(\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ denotes some (asymptotically normal) estimator of $(\theta_1^\star, \theta_2^\star)$ and $(X_i)_{i \in \mathbb{N}}$ denotes an i.i.d. sequence of samples from X. Moreover, we provide estimators for the asymptotic variance of T_n as well as its bootstrap consistency. The core assumption which is needed to deduce these results is

$$A: \ \lim_{n \to \infty} \mathbb{E}\left[\left(\mathsf{C}(f_1(X, \theta_1^\star), f_2(X, \theta_2^\star), X) - \mathsf{C}(f_1(X, \hat{\theta}_{1,n}), f_2(X, \hat{\theta}_{2,n}), X)\right)^2\right] \to 0$$

in probability, which essentially ensures that the fluctuation of $\mathsf{C}(f_1(\cdot, \hat{\theta}_{1,n}), f_2(\cdot, \hat{\theta}_{2,n}), \cdot)$ around $\mathsf{C}(f_1(\cdot, \theta_{1,n}^\star), f_2(\cdot, \theta_{2,n}^\star), \cdot)$ is small with high probability.

A particularly important example of a non-smooth function C is given by $\mathsf{C}(f_1, f_2) = \mathbf{1}_{\{\log(f_1) - \log(f_2) > 0\}} - 1/2$, which yields the null hypothesis of the Clarke test [15, 14] for non-nested model comparison of competing families of parametric densities $f_1(\cdot, \theta_1)$ and $f_2(\cdot, \theta_2)$ given by

$$H_0^C: \ P\big(\log(f_1(X, \theta_1^\star)) - \log(f_2(X, \theta_2^\star)) > 0\big) = \frac{1}{2}.$$

The author of [15, 14] claimed that $\sum_{i=1}^{n} \mathbf{1}_{\{\log(f_1(X_i, \hat{\theta}_{1,n})) - \log(f_2(X_i, \hat{\theta}_{2,n})) > 0\}}$ is Binomial distributed under H_0^C. However, we show that this claim is not even asymptotically correct and our results imply that $\sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\mathbf{1}_{\{\log(f_1(X_i, \hat{\theta}_{1,n})) - \log(f_2(X_i, \hat{\theta}_{2,n})) > 0\}} - 1/2\right)$ asymptotically follows

a normal distribution under H_0^C. Furthermore, it is easy to see that H_0^C is not meaningful if $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$, i.e. when the "optimal" models in both families are identical. Therefore, we have introduced the so-called modified Clarke null hypothesis

$$\tilde{H}_0^C : \ P\big(\log\left(f_1(X, \theta_1^\star)\right) - \log\left(f_2(X, \theta_2^\star)\right) > 0\big) = P\big(\log\left(f_1(X, \theta_1^\star)\right) - \log\left(f_2(X, \theta_2^\star)\right) < 0\big),$$

which may be embedded into our general framework by choosing $\mathsf{C}(f_1, f_2) = \mathbf{1}_{\{\log(f_1) - \log(f_2) > 0\}} - \mathbf{1}_{\{\log(f_1) - \log(f_2) < 0\}}$. This null hypothesis sensibly generalizes H_0^C to situations where $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$ may not be excluded. However, assumption A may not be satisfied in some situations when $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$, which is why we propose an alternative testing procedure of $\tilde{H}_0^C$ that essentially uses a smoothed version of C.

A simulation study illustrates the finite sample properties of our testing framework in a linear and quantile regression setting. Moreover, we illustrate the finite sample properties of our proposed test statistic for the modified Clarke null hypothesis in the case $f_1(\cdot, \theta_1^\star) = f_2(\cdot, \theta_2^\star)$. An empirical analysis of stock return data illustrates that the original Clarke test proposed in [15, 14] may lead to erroneous conclusions in practice and that our framework mitigates these problems.

As a side product, we provide sufficient conditions such that the empirical variance of bootstrap replicates of an arbitrary test statistic converges to the true asymptotic variance of this test statistic. This way of estimating asymptotic variances seems to be a common approach in practice, but, to the best of our knowledge, has not been theoretically verified in a general framework before.

Individual contributions

In my master thesis [7] I solely investigated the asymptotic behavior of the Clarke test for non-nested models. I proved that the stated asymptotic distribution of the Clarke test in [15, 14] is wrong and provided the asymptotic normality, bootstrap consistency and consistency of the estimators of the asymptotic variance for the particular case of the Clarke test for non-nested models. The idea to investigate the asymptotic behavior of the Clarke test in my master thesis was developed by Jean-David Fermanian and Aleksey Min, the proofs were all developed and carried out by myself, relying on some hints of Jean-David Fermanian. For my doctoral thesis, to explicitly prevent double-counting, only those parts of this paper that go beyond my master thesis are to be accredited.

During my PhD we have substantially generalized the scope of the paper. We now consider a general, possibly non-smooth, criterion function. The Clarke test is merely a particular example which may be deduced from a specific choice of a non-smooth criterion function and is discussed in a separate section of the paper. All the proofs and results have been adapted to the generalized framework. Jean-David Fermanian and I were responsible for extending the framework and the proofs.

Moreover, we have introduced a modified Clarke test, which is also applicable to nested models.

The idea for the modified Clarke test was generated by me and the test statistic proposed to test the modified Clarke null hypothesis was developed by Jean-David Fermanian. Additionally, we have provided the consistency of an estimator for the asymptotic variance of an arbitrary test statistic, which is based on the empirical variance of the bootstrap replicates of this test statistic. The formulation and the proof of this result were carried out by me.

Several examples for possible applications of our tests were developed by Jean-David Fermanian and refined by me. The simulation section was mainly developed by Aleksey Min, while Jean-David Fermanian and I were significantly involved in selecting the provided examples and I was mainly responsible for theoretical verification of the applicability of our testing procedure to the stated examples. The empirical analysis was developed by Jean-David Fermanian and the corresponding simulations were carried out by Aleksey Min. Everybody was significantly involved in the writing of all parts of the paper so that it is impossible to exactly determine which parts were written by which person.

I am the main author of this article.

Permission to include the article

Permissions

As a general rule, permission should be sought from the rights holder to reproduce any substantial part of a copyrighted work. This includes any text, illustrations, charts, tables, photographs, or other material from previously published sources. Obtaining permission to re-use content published by Elsevier is simple. Follow the guide below for a quick and easy route to permission.

Permission guidelines   ScienceDirect content   ClinicalKey content   Tutorial videos   Help and support

Permission guidelines

For further guidelines about obtaining permission, please review our Frequently Asked Questions below:

When is permission required?	⌄
When is permission not required?	⌄
From whom do I need permission?	⌄
How do I obtain permission to use photographs or illustrations?	⌄
Do I need to obtain permission to use material posted on a website such as Blogs/Google images/e-commerce websites?	⌄
What rights does Elsevier require when requesting permission?	⌄
How do I obtain permission from another publisher?	⌄
What is RightsLink/CCC?	⌄
What should I do if I am not able to locate the copyright owner?	⌄
Can I obtain permission from a Reproduction Rights Organization (RRO)?	⌄
Is Elsevier an STM signatory publisher?	⌄
Do I need to request permission to re-use work from another STM publisher?	⌄
Do I need to request permission to text mine Elsevier content?	⌄
Can I include/use my article in my thesis/dissertation?	⌃

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.

For any further clarifications, you can submit your query via our online form (https://service.elsevier.com/app/contact/supporthub/permissions-helpdesk)

Which uses of a work does Elsevier view as a form of 'prior publication'?	⌄
How do I obtain permission to use Elsevier Journal material such as figures, tables, or text excerpts, if the request falls within the STM permissions guidelines?	⌄
How do I obtain permission to use Elsevier Journal material such as figures, tables, or text excerpts, if the amount of material I wish to use does not fall within the free limits set out in the STM permissions guidelines?	⌄
How do I obtain permission to use Elsevier Book material such as figures, tables, or text excerpts?	⌄
How do I obtain permission to use Elsevier material that is NOT on ScienceDirect or Clinical Key?	⌄
Can I use material from my Elsevier journal article within my thesis/dissertation?	⌃

As an Elsevier journal author, you have the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see the Copyright page (https://www.elsevier.com/about/policies/copyright)for more information. No written permission from Elsevier is necessary.

This right extends to the posting of your thesis to your university's repository provided that if you include the published journal article, it is embedded in your thesis and not separately downloadable.

For any further clarifications, you can submit your query via our online form (https://service.elsevier.com/app/contact/supporthub/permissions-helpdesk)

Can I modify a figure when I have received permission to use it?	⌄
Can I modify Lancet material?	⌄
How do I obtain copies of a book in a format suitable for a person with a disability?	⌄
What is the National Institute of Health's (NIH) policy for depositing articles to PubMed Central?	⌄
How should I acknowledge Elsevier material?	⌄
How do I set up a RightsLink account?	⌄
As an Elsevier book author can I post my chapter on a website?	⌄

Search by keyword, title, subject area   🔍

Copyright

Overview   Author rights   Institution rights   Government rights   Find out more

Overview

In order for Elsevier to publish and disseminate research articles, we need certain publishing rights from authors, which are determined by a publishing agreement between the author and Elsevier.

For articles published open access, the authors license exclusive rights in their article to Elsevier where a CC BY-NC-ND end user license is selected, and license non-exclusive rights where a CC BY end user license is selected.

For articles published under the subscription model, the authors typically transfer copyright to Elsevier. In some circumstances, authors may instead grant us (or the learned society for whom we publish) an exclusive license to publish and disseminate their work.

Regardless of whether they choose to publish open access or subscription with Elsevier, authors have many of the same rights under our publishing agreement, which support their need to share, disseminate and maximize the impact of their research.

For open access articles, authors will also have additional rights, depending on the Creative Commons end user license that they select. This Creative Commons license sets out the rights that readers (as well as the authors) have to re-use and share the article: please see here (https://www.elsevier.com/about/policies/open-access-licenses) for more information on how articles can be re-used and shared under these licenses.

This page aims to summarise authors' rights when publishing with Elsevier; these are explained in more detail in the publishing agreement between the author and Elsevier.

Irrespective of how an article is published, Elsevier is committed to protect and defend authors' works and their reputation. We take allegations of infringement, plagiarism, ethical disputes, and fraud very seriously.

Author rights

The below table explains the rights that authors have when they publish with Elsevier, for authors who choose to publish either open access or subscription. These apply to the corresponding author and all co-authors.

Author rights in Elsevier's proprietary journals	Published open access	Published subscription
Retain patent and trademark rights	√	√
Retain the rights to use their research data freely without any restriction	√	√
Receive proper attribution and credit for their published work	√	√
Re-use their own material in new works without permission or payment (with full acknowledgement of the original article): 1. Extend an article to book length 2. Include an article in a subsequent compilation of their own work 3. Re-use portions, excerpts, and their own figures or tables in other works.	√	√
Use and share their works for scholarly purposes (with full acknowledgement of the original article): 1. In their own classroom teaching. Electronic and physical distribution of copies is permitted 2. If an author is speaking at a conference, they can present the article and distribute copies to the attendees 3. Distribute the article, including by email, to their students and to research colleagues who they know for their personal use 4. Share and publicize the article via Share Links, which offers 50 days' free access for anyone, without signup or registration 5. Include in a thesis or dissertation (provided this is not published commercially) 6. Share copies of their article privately as part of an invitation-only work group on commercial sites with which the publisher has a hosting agreement	√	√
Publicly share the preprint on any website or repository at any time.	√	√
Publicly share the accepted manuscript on non-commercial sites	√	√ using a CC BY-NC-ND license and usually only after an embargo period (see Sharing Policy (https://www.elsevier.com/about/policies/sharing) for more information)
Publicly share the final published article	√ in line with the author's choice of end user license	×

A corrected Clarke test for model selection and beyond

Florian Brück [a,*], Jean-David Fermanian [b,1], Aleksey Min [a]

[a] *Technical University of Munich, Chair of Mathematical Finance, Parkring 11, 85748 Garching b. München, Germany*
[b] *Crest-ENSAE, 5, avenue Henry Le Chatelier, 91764 Palaiseau cedex, France*

ABSTRACT

We introduce a large family of model selection tests based on the expectation of an arbitrary, possibly non-smooth, parametric criterion function of the data. The considered methodology is illustrated for several econometric problems, including linear and quantile regression. It covers the case of strictly locally non-nested models and some overlapping models. The asymptotic theory of the proposed test statistic is stated. A general exchangeable bootstrap scheme allows the evaluation of its limiting law as well as its asymptotic variance. Our framework includes the tests for non-nested model selection of Vuong (1989) and Clarke (2007) as particular cases. We show that the statistic of the latter test is not Binomial distributed as originally stated and we provide its corrected limiting law. In a simulation study, we empirically verify the distributional approximation of our test statistic in a finite sample and examine the empirical level and power of the corresponding model selection tests in various settings. Finally, an analysis of a financial dataset illustrates the proposed model selection procedure at work.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

In econometrics and statistics, the problem of model selection is usually tackled by likelihood-based comparisons. The most famous criteria are the AIC and the BIC proposed by Akaike (1998) and Schwarz (1978), respectively. They are based on the Kullback–Leibler Information Criterion (KLIC) (Kullback and Leibler, 1951), which measures a pseudo-distance between a proposed density and the true density of the observations. Both criteria penalize (minus) the sample log-likelihood by the number of estimated parameters. Among several candidates, the model with the lowest AIC or BIC is selected. Both criteria share nice properties as "dimension-consistency" and asymptotic optimality (see Burnham and Anderson (2002), Section 6). However, neither approach provides any information about the statistical significance of the choice between two or more models.

This drawback of AIC and BIC criteria can be mitigated by some tests for model comparison, following the seminal paper Vuong (1989), which was also based on the KLIC. Vuong (1989) derived the asymptotic distribution of the log-likelihood ratio of two competing densities, whose parameters are estimated in a preliminary stage. Based on this result, it is possible to formally test whether two competing families of densities are equally well approximating the unknown density of the observations. The Vuong test is applicable to overlapping or non-nested model comparisons, meaning that the competing families share some common densities or not, respectively. It has been applied in many empirical works due to its simple interpretation and implementation (see e.g. Fafchamps (1993), Bonnal et al. (1997), Cameron and Heckman

---

(1998), Caballero and Engel (1999), Nyarko and Schotter (2002), Karaivanov and Townsend (2014) and Kendall et al. (2015), among many others).

The seminal paper of Vuong (1989) has induced many contributions and extensions. In particular and without being exhaustive, Lavergne and Vuong (1996) studied a MSE-based testing procedure for discriminating between two sets of regressors. Rivers and Vuong (2002) extended Vuong (1989) to deal with weakly dependent observations, allowing for a wide variety of estimation techniques besides the KLIC. Li (2009) falls within this framework using a simulated mean squared error of predictions as another lack-of-fit criterion. In a semiparametric model, Chen and Fan (2005) used likelihood ratios to discriminate between two parametric copula families. Chen et al. (2007) proposed a nonparametric likelihood ratio testing procedure to choose between a parametric model and a moment condition model when both models could be misspecified. The relation between Bayesian forecast averaging and frequentist model selection and prediction has been studied in Hong and Preston (2012). In Vuong (1989), the difference between the KLIC of two competing models exhibits a variety of limiting distributions depending on whether the two models are overlapping or not. Therefore, using the KLIC typically requires pre-testing to decide which distribution should be used for the computation of critical values. Shi (2015b) showed that this could generate severe size distortions in finite samples. She proposed a one-step "nondegenerate" test that achieves uniform asymptotic size control in the overlapping and the non-overlapping cases.

Clarke (2007) combined the ideas of the Vuong test and the paired sign test: the selection between two non-nested parametric models is based on the median of the log-ratio of the two competing densities. Clarke's intuition is that the better model will yield higher likelihoods than the worse, on average and under the true underlying DGP. Then, the Clarke test statistic simply counts the number of likelihood ratios greater than one among all observations. Therefore, this statistic is claimed to be Binomial distributed, a trap one can easily fall into. A nice feature of the Clarke test is its wide range of applicability (the median of any distribution always exists) and a (falsely assumed) known law under the null. However, Clarke (2007) lacks mathematical rigor. In particular, the assessed law of the Clarke test statistic under the null is wrong due to the influence of first-stage estimators (see Appendix B).

Nonetheless, the Clarke test has been extensively used and invoked in the academic literature. Besides statistics (Panagiotelis et al., 2012), this test has been applied in many different fields: political science (Imai and Tingley, 2012; Martin and Vanberg, 2014), economics (Harrison et al., 2015; Czajkowski et al., 2009), finance (Agarwal and Taffler, 2008; Markwat et al., 2009), financial econometrics (Czado et al., 2012), actuarial science (Erhardt and Czado, 2012), psychometrics (New and Grainger, 2011; Douven and Schupbach, 2015), accounting (Barth et al., 2012; Verbeeten and Vijn, 2010), medicine (Hofbauer et al., 2015; Zhou et al., 2016), etc. For instance, in the context of decision-making under risk, Hey and Panaccione (2011) implemented an experimental design in which subjects are asked to take two sequential decisions concerning the allocation of a given sum of money. Individuals are then classified as resolute, sophisticated, myopic or naive through an intensive use of the Clarke test (see Table 5 in Hey and Panaccione (2011)).

The empirical analyses with the Clarke test usually deal with a comparison of either regression-type models with a specified log-likelihood or parametric distributions. Many applications simultaneously use the Clarke and Vuong tests and the reported p-values of the former are usually significantly lower than those of the latter. As a consequence, contradictory empirical assessments may easily be drawn based on the two tests: e.g. see Koliai (2016) or Panagiotelis et al. (2012). We will explain the source of such possible inconsistencies and our empirical analysis will illustrate this issue once more. Currently, the seminal paper Clarke (2007) is cited in roughly 300 academic works and new papers regularly use it incorrectly. Therefore, there is an urgent need to rectify its use under a theoretical and practical point of view.

In this paper, we propose a family of general model selection tests, which can be applied to many different settings. The Vuong and Clarke tests for strictly locally non-nested models (see Definition 4.1 in Section 4) are covered by the proposed framework. Using techniques from the empirical process theory, we show that a properly normalized test statistic is asymptotically normal. As a corollary, we revisit the original Clarke test and provide its corrected asymptotic distribution under the null hypothesis. Since its asymptotic variance is usually not available in closed form, we provide a weakly consistent estimator based on estimated numerical derivatives. Additionally, we propose an exchangeable bootstrap scheme to evaluate the limiting law of our general test statistic and its respective asymptotic variance.

The paper is organized as follows. In Section 2, we introduce a general framework for model selection with a possibly non-smooth criterion function and also motivate the original Clarke test. Moreover, we discuss several potential applications of the proposed model selection methodology in econometrics. Section 3 discusses the asymptotic normality of a test statistic for the general model selection procedure and provides a consistent numerical estimator for its unknown limiting variance. In Section 4, we deduce the asymptotic normality of the corrected Clarke test for strictly locally non-nested models. Section 5 discusses a modified Clarke-type test approach when the optimal competing models may coincide partially or completely. The validity of a general exchangeable bootstrap procedure is stated in Section 6. Additionally, we prove that the empirical variance of bootstrap replications of the test statistic also constitutes a consistent estimator of the unknown asymptotic variance. In Section 7, a simulation study investigates the finite sample performance of the test statistic in linear and quantile regression settings. Additionally, a short empirical investigation of a modified Clarke test for identical optimal models is provided. Section 8 provides a real data application of the corrected Clarke test and compares it to the original test by Clarke (2007). Section 9 summarizes and concludes the paper. The proofs have been postponed to Appendix A. In Appendix B, an illustrative example shows that the original Clarke test statistic is generally not Binomial distributed.

2. The Vuong test, the Clarke test and beyond

Let us introduce our mathematical framework for model selection, define the main test statistic of interest and set some notations. Consider a probability space $(\Omega, \mathcal{A}, P)$ and a random vector $X : \Omega \to \mathbb{R}^d$. Expectations $\mathbb{E}[\cdot]$ refer to the law of X, by default. We cover the case of continuous, discrete or mixed distributions. In many applications the vector X can be split into two subvectors Y, the "explained" variable, and Z, the "explanatory" variable (also called "covariates"). The conditional distribution of Y given Z induces a law $P_{Y|Z}$ on $\mathbb{R}^{d'}$ with $d' \leq d$, which becomes the topic for modeling purposes when the marginal law of the covariates Z is assumed to be the same for all models.

The paradigm of our model selection can be summarized as follows: given two competing model specifications, which one is best-suited to the true Data Generating Process (DGP)? Both families, only one, or neither are allowed to be correctly specified or misspecified. Formally, consider two families of models $\mathcal{F} := \{f_\alpha : \mathbb{R}^d \to \mathbb{R}^{d_\mathcal{F}} \mid \alpha \in \Theta_\alpha\}$ and $\mathcal{G} := \{g_\beta : \mathbb{R}^d \to \mathbb{R}^{d_\mathcal{G}} \mid \beta \in \Theta_\beta\}$, respectively, for two compact parameter sets $\Theta_\alpha \subset \mathbb{R}^{d_\alpha}$ and $\Theta_\beta \subset \mathbb{R}^{d_\beta}$. Typically, the latter maps are Radon–Nikodym densities of X (or of $Y \mid Z$) w.r.t. a common given dominating measure μ. Nonetheless, nothing precludes from considering regressions of Y on Z, conditional quantiles or others.

Further, assume that there exist some "pseudo-true" values (see Sawa (1978) and Monfort (1996)) $\alpha^\star$ and $\beta^\star$ that belong to the interior of Θ_α and Θ_β, respectively, and sequences of estimators $(\hat{\alpha}_n)_{n\in\mathbb{N}}$ and $(\hat{\beta}_n)_{n\in\mathbb{N}}$ such that $\hat{\alpha}_n = \alpha^\star + o_P(1)$ and $\hat{\beta}_n = \beta^\star + o_P(1)$. In general, $\alpha^\star$ and/or $\beta^\star$ are the parameter values which minimize a criterion (a divergence, or even a true distance) between f_α and/or g_β and a corresponding map given the true DGP. For instance, when $\mathcal{F}$ and $\mathcal{G}$ are sets of densities, $\alpha^\star$ (resp. $\beta^\star$) may be defined as the minimizer of $\mathbb{E}[\ln f_\alpha(X)]$ (resp. $\mathbb{E}[\ln g_\beta(X)]$), i.e. the minimizer of the Kullback–Leibler distance between the first (resp. second) density model and the true DGP. Nonetheless, it is not necessary to be restricted to such definitions in the considered framework. In other words, we do not assume anything on $\alpha^\star$ and $\beta^\star$, except the following zero assumption:

$$\mathcal{H}_0 : \mathbb{E}\Big[H\big(f_{\alpha^\star}(X), g_{\beta^\star}(X), X\big)\Big] = 0, \tag{1}$$

where $H : \mathbb{R}^{d_\mathcal{F}} \times \mathbb{R}^{d_\mathcal{G}} \times \mathbb{R}^d \to \mathbb{R}$ is a known measurable map. Most often, $\mathbb{E}[H(f(X), g(X), X)]$ is interpreted as a "distance" between f and g even if it is not required that the latter quantity has to satisfy the corresponding mathematical definition. Under $\mathcal{H}_0$, one considers the models $\mathcal{F}$ and $\mathcal{G}$ as equally well-suited given the true DGP. The third argument of H increases the range of potential applications. In particular, it allows the use of different "weights" depending on the location of X. Moreover, we cover the case of non-differentiable maps H, traditionally a source of technical difficulties.

In the literature, it is customary to distinguish between several situations: two models $\mathcal{F}$ and $\mathcal{G}$ are non-nested when $\mathcal{F} \cap \mathcal{G} = \emptyset$, overlapping when $\mathcal{F} \cap \mathcal{G} \neq \emptyset$ and nested when $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$. In this paper, we consider two slightly different situations: the case of strictly locally non-nested competing models when (α, β) belongs to a neighborhood of $(\alpha^\star, \beta^\star)$ (see Definition 4.1 in Section 4), or the case $P\big(f_{\alpha^\star}(X) = g_{\beta^\star}(X)\big) > 0$. When the two competing models $\mathcal{F}$ and $\mathcal{G}$ are strictly locally non-nested, a test of $\mathcal{H}_0$ is typically based on the test statistic

$$\hat{C}_n := \frac{1}{n} \sum_{i=1}^n H\big(f_{\hat{\alpha}_n}(X_i), g_{\hat{\beta}_n}(X_i), X_i\big), \tag{2}$$

where $(X_i)_{i\in\mathbb{N}}$ denotes an i.i.d. sample from X and $\hat{\alpha}_n$ and $\hat{\beta}_n$ denote the estimators of $\alpha^\star$ and $\beta^\star$, respectively.

In econometrics, the most common case by far is to assume that $f_\alpha(\cdot)$ and $g_\beta(\cdot)$ are some candidate densities of X, parameterized by α and β, respectively. If $H(f(\cdot), g(\cdot), \cdot) = \log(f(\cdot)/g(\cdot))$, then the two competing models will be considered as equivalent (i.e. $\mathcal{H}_0$ is satisfied) when their Kullback–Leibler divergences from the true law of X are the same. When allowing for covariates Z whose law does not depend on the parameters, this is equivalent to consider $f_\alpha(\cdot \mid z)$ and $g_\beta(\cdot \mid z)$, the candidate densities of Y given $Z = z$. We obtain the null hypothesis

$$\mathcal{H}_0^V : \quad \mathbb{E}\big[\log\big(f_{\alpha^\star}(X)\big)\big] = \mathbb{E}\big[\log\big(g_{\beta^\star}(X)\big)\big].$$

In the case of covariates, we have written $f_\alpha(X)$ instead of $f_\alpha(Y \mid Z)$, since $f_\alpha(Y \mid Z)$ can always be viewed as a function of X. Obviously, when $\mathbb{E}\big[H(f_{\alpha^\star}(X), g_{\beta^\star}(X), X)\big] > 0$, the first model is preferred over the second one, and inversely when $\mathbb{E}\big[H(f_{\alpha^\star}(X), g_{\beta^\star}(X), X)\big] < 0$. In case $\hat{\alpha}_n$ and $\hat{\beta}_n$ are pseudo-maximum likelihood estimators, $\mathcal{H}_0^V$ is the null hypothesis of the popular test of Vuong (1989). When dealing with non-nested models its test statistic

$$n^{-1/2} \sum_{i=1}^n \{\log\big(f_{\hat{\alpha}_n}(X_i)\big) - \log\big(g_{\hat{\beta}_n}(X_i)\big)\}$$

is asymptotically normal under $\mathcal{H}_0^V$.

In the presence of covariates, the same result applies. In the case of overlapping models, the latter asymptotic variance may be equal to zero, precluding the building of usual confidence intervals. That is why Vuong (1989) promotes a two-stage testing procedure with a pre-test of the zero assumption $\mathcal{H}_{00}^V : f_{\alpha^\star}(\cdot) = g_{\beta^\star}(\cdot)$.

To avoid confusion between the Vuong tests for non-nested and overlapping models we refer to the one-(resp. two-) step Vuong test when a test of $\mathcal{H}_0^V$ is conducted without (resp. with) a pre-test of $\mathcal{H}_{00}^V$. The pre-testing procedure to

manage overlapping situations can be seen as a weakness, due to a loss of power and practical inconvenience. To solve it, some authors have tackled the case of non-nested and overlapping models simultaneously. Notably, Shi (2015b) and Liao and Shi (2020) add higher-order adjustments to the numerator and/or the denominator of the Vuong Studentized log-likelihood ratio. By cleverly weighting individual log-likelihoods, Schennach and Wilhelm (2017) obtained another pivotal test statistic. Hsu and Shi (2017) and Shi (2015a) use similar techniques to manage the case of conditional moment identities and/or inequalities. Liao and Shi (2020) extended the conceptual ideas of Shi (2015b) to semi/non-parametric models. Due to the additional amount of randomness induced by simulation techniques, the test statistic in Li (2009) is never degenerate and allows to choose among competing structural models.

Generally speaking, all Vuong-type tests try to check $\mathcal{H}_0 : D(P, P_{\alpha^\star}) = D(P, P_{\beta^\star})$, for some pseudo-distance D, as noticed by many authors (see Li (2009), Liao and Shi (2020), among others). This particular case of $\mathcal{H}_0$ is obtained by comparing a measure of discrepancy between each competing model and the true underlying distribution. One can say that such procedures "separate" the features of the two models, by dealing with them independently. In our notation, separability means that one can write $H(f(\cdot), g(\cdot), \cdot) = H_1(f(\cdot)) - H_2(g(\cdot))$ for two measurable maps H_1 and H_2.

The previous situation of separability is more the exception than the rule. For instance, Clarke (2007) proposed to choose $H(f(\cdot), g(\cdot), \cdot) = \mathbf{1}\{\log(f(\cdot)/g(\cdot)) > 0\} - 1/2$, yielding a non-separable criterion. As in Vuong (1989), the maps f_α and g_β are densities, assuming that the competing models are non-nested. In this case, $\mathcal{H}_0$ translates to Clarke's null hypothesis[2]

$$\mathcal{H}_0^C : \ P\left(\log\left(\frac{f_{\alpha^\star}(X)}{g_{\beta^\star}(X)}\right) > 0\right) = \frac{1}{2} \cdot \tag{3}$$

The "separability feature" of Vuong-type tests cannot be exploited here. Instead of comparing two models in terms of their KLIC, Clarke proposed a "distribution-free" test statistic by considering the median of $\log(f_\alpha(Y \mid Z)) - \log(g_\beta(Y \mid Z))$. Then, under Clarke's approach, both models $\mathcal{F}$ and $\mathcal{G}$ are equally suited to describe the distribution of X if the median of $\log(f_{\alpha^\star}(X)) - \log(g_{\beta^\star}(X))$ is equal to 0. Note that the considered function H is a discontinuous function due to its jump at zero, which induces a technical hurdle to manage the situation $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Under such circumstances, it is impossible to control the regularity of the map $(\alpha, \beta) \mapsto P(\log(f_\alpha(X)/g_\beta(X)) > 0)$ in a neighborhood of $(\alpha^\star, \beta^\star)$ without too restrictive assumptions.

To test $\mathcal{H}_0^C$, Clarke proposed the test statistic

$$\hat{B}_n := \sum_{i=1}^n \mathbf{1}\left\{\log\left(\frac{f_{\hat{\alpha}_n}(X_i)}{g_{\hat{\beta}_n}(X_i)}\right) > 0\right\} = \sum_{i=1}^n \mathbf{1}\left\{\log\left(\frac{f_{\hat{\alpha}_n}(Y_i \mid Z_i)}{g_{\hat{\beta}_n}(Y_i \mid Z_i)}\right) > 0\right\},$$

and he claimed that it is Binomial distributed. The theoretical counterpart

$$B_n := \sum_{i=1}^n \mathbf{1}\left\{\log\left(\frac{f_{\alpha^\star}(X_i)}{g_{\alpha^\star}(X_i)}\right) > 0\right\} = \sum_{i=1}^n \mathbf{1}\left\{\log\left(\frac{f_{\alpha^\star}(Y_i \mid Z_i)}{g_{\beta^\star}(Y_i \mid Z_i)}\right) > 0\right\}$$

of $\hat{B}_n$ is indeed Binomial distributed. However, as it is illustrated in Appendix B, $\hat{B}_n$ is generally not Binomial distributed, neither for fixed n nor asymptotically, due to the additional randomness and dependence induced by the estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$. Necessary modifications of some asymptotic laws due to first-stage estimators have been pointed out for a long time in the literature (see Durbin (1973), e.g.). Apparently and surprisingly, this erroneous statement on the limiting law of the Clarke test has never been noticed in the literature.

In this paper, we provide the corrected asymptotic distribution of a normalized Clarke test statistic in the case of strictly locally non-nested models. If $f_{\alpha^\star} = g_{\beta^\star}$, i.e. the two "optimal models" are identical, $\mathcal{H}_0^C$ is not satisfied. This may be considered as a drawback of the Clarke approach. To make progress on this problematic case, we propose to test $\mathcal{H}_0^C$ with a modified null hypothesis — see Section 5, or to test $\mathcal{H}_0$ given in (1) for convenient maps H, which are smoother than $(f, g) \mapsto \mathbf{1}\{\log(f/g) > 0\}$ and may allow to handle the case $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Indeed, assume that H in (1) is "sufficiently regular" and satisfies $H(f, f, \cdot) = 0$. Then, $\mathcal{H}_0$ may be satisfied for some strictly locally non-nested models and for some models with $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. A typical choice for such a smooth function H is given by $H(f, g, \cdot) = \Psi(f - g) - 1/2$, where Ψ denotes some distribution function which satisfies $\Psi(0) = 1/2$. For example, set $\Psi(t) = \Phi(t/\chi)$, with a sufficiently small tuning parameter $\chi > 0$, where Φ denotes the cumulative distribution function of the standard normal distribution. Compared to (3), this means softening a non-differentiable discontinuous map, as usual in Machine Learning theory. Actually, both cases are complementary and will be managed in the next sections. To be specific, we study the asymptotic law of our general test statistic defined in (2) which covers both aforementioned situations. We state the limiting law of $\sqrt{n}\,\hat{C}_n$ under $\mathcal{H}_0$. In particular, setting $H(f(\cdot), g(\cdot), \cdot) = \mathbf{1}\{f(\cdot) > g(\cdot)\} - 1/2$, we recover Clarke's test statistic $\hat{B}_n$.

Let us illustrate the relevance of the original Clarke null hypothesis $\mathcal{H}_0^C$ and other Clarke-type null hypotheses in econometrics. We focus on the case of "no-separability" which cannot be managed by Vuong-type tests.

---

[2] Since f_α and g_β do not need to have the same support, it can occur that $\log(f_\alpha(X)/g_\beta(X))$ is not well-defined when one or both densities are zero. In such cases, formally set $\log(f_\alpha(X)/0) := +\infty$, $\log(0/g_\beta(X)) := -\infty$ and $\log(0/0) := 0$. According to these definitions $\mathcal{H}_0^C$ can be rewritten as $H_0^C : \ P(f_{\alpha^\star}(X) > g_{\beta^\star}(X)) = 1/2$.

- Linear regression: Assume two competing linear regression models

$$Y = r_{\alpha^\star}(Z) + \varepsilon_{\alpha^\star}, \ \mathbb{E}[\varepsilon_{\alpha^\star} \mid Z] = 0, \ \text{or} \ Y = \bar{r}_{\beta^\star}(Z) + \bar{\varepsilon}_{\beta^\star}, \ \mathbb{E}[\bar{\varepsilon}_{\beta^\star} \mid Z] = 0.$$

The traditional way of discriminating between such models is to compare the variances of their residuals. The two models would then be similar if $\mathbb{E}[\varepsilon_{\alpha^\star}^2] = \mathbb{E}[\bar{\varepsilon}_{\beta^\star}^2]$. In line with $\mathcal{H}_0^C$, our zero assumption would rather be $\mathcal{H}_0 : \mathbb{P}(|\varepsilon_{\alpha^\star}| > |\bar{\varepsilon}_{\beta^\star}|) = 1/2$, which allows to test for structural differences in the error terms of both models. Clearly, a violation of $\mathcal{H}_0$ would imply that the prediction error of one model tends to be smaller than the prediction error of the other model. In contrast to standard variance comparisons, which only involve the marginal second moment of $(\varepsilon_{\alpha^\star}, \bar{\varepsilon}_{\beta^\star})$, $\mathcal{H}_0$ accounts for the joint distribution of the error terms $(\varepsilon_{\alpha^\star}, \bar{\varepsilon}_{\beta^\star})$, independently of their moments. Moreover, $\mathcal{H}_0$ is insensitive to rare occurrences of large values (outliers) of $\varepsilon_{\alpha^\star}$, resp. $\bar{\varepsilon}_{\beta^\star}$, which usually have a large impact on the magnitude of their second moments.
Setting $f_\alpha(X) = Y - r_\alpha(Z)$ and $g_\beta(X) = Y - \bar{r}_\beta(Z)$ the most general case of $\mathcal{H}_0$ is given by

$$\mathbb{E}\Big[H\big(Y - r_{\alpha^\star}(Z), Y - \bar{r}_{\beta^\star}(Z), (Y, Z)\big)\Big] = 0,$$

for some map H such that $H(f, f, \cdot) = 0$ for every real f. In particular, such tests would provide a tool for selecting a subset of relevant explanatory variables.

- Quantile regression: The two competing models aim to predict the conditional quantiles of a random variable $Y \mid Z$, where Y given $Z = z$ is assumed to be a continuous random variable for every z. For a given quantile level $\tau \in (0, 1)$ they are defined as the function $q(z)$ such that $\mathbb{P}(Y \leq q(z, \alpha^\star) \mid Z = z) = \tau$ for almost every z. Define the loss function $\rho_\tau(t) = t(\tau - \mathbf{1}(t \leq 0))$. The two competing models $q_1(z, \alpha)_{\alpha \in \Theta_\alpha}$ and $q_2(z, \beta)_{\beta \in \Theta_\beta}$ seek to minimize their corresponding expected loss given by $\mathbb{E}[\rho_\tau(Y - q_i(Z, \cdot))]$, $i = 1, 2$. Let $\alpha^\star$ and $\beta^\star$ denote the respective minimizers of this criterion. Then, similarly to linear regression and in the same spirit as Clarke's $\mathcal{H}_0^C$, it would make sense to test

$$P\Big(\rho_\tau\big(Y - q_1(Z, \alpha^\star)\big) > \rho_\tau\big(Y - q_2(Z, \beta^\star)\big)\Big) = 1/2 \tag{4}$$

for structural differences in the error terms, or even to test

$$P\big(q_1(Z, \alpha^\star) > q_2(Z, \beta^\star)\big) = 1/2 \tag{5}$$

for structural differences in the predicted quantiles. Nonetheless, when the former assumption is not satisfied, this does not provide a way of choosing among the two models $\mathcal{F}$ and $\mathcal{G}$. In other terms, (4) is more relevant than (5) for model selection purposes. The particular case $\tau = 1/2$ yields the Least Absolute Deviation estimator for which $\rho_{1/2}(t) = |t|$ and we would test

$$P\big(|Y - q_1(Z, \alpha^\star)| > |Y - q_2(Z, \beta^\star)|\big) = 1/2.$$

Model selection in quantile regression is usually based on some distorted first moment of predicted absolute errors (check functions, as in Liao and Shi (2020) e.g.). On the other side, model selection procedures based on Clarke-type null hypotheses have a similar interpretation as in the linear regression setting. Particularly, they involve the joint distribution of the predicted absolute errors, are independent of the moments of the error term, and are less sensitive to outliers.
As before, the general null hypothesis $\mathcal{H}_0$ may be rewritten as

$$\mathbb{E}\Big[H\big(Y - q_1(Z, \alpha^\star), Y - q_2(Z, \beta^\star), (Y, Z)\big)\Big] = 0.$$

Finally, choosing $H(f(\cdot), g(\cdot), \cdot) = H_0(f(\cdot), g(\cdot))\mathbf{1}(\cdot \in \mathcal{C})$ for some map H_0 and some appropriate Borelian subset $\mathcal{C}$ could provide a way of choosing between two measures of systemic risk as CoVaR (see Adrian and Brunnermeier (2016)).

- Conditional expectiles: For any $\tau \in (0, 1)$, the τ-th expectile of Y given $Z = z$ (Newey and Powell, 1987) is a real number $e(\tau, z)$ that minimizes $\mathbb{E}[\tilde{\rho}_\tau(Y - m) \mid Z = z]$ over m for almost all z. Here, $\tilde{\rho}_\tau(t) := t^2(\tau - \mathbf{1}(t \leq 0))$, justifying the terminology "asymmetric least squares". Nonetheless, more general maps $\tilde{\rho}_\tau(t) := \psi(t)(\tau - \mathbf{1}(t \leq 0))$ have been proposed too (Breckling and Chambers, 1988). As for quantile regressions, consider two competing parametric models for conditional τ-expectiles, say $f_\alpha(z)$ and $g_\beta(z)$. The performances of the two latter models can be considered as similar if

$$P\Big(\tilde{\rho}_\tau\big(Y - f_{\alpha^\star}(Z)\big) > \tilde{\rho}_\tau\big(Y - g_{\beta^\star}(Z)\big)\Big) = 1/2.$$

- Limited-dependent variables: Assume that $Y \in \{0, 1\}$ is discrete. Typically, this arises in usual binary classification problems (Logit-Probit/SVM/trees, etc.). Denote $f_\alpha(X)$ and $g_\beta(X)$ as the likelihoods of $Y \mid Z$ under both model specifications: with a slight abuse of notation set

$$f_\alpha(X) := P(Y = 1 \mid Z, \alpha)^Y P(Y = 0 \mid Z, \alpha)^{1-Y}, \ \text{and}$$

$$g_\beta(X) := P(Y = 1 \mid Z, \beta)^Y P(Y = 0 \mid Z, \beta)^{1-Y}.$$

Note that

$$P\left(\log\left(\frac{f_{\alpha^\star}(X)}{g_{\beta^\star}(X)}\right) > 0\right) = \mathbb{E}_Z\Big[\mathbf{1}\{P(Y = 1 \mid Z, \alpha^\star) > P(Y = 1 \mid Z, \beta^\star)\}P(Y = 1 \mid Z)$$

$$+ \mathbf{1}\{P(Y = 0 \mid Z, \alpha^\star) > P(Y = 0 \mid Z, \beta^\star)\}P(Y = 0 \mid Z)\Big].$$

Clearly, $P\left(\log\left(f_{\alpha^\star}(X)/g_{\beta^\star}(X)\right) > 0\right) > 1/2$ means that the model $\mathcal{F}$ makes better predictions than the model $\mathcal{G}$ on average. Therefore, it makes sense to consider the two families of classifiers $\mathcal{F}$ and $\mathcal{G}$ as equally well suited classifiers if Clarke's null hypothesis $\mathcal{H}_0^C : P\left(\log\left(f_{\alpha^\star}(X)/g_{\beta^\star}(X)\right) > 0\right) = 1/2$ is satisfied. Alternatively, the expected misclassification rate (EMR) under $\mathcal{F}$ is given by

$$EMR(\mathcal{F}) := \mathbb{E}\Big[(1 - Y)\mathbf{1}\{P(Y = 1 \mid Z, \alpha^\star) > P(Y = 0 \mid Z, \alpha^\star)\}$$

$$+ Y\mathbf{1}\{P(Y = 1 \mid Z, \alpha^\star) < P(Y = 0 \mid Z, \alpha^\star)\}\Big]. \tag{6}$$

Thus, it would make sense to test $\mathcal{H}_0 : EMR(\mathcal{F}) = EMR(\mathcal{G})$, where $EMR(\mathcal{G})$ is defined similarly. In other words, both model specifications will be considered as equivalent if their expected proportions of misclassified outcomes are similar. Note that most binary classifiers are based on a rule such as: "$Y = 1$ iff $h(Z \mid \theta) \geq 0$ for some known parametric family h and some finite dimensional parameter θ". Therefore, a more general version of (6) would be $EMR(\mathcal{F}) := \mathbb{E}[f_{\alpha^\star}(X)]$, where $f_{\alpha^\star}(X) := (1 - Y)\mathbf{1}\{h(Z \mid \alpha^\star) \geq 0\} + Y\mathbf{1}\{h(Z \mid \alpha^\star) < 0\}$. Similarly, $EMR(\mathcal{G}) := \mathbb{E}[g_{\beta^\star}(X)]$, where $g_{\beta^\star}(X) := (1 - Y)\mathbf{1}\{\bar{h}(Z \mid \beta^\star) > 0\} + Y\mathbf{1}\{\bar{h}(Z \mid \beta^\star)\}$. All the latter points of view are particular cases of (1) with non-smooth H and we could conduct a test of $\mathcal{H}_0$ by setting $H\big(f_\alpha(\cdot), g_\beta(\cdot), \cdot\big) := f_\alpha(\cdot) - g_\beta(\cdot)$.

- Stochastic dominance: a cumulative distribution function F_1 first-order dominates another cumulative distribution function F_2 when $F_1(x) \leq F_2(x)$ for every x. If we consider that f_α and g_β are cumulative distribution functions, one says that the model $\mathcal{F}$ first-order dominates the model $\mathcal{G}$ if $P\big(f_{\alpha^\star}(X) \leq g_{\beta^\star}(X)\big) = 1$. This is a particular case of $\mathcal{H}_0$, with $H(f, g) = \mathbf{1}(f \leq g) - 1$. However, without modification, Theorem 3.1 only applies to tests of the null hypotheses $P\big(f_{\alpha^\star}(X) \leq g_{\beta^\star}(X)\big) = 1 - \epsilon$ for some small $\epsilon > 0$, which can be seen as an "approximated" first-order stochastic dominance property. Such ideas are close to the concept of "Almost Stochastic Dominance" (Leshno and Levy, 2002), even if the latter property cannot easily be written as (1). Alternatively, one could test the equivalent null hypothesis $\mathcal{H}_0 : \mathbb{E}\big[\Psi\big(\max\{0; f_{\alpha^\star}(X) - g_{\beta^\star}(X)\}\big)\big] = 0$, where Ψ denotes a sufficiently smooth distribution function with $\Psi(0) = 0$, for which it is likely that the assumptions of Theorem 3.1 are satisfied.

3. Limiting behaviors for an omnibus test

We first study our general-purpose statistic $\sqrt{n}\hat{C}_n$ defined in (2). If not explicitly stated otherwise (e.g. as in Theorem 3.1), we will always assume that the null hypothesis $\mathcal{H}_0$ from (1) is satisfied.

Following the usual notations in the field of empirical processes (e.g. see van der Vaart and Wellner (1996)), denote $Pf = \int_\Omega f(X(\omega))dP(\omega) = \mathbb{E}[f(X)]$ for any measurable function $f : \mathbb{R}^d \mapsto \mathbb{R}$. Let $(X_i)_{i\in\mathbb{N}}$ denote a sequence of i.i.d. realizations of the random vector X whose law is P. We refer to the law of the sequence $(X_i)_{i\in\mathbb{N}}$ by $\mathbb{P}$. The empirical measure associated with $(X_i)_{1\leq i\leq n}$ is denoted as $\mathbb{P}_n := n^{-1}\sum_{i=1}^n \delta_{X_i}$, where δ_{X_i} denotes the Dirac measure at X_i. Similarly to Pf, define $\mathbb{P}_n f := n^{-1}\sum_{i=1}^n f(X_i)$. In the proofs, we frequently use the expression $\mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f$. Further, $u \cdot v$ denotes the Euclidean scalar product of two vectors u and v. For any $\delta > 0$, denote $E_\delta := [\alpha_1^\star - \delta, \alpha_1^\star + \delta] \times \cdots \times [\alpha_{d_\alpha}^\star - \delta, \alpha_{d_\alpha}^\star + \delta] \times [\beta_1^\star - \delta, \beta_1^\star + \delta] \times \cdots \times [\beta_{d_\beta}^\star - \delta, \beta_{d_\beta}^\star + \delta]$ and

$$\phi(x, \alpha, \beta) := H\big(f(x, \alpha), g(x, \beta), x\big).$$

Moreover, for $\gamma > 0$ and $E_\gamma \subseteq \Theta_\alpha \times \Theta_\beta$, we denote $\mathcal{H}_\gamma := \{\phi(\cdot, \alpha, \beta) \mid (\alpha, \beta) \in E_\gamma\}$. Since $(\alpha^\star, \beta^\star)$ is assumed to be an interior point of $\Theta_\alpha \times \Theta_\beta$, such a $\gamma > 0$ always exists and $\mathcal{H}_\gamma$ is well-defined. Define the random function $\phi_n := H\big(f_{\hat{\alpha}_n}(\cdot), g_{\hat{\beta}_n}(\cdot), \cdot\big) = \phi(\cdot, \hat{\alpha}_n, \hat{\beta}_n)$ as well as $\phi_\star := H\big(f_{\alpha^\star}(\cdot), g_{\beta^\star}(\cdot), \cdot\big) = \phi(\cdot, \alpha^\star, \beta^\star)$. Finally, we use $\rightsquigarrow$ to denote weak convergence in $l^\infty(\mathcal{H}_\gamma) := \{\zeta \mid \zeta : \mathcal{H}_\gamma \mapsto \mathbb{R}; \sup_{f\in\mathcal{H}_\gamma} |\zeta(f)| < \infty\}$ equipped with the supremum norm, according to the standard theory of van der Vaart and Wellner (1996). The following assumptions are required to prove the asymptotic normality of $\sqrt{n}\hat{C}_n$.

Assumptions.

B1. $\alpha^\star$ and $\beta^\star$ belong to the interior of Θ_α and Θ_β, respectively.

B2. The (measurable) estimators satisfy $\sqrt{n}(\hat{\alpha}_n - \alpha^\star) = \mathbb{G}_n(s_1) + o_\mathbb{P}(1)$ and $\sqrt{n}(\hat{\beta}_n - \beta^\star) = \mathbb{G}_n(s_2) + o_\mathbb{P}(1)$ for some measurable functions s_1 and s_2 with $\mathbb{E}\big[s_1(X)^2\big] < \infty$ and $\mathbb{E}\big[s_2(X)^2\big] < \infty$.

B3. $\int (\phi_n - \phi_\star)^2 \, dP = \mathbb{E}_X \left[(\phi_n(X) - \phi_\star(X))^2 \right]$ tends to zero in probability.

B4. There exists some $\gamma > 0$ such that the function

$$h : E_\gamma \to \mathbb{R} \ ; \ (\alpha, \beta) \mapsto \int \phi(x, \alpha, \beta) \, dP(x) = \int H\big(f_\alpha(x), g_\beta(x), x\big) \, dP(x)$$

is continuously differentiable at $(\alpha^\star, \beta^\star)$.

Note that, by Dominated Convergence, Assumption B3 is satisfied when the map H is continuous and bounded, assuming B2 and strong consistency of the first-stage estimators. Otherwise, when H is discontinuous, it is necessary to directly check B3, which is the case for the Clarke test (see the proof of Corollary 4.2). Moreover, note that $\mathcal{H}_0$ simply means $h(\alpha^\star, \beta^\star) = 0$. Assumption B2 is true for many M- and Z-estimators: e.g. see Theorems 2.11–2.13 in Kosorok (2008). In other words, we do not restrict ourselves to ML or quasi-ML parameter estimators, contrary to Vuong (1989) and Clarke (2007). We denote

$$h_1(\alpha, \beta) := \left[\frac{\partial}{\partial \alpha_1}, \dots, \frac{\partial}{\partial \alpha_{d_\alpha}} \right]^\top h(\alpha, \beta), \ \text{and} \ h_2(\alpha, \beta) := \left[\frac{\partial}{\partial \beta_1}, \dots, \frac{\partial}{\partial \beta_{d_\beta}} \right]^\top h(\alpha, \beta),$$

where $x^\top$ denotes the transpose of the vector x. The first part of the next theorem shows the asymptotic normality of $\sqrt{n} \hat{C}_n$ under $\mathcal{H}_0$. Parts (ii) and (iii) handle the two possible cases under the alternative.

Theorem 3.1. *Assume that there exists $\gamma > 0$ such that $\mathcal{H}_\gamma$ is P-Donsker and Assumptions B1–B4 are satisfied. Then the following statements are valid.*

(i) Under $\mathcal{H}_0$, i.e. $\mathbb{E}\left[H\big(f_{\alpha^\star}(X), g_{\beta^\star}(X), X\big) \right] = 0$, we have $\sqrt{n} \hat{C}_n \rightsquigarrow \mathcal{N}(0, \sigma^2)$, where

$$\sigma^2 = \mathrm{Var}\big(\phi(X, \alpha^\star, \beta^\star) + h_1(\alpha^\star, \beta^\star) \cdot s_1(X) + h_2(\alpha^\star, \beta^\star) \cdot s_2(X)\big).$$

(ii) If $\mathbb{E}\left[H\big(f_{\alpha^\star}(X), g_{\beta^\star}(X), X\big) \right] < 0$, then $\sqrt{n} \hat{C}_n \to -\infty$ P-almost surely.

(iii) If $\mathbb{E}\left[H\big(f_{\alpha^\star}(X), g_{\beta^\star}(X), X\big) \right] > 0$, then $\sqrt{n} \hat{C}_n \to +\infty$ P-almost surely.

The proof of this theorem has been postponed to Appendix A.1. Under (ii), the optimal model from $\mathcal{G}$ should be preferred over the optimal model from $\mathcal{F}$, and the opposite under (iii). Note that we do not explicitly require any nested/non-nested/overlapping assumption concerning the two competing models. Nonetheless, such features would be relevant to verify (or falsify) the validity of the key Assumption B3 and B4 for a particular map H. Moreover, it may happen that $\phi_\star = h_1 = h_2 = 0$ which implies $\sigma = 0$. In such situations, Theorem 3.1 states convergence to a degenerate distribution and the corresponding testing procedures cannot be conducted. For example, the degenerate limiting distribution appears when $f_{\alpha^\star} = g_{\beta^\star}$ and one would falsely apply the one-step Vuong test for non-nested model selection. Moreover, as for the one-step Vuong test, a test of $\mathcal{H}_0$ based on $\sqrt{n} \hat{C}_n$ can suffer from finite sample size distortions when σ is close to zero, because Theorem 3.1 is not stated uniformly w.r.t. the unknown DGP nor the competing models. Some uniformly exact asymptotic size results of this type have been obtained by Shi (2015b), Schennach and Wilhelm (2017) and Liao and Shi (2020) in the framework of Vuong (1989) and are left for further research in the framework of Clarke (2007). For more details on the situation $f_{\alpha^\star} = g_{\beta^\star}$, we refer to the discussion around the Clarke test in Section 5.

In the latter theorem, we require that $\mathcal{H}_\gamma$ is P-Donsker, a standard assumption in empirical process theory. It can be explicitly verified in many cases: see Van der Vaart (2000) (Section 19) and Brück (2019) for a verification of this assumption for the corrected Clarke test. Essentially, it just means that the complexity of $\mathcal{H}_\gamma$ is not too high.

Considering $\mathbb{E}[H(f_{\alpha^\star}(X), g_{\beta^\star}(X), X)]$ as an unknown parameter C_∞ that is empirically estimated by $\hat{C}_n$, our inference strategy is exactly that of two-step parametric estimators: in a first stage, estimate $(\alpha^\star, \beta^\star)$ by $(\hat{\alpha}_n, \hat{\beta}_n)$; in a second stage, plug-in the latter estimates in an empirical moment formula to evaluate C_∞. The asymptotic theory of such two-step parametric estimators is well-known: see some textbooks as White (1996) (Section 6.3) or Wooldridge (2010) (Section 12.4). In particular, when the first-stage estimators $(\hat{\alpha}_n, \hat{\beta}_n)$ are based on some moment conditions, the limiting law of $\sqrt{n} \hat{C}_n$ is deduced from the usual theory of Z-estimators (Newey (1984), Murphy and Topel (1985), e.g.). Unfortunately, the latter results cannot be applied here because they require more restrictive regularity assumptions than ours. Indeed, the strength of Theorem 3.1 is due to the fact that we do not impose the differentiability of $(\alpha, \beta) \mapsto H\big(f_\alpha(x), g_\beta(x), x\big)$ for every x but rather of h, its expectation under P. In many cases h will be differentiable, even though H may not be differentiable, since integrating over a probability measure is similar to a smoothing procedure.

In practical applications, the asymptotic variance σ^2 of the test statistic in Theorem 3.1 is unknown. To build a consistent estimator of σ^2, we numerically estimate the partial derivatives of the function h. To this goal, let $e(n)$ denote some positive function and denote by u_j^α and u_j^β the jth unit vectors in $\mathbb{R}^{d_\alpha}$ and $\mathbb{R}^{d_\beta}$, respectively. As an estimator of h will use the map $h_n : E_\gamma \to \mathbb{R}, h_n(\alpha, \beta) := \int \phi(x, \alpha, \beta) \, d\mathbb{P}_n(x)$, and set

$$\hat{h}_{1,n,i} := \frac{1}{2e(n)} \big(h_n(\hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) - h_n(\hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n)\big), \ 1 \le i \le d_\alpha,$$

$$\hat{h}_{2,n,j} := \frac{1}{2e(n)}\big(h_n(\hat{\alpha}_n, \hat{\beta}_n + e(n)u_j^{\beta}) - h_n(\hat{\alpha}_n, \hat{\beta}_n - e(n)u_j^{\beta})\big), \ 1 \le j \le d_{\beta}.$$

With this notation, we can propose the following estimator of σ^2.

Theorem 3.2. *Consider a positive function $e(n)$ such that $\lim_{n\to\infty} e(n) = 0$ and $\lim_{n\to\infty} \sqrt{n}\,e(n) = +\infty$. Under the assumptions of Theorem 3.1 (or Corollary 4.2) and if $\sup_{(\alpha,\beta)\in E_{\gamma}} \mathbb{E}\left[H\big(f_{\alpha}(X), g_{\beta}(X), X\big)\right] < \infty$, the variance estimator*

$$\hat{\sigma}^2 := \mathbb{P}_n\left(\phi_n + \hat{h}_{1,n} \cdot s_1 + \hat{h}_{2,n} \cdot s_2 - \mathbb{P}_n\big(\phi_n + \hat{h}_{1,n} \cdot s_1 + \hat{h}_{2,n} \cdot s_2\big)\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(H\big(f_{\hat{\alpha}_n}(X_i), g_{\hat{\beta}_n}(X_i), X_i\big) + \hat{h}_{1,n} \cdot s_1(X_i) + \hat{h}_{2,n} \cdot s_2(X_i)\right)^2$$

$$- \left(\frac{1}{n}\sum_{i=1}^{n} H\big(f_{\hat{\alpha}_n}(X_i), g_{\hat{\beta}_n}(X_i), X_i\big) + \hat{h}_{1,n} \cdot s_1(X_i) + \hat{h}_{2,n} \cdot s_2(X_i)\right)^2$$

tends to σ^2 in probability.

Again, the proof is postponed to Appendix A.2. By Slutsky's theorem, $\sqrt{n}\,\hat{C}_n/\hat{\sigma}$ is asymptotically standard normally distributed under $\mathcal{H}_0$. Moreover, by inspecting the proof of Theorem 3.1, we see that Theorem 3.2 is also valid if the null hypothesis $\mathcal{H}_0$ is not satisfied. In the latter case, σ^2 is the asymptotic variance of

$$n^{-1/2}\sum_{i=1}^{n}\big\{H(f_{\hat{\alpha}_n}(X_i), g_{\hat{\beta}_n}(X_i), X_i) - \mathbb{E}\left[H(f_{\alpha^\star}(X), g_{\beta^\star}(X), X)\right]\big\}.$$

This result is important if we apply our test under the alternative, since it shows that the variance estimator $\hat{\sigma}^2$ converges to some finite real number and that the test statistic $\sqrt{n}\,\hat{C}_n/\hat{\sigma}$ converges to $+\infty$ or $-\infty$, depending on the sign of $\mathbb{E}\left[H(f_{\alpha^\star}(X), g_{\beta^\star}(X), X)\right]$.

Remark 1. The presented estimator $\hat{\sigma}^2$ depends on the tuning parameter $e(n)$ that has to be calibrated. In our empirical sections, we alternatively estimate the unknown variance σ^2 by a bootstrap procedure. The latter procedure does not require any tuning parameter but, on the other hand, is more computationally costly.

4. Application to the Clarke test and strictly locally non-nested models

With the help of Theorem 3.1, we are ready to state the corrected limiting law of the Clarke test statistic $\hat{B}_n$. As announced before, we need to assume strictly locally non-nested models in a neighborhood of the pseudo-true values. The following definition excludes the possibility of $f_{\alpha^\star}(X)/g_{\beta^\star}(X)$ having an atom at 1.

Definition 4.1. A couple of parametric families $\mathcal{F} = (f_{\alpha})_{\alpha\in\Theta_{\alpha}}$ and $\mathcal{G} = (g_{\beta})_{\beta\in\Theta_{\beta}}$ is strictly locally non-nested on $\overline{\Theta}_{\alpha} \times \overline{\Theta}_{\beta} \subseteq \Theta_{\alpha} \times \Theta_{\beta}$ with non-empty interior if $f_{\alpha}(X) \ne g_{\beta}(X)$ a.s. for all $\alpha \in \overline{\Theta}_{\alpha}$ and $\beta \in \overline{\Theta}_{\beta}$.

Note that the latter definition depends on the unknown law of the true DGP and is slightly stronger than the definition of non-nested models in Vuong (1989), since we do not allow that two rival models f_{α} and g_{β} coincide on a set with positive probability. This was not excluded in Vuong (1989), who only required that the functions $f_{\alpha}(\cdot \mid Z = z)$ and $g_{\beta}(\cdot \mid Z = z)$ are not equal for P_Z-almost all z and every $(\alpha, \beta) \in \Theta_{\alpha} \times \Theta_{\beta}$ ("strictly non-nested" models, Definition 2), which allows to verify his notion of non-nestedness by checking that $\mathcal{F} \cap \mathcal{G} = \emptyset$. On the other hand, the assumption of strictly locally non-nested models may be hard to validate empirically, since the set $\Theta_{\alpha=\beta}^{\delta} := \{x \in \mathbb{R}^d \mid \exists\, (\alpha, \beta) \in [\alpha^\star - \delta, \alpha^\star + \delta] \times [\beta^\star - \delta, \beta^\star + \delta]\ s.t.\ f_{\alpha}(x) = g_{\beta}(x)\}$ may theoretically have positive probability for every $\delta > 0$, even if the optimal models are strictly locally non-nested. Therefore, the possibility of $P(\Theta_{\alpha=\beta}^{\delta}) > 0$ suggests that a formal consistent test for strict local non-nestedness seems to be difficult, if not impossible.

We want to emphasize that overlapping families of densities, i.e. families for which $\mathcal{F} \cap \mathcal{G} \ne \emptyset$, may be strictly locally non-nested in the sense of Definition 4.1. Indeed, the competing families $\mathcal{F}$ and $\mathcal{G}$ can intersect even if they are strictly locally non-nested in a neighborhood of the optimal parameter $(\alpha^\star, \beta^\star)$. Actually, there are only two cases of theoretical interest for us: strictly locally non-nested models and models for which $P\big(f_{\alpha^\star}(X) = g_{\beta^\star}(X)\big) > 0$. In the former case, the Clarke test null hypothesis $\mathcal{H}_0^C$ can be tested. The latter situation is discussed in Section 5. Let us specify Assumption $B3$ in the particular case of the Clarke test.

Assumption.

B3'. There exists $\gamma > 0$ such that, for any $x \in Range(X)$, the functions $\alpha \mapsto f_{\alpha}(x)$ and $\beta \mapsto g_{\beta}(x)$ are continuous on E_{γ}. Moreover, the elements of $\mathcal{F}$ and $\mathcal{G}$ are Radon–Nikodym derivatives with respects to a common dominating measure and there exists some $\gamma > 0$ such that the models $\mathcal{F}$ and $\mathcal{G}$ are strictly locally non-nested on E_{γ}. Moreover, $\hat{\alpha}_n$ and $\hat{\beta}_n$ are strongly consistent.

Now, let us state the corrected asymptotic distribution of $n^{-1/2}(\hat{B}_n - n/2)$. To lighten notations, set

$$\psi(x, \alpha, \beta) := \log\big(f_\alpha(x)/g_\beta(x)\big),$$

and particularize the previous ϕ-related notations: $\phi^C(x, \alpha, \beta) := \mathbf{1}\big\{\psi(x, \alpha, \beta) > 0\big\}$, $\phi_n^C := \phi^C(X, \hat{\alpha}_n, \hat{\beta}_n)$, and $\phi_\star^C := \phi^C(X, \alpha^\star, \beta^\star)$.

Corollary 4.2. *Assume that there exists $\gamma > 0$ such that $\mathcal{H}_\gamma^C := \{\mathbf{1}\{\psi(\cdot, \alpha, \beta) > 0\} \mid (\alpha, \beta) \in E_\gamma\}$ is P-Donsker and that Assumptions B1, B2, B3′, B4 are satisfied. Then the following statements are valid.*

(i) Under $\mathcal{H}_0^C$, i.e. if $P(\psi(X, \alpha^\star, \beta^\star) > 0) = 1/2$, we have

$$\frac{1}{\sqrt{n}}\left(\hat{B}_n - \frac{n}{2}\right) \rightsquigarrow \mathcal{N}(0, \sigma_\psi^2), \tag{7}$$

where $\sigma_\psi^2 = \text{Var}\big(\mathbf{1}\{\psi(X, \alpha^\star, \beta^\star) > 0\} + h_1(\alpha^\star, \beta^\star) \cdot s_1(X) + h_2(\alpha^\star, \beta^\star) \cdot s_2(X)\big)$.

(ii) If $P(\psi(X, \alpha^\star, \beta^\star) > 0) < 1/2$, then $n^{-1/2}(\hat{B}_n - n/2) \to -\infty$ P-almost surely.

(iii) If $P(\psi(X, \alpha^\star, \beta^\star) > 0) > 1/2$, then $n^{-1/2}(\hat{B}_n - n/2) \to +\infty$ P-almost surely.

The proof of this corollary has been postponed to Appendix A.3. Note that, assuming strictly locally non-nested models, $\mathcal{H}_0^C$ can be rewritten in the spirit of (1) with

$$H(f_\alpha(x), g_\beta(x), x) := \mathbf{1}\{f_\alpha(x) \geq g_\beta(x)\} - \mathbf{1}\{f_\alpha(x) \leq g_\beta(x)\}, \tag{8}$$

yielding the so-called "modified Clarke null hypothesis":

$$\bar{\mathcal{H}}_0^C : P\big(f_{\alpha^\star}(X) \geq g_{\beta^\star}(X)\big) = P\big(f_{\alpha^\star}(X) \leq g_{\beta^\star}(X)\big). \tag{9}$$

Under $\bar{\mathcal{H}}_0^C$ (or $\mathcal{H}_0^C$, equivalently), a similar result as in Corollary 4.2 applies, replacing $\hat{B}_n - n/2$ by

$$\bar{B}_n := \frac{1}{2}\sum_{i=1}^n \Big(\mathbf{1}\{f_{\hat{\alpha}_n}(X_i) \geq g_{\hat{\beta}_n}(X_i)\} - \mathbf{1}\{f_{\hat{\alpha}_n}(X_i) \leq g_{\hat{\beta}_n}(X_i)\}\Big).$$

According to a counter-example from Appendix B, Clarke (2007) misleadingly stated that $\hat{B}_n$ is Binomial distributed. This erroneous statement implied that the asymptotic variance of $(\hat{B}_n - n/2)/\sqrt{n}$ is $\text{Var}\big(\mathbf{1}\{\psi(X, \alpha^\star, \beta^\star) > 0\}\big)$, which is equal to $1/4$ under $\mathcal{H}_0^C$. Corollary 4.2 rectifies this incorrect statement. Further, note that $f_\alpha(X)$ and $g_\beta(X)$ in Corollary 4.2 are general likelihoods associated to the observation X. This extends Clarke's initial framework, who only considered densities w.r.t. the Lebesgue measure. In particular, we now cover the case of limited-dependent variables. Moreover, from the proof of Theorem 3.1, we easily deduce the asymptotic "distance" from the Binomial distributed test statistic B_n to $\hat{B}_n$.

Corollary 4.3. *Under the assumptions of Corollary 4.2, we have*

$$\frac{1}{\sqrt{n}}\big(\hat{B}_n - B_n\big) \rightsquigarrow \mathcal{N}(0, \sigma_h^2), \text{ with } \sigma_h^2 := \text{Var}\big(h_1(\alpha^\star, \beta^\star) \cdot s_1(X) + h_2(\alpha^\star, \beta^\star) \cdot s_2(X)\big).$$

Corollary 4.3 tells us that the difference between the Clarke test statistic $\hat{B}_n$ and its Binomial distributed theoretical counterpart B_n, once properly normalized, converges to a non-degenerate continuous random variable if $h_1(\alpha^\star, \beta^\star) \cdot s_1(X) + h_2(\alpha^\star, \beta^\star) \cdot s_2(X)$ is not equal to 0. This is the statistical price to be paid for the estimation of the unknown quantities $\alpha^\star$ and $\beta^\star$. The calculation of s_1 and s_2 depends on the way $\alpha^\star$ and $\beta^\star$ are estimated. A standard choice is to use pseudo-maximum likelihood estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ i.e.

$$\hat{\alpha}_n = \underset{\alpha \in \Theta_\alpha}{\arg\max} \sum_{i=1}^n \log\big(f_\alpha(Y_i \mid Z_i)\big) \text{ and } \hat{\beta}_n = \underset{\beta \in \Theta_\beta}{\arg\max} \sum_{i=1}^n \log\big(g_\beta(Y_i \mid Z_i)\big).$$

Under some usual conditions of regularity (White (1982), Theorem 3.1), one obtains

$$s_1(X) := -Hess_f(\alpha^\star)^{-1}\frac{\partial \log\big(f_{\alpha^\star}(X)\big)}{\partial \alpha}, \ s_2(X) := -Hess_g(\beta^\star)^{-1}\frac{\partial \log\big(g_{\beta^\star}(X)\big)}{\partial \beta},$$

where $Hess_f(\alpha)$ and $Hess_g(\beta)$ are the Hessian matrices of the two competing models.

In practice, the asymptotic variance σ_ψ^2 from Corollary 4.2 is unknown. The asymptotic variance of the one-step Vuong test statistic could be consistently estimated by usual sample counterparts, i.e. by $\mathbb{P}_n\phi_n^2 - (\mathbb{P}_n\phi_n)^2$ in our notations (Vuong (1989), Equation (4.2)). This is due to an orthogonality property, coming from pseudo-MLE first-stage estimators. Since we consider general moment relationships and general estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$, we cannot hope to fulfill such orthogonality conditions in most cases. In the particular case of the Clarke test, $\phi_n^C = \big(\phi_n^C\big)^2$ and the empirical variance estimator would

be given by $\mathbb{P}_n\phi_n^C - (\mathbb{P}_n\phi_n^C)^2$, which tends to $1/4$ in probability. However, Example 2 in Brück (2019) shows that σ_ψ^2 is generally not equal to $1/4$, which can be explained by the additional noise that is generated by using $n/2$ instead of the unknown quantity $\mathbb{E}[\hat{B}_n] = nP(\psi(X, \hat{\alpha}_n, \hat{\beta}_n) > 0)$ to center $\hat{B}_n$ in (7). In particular, Theorem 3.2 implies that a feasible estimator of the asymptotic variance of the Clarke test $n^{-1/2}(\hat{B}_n - n/2)$ is given by

$$\hat{\sigma}_\psi^2 := \frac{1}{n}\sum_{i=1}^n \left(\mathbf{1}\{f_{\hat{\alpha}_n}(X_i) > g_{\hat{\beta}_n}(X_i)\} + \hat{h}_{1,n} \cdot s_1(X_i) + \hat{h}_{2,n} \cdot s_2(X_i)\right)^2$$

$$- \left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{f_{\hat{\alpha}_n}(X_i) > g_{\hat{\beta}_n}(X_i)\} + \hat{h}_{1,n} \cdot s_1(X_i) + \hat{h}_{2,n} \cdot s_2(X_i)\right)^2.$$

5. A modified Clarke test in case $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$

The formulation of Clarke's null hypothesis $\mathcal{H}_0^C$ in (3) is not suitable for models which satisfy $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Indeed, in such situations, it is possible that $P(f_{\alpha^\star}(X) < g_{\beta^\star}(X))$ and $P(f_{\alpha^\star}(X) > g_{\beta^\star}(X))$ are both strictly smaller than $1/2$. When the optimal models $f_{\alpha^\star}$ and $g_{\beta^\star}$ are identical (the extreme situation), $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) = 1$ and $\mathcal{H}_0^C$ would not be satisfied. This shows that the usual Clarke null hypothesis $\mathcal{H}_0^C$ is not well-suited under such circumstances.

To rectify this shortcoming of Clarke's null hypothesis, the "modified Clarke null hypothesis" $\bar{\mathcal{H}}_0^C$ (9) can be invoked. In the spirit of the Clarke test, $\bar{\mathcal{H}}_0^C$ naturally expresses that both competing models are equally well suited to describe the modeling problem, even if the competing models satisfy $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Moreover, $\bar{\mathcal{H}}_0^C$ and $\mathcal{H}_0^C$ are equivalent if the competing models are strictly locally non-nested. Thus, $\bar{\mathcal{H}}_0^C$ sensibly generalizes Clarke's null hypothesis to a meaningful null hypothesis for general model specifications. It is worth noting that overlapping models, i.e. $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, potentially satisfy $f_{\alpha^\star} = g_{\beta^\star}$ and the modified Clarke test should be invoked in practice.

It is tempting to directly apply Theorem 3.1 to the respective map H from (8) and use its empirical counterpart as a test statistic of $\bar{\mathcal{H}}_0^C$. In theory, Theorem 3.1 should provide a Gaussian limiting law even when $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Unfortunately, when the event $\{f_{\alpha^\star}(X) = g_{\beta^\star}(X)\}$ has a strictly positive probability under P, it seems impossible to satisfy the required regularity assumptions and to obtain a non-degenerate limiting law. The assumptions $B3$ and $B4$ are particularly problematic. To illustrate the technical difficulties assume that the underlying laws are continuous and that $f_{\alpha^\star}(\cdot) = g_{\beta^\star}(\cdot)$. Then, using the same notation as in Section 3, we have

$$\phi_n(\cdot) = \mathbf{1}(f_{\hat{\alpha}_n}(\cdot) \geq g_{\hat{\beta}_n}(\cdot)) - \mathbf{1}(f_{\hat{\alpha}_n}(\cdot) \leq g_{\hat{\beta}_n}(\cdot)), \text{ and}$$

$$\phi_\star(\cdot) = \mathbf{1}(f_{\alpha^\star}(\cdot) \geq g_{\beta^\star}(\cdot)) - \mathbf{1}(f_{\alpha^\star}(\cdot) \leq g_{\beta^\star}(\cdot)) = 0.$$

As a consequence, to check Assumption B3, we need to verify that

$$\mathbb{E}_X\left[(\phi_n - \phi_\star)^2(X)\right] = \mathbb{E}_X\left[\mathbf{1}\{f_{\hat{\alpha}_n}(X) > g_{\hat{\beta}_n}(X)\} + \mathbf{1}\{f_{\hat{\alpha}_n}(X) < g_{\hat{\beta}_n}(X)\}\right]$$

tends to zero in probability. However, if $P(f_{\hat{\alpha}_n}(X) = g_{\hat{\beta}_n}(X)) = 0$ – a standard situation – we deduce that $\mathbb{E}_X[(\phi_n - \phi_\star)^2(X)] = 1$, which shows that Assumption B3 cannot be satisfied in general. Moreover, for many simple models, the partial derivatives h_1 and h_2 from Assumption B4 do not exist. This can be easily checked with the Joint Normal Location model in Example 1 of Schennach and Wilhelm (2017).

More fundamentally, the original Clarke test perspective or even its modified version $\bar{\mathcal{H}}_0^C$ are not well-suited for models with $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$. Indeed, when $f_{\alpha^\star}$ and $g_{\beta^\star}$ are identical, focusing on the relative positions of $f_{\hat{\alpha}_n}(X)$ and $g_{\hat{\beta}_n}(X)$ is not very relevant due to their respective statistical uncertainties. It is easy to build some situations where simply comparing $f_{\hat{\alpha}_n}(X)$ and $g_{\hat{\beta}_n}(X)$ may lead to wrong conclusions. For instance, assume that the true DGP is a standard normal distribution and consider the competing models $\mathcal{F} = \{f_\alpha\} = \{f_{\alpha^\star}\} = \{\varphi\}$, where φ denotes the density of the standard normal distribution, and

$$\mathcal{G} := \left\{x \mapsto (1+\beta)\varphi(x)\mathbf{1}_{\{|x|>1\}} + (\varphi(x) - \beta\Phi(-1))\mathbf{1}_{\{|x|\leq 1\}}, \beta \in [-1/2, 1/2]\right\}.$$

It is easy to see that $\beta^\star = 0$, which implies that $f_{\alpha^\star} = g_{\beta^\star}$. Thus, $\hat{\alpha}_n = \alpha^\star$ for every n, $P(f_{\hat{\alpha}_n}(X) \geq g_{\hat{\beta}_n}(X)) = 2\Phi(-1)$ if $\hat{\beta}_n > 0$, $P(f_{\hat{\alpha}_n}(X) \geq g_{\hat{\beta}_n}(X)) = 1 - 2\Phi(-1)$ if $\hat{\beta}_n < 0$ and $P(f_{\hat{\alpha}_n}(X) = g_{\hat{\beta}_n}(X)) = 1$ if $\hat{\beta}_n = 0$. In other words, the likelihood of the events of interest for conducting a (modified or not) Clarke test strongly depends on the way the parameter $\hat{\beta}_n$ tends to zero, which cannot be controlled.

Therefore, to deal with the possibility of $P(f_{\alpha^\star}(X) = g_{\beta^\star}(X)) > 0$, we rather promote the use of a "smoothing trick", as introduced in Section 2. The idea is to replace the non-differentiable indicator functions appearing in the modified Clarke null hypothesis by differentiable approximations. The smoothing trick can account for the uncertainty of the event $\{f_{\alpha^\star}(X) = g_{\beta^\star}(X)\}$ when only looking at $f_{\hat{\alpha}_n}$ and $g_{\hat{\beta}_n}$ and by putting less emphasis on those observations which are not clearly in favor of either model. With the previous notations, set

$$H_\chi(f, g, x) = \Psi\left(\frac{f(x) - g(x)}{\chi}\right) - \Psi\left(\frac{g(x) - f(x)}{\chi}\right), \tag{10}$$

for some constant $\chi > 0$ and some increasing map $\Psi : \mathbb{R} \to \mathbb{R}$ such that $\Psi(0) = 1/2$, $\lim_{t \to -\infty} \Psi(t) = 0$ and $\lim_{t \to +\infty} \Psi(t) = 1$. A natural example is to set $\Psi = \Phi$, the cumulative distribution function of a standard normal distribution. The new null hypothesis is

$$\mathcal{H}_{0,\chi} : \mathbb{E}\big[H_\chi(f_{\alpha^\star}, g_{\beta^\star}, X)\big] = 0. \tag{11}$$

Now, under $\mathcal{H}_{0,\chi}$, Assumptions *B3* and *B4* will be satisfied in many situations and the associated test statistic $\sqrt{n}\,\hat{C}_n$ will be generally asymptotically normal by Theorem 3.1. The latter result is true for strictly locally non-nested models and for most models which satisfy $P\big(f_{\alpha^*}(X) = g_{\beta^*}(X)\big) > 0$, as soon as (11) is satisfied. Obviously, testing $\bar{\mathcal{H}}_0^C$ is not equivalent to testing $\mathcal{H}_{0,\chi}$ in general, except the optimal models are identical, i.e. $f_{\alpha^\star} = g_{\beta^\star}$. Indeed, particularly in the strictly locally non-nested case, it is possible to build some examples where $\bar{\mathcal{H}}_0^C$ is satisfied but $\mathcal{H}_{0,\chi}$ is not satisfied for some χ. Nonetheless, in every case, $\mathcal{H}_{0,\chi}$ and $\bar{\mathcal{H}}_0^C$ reflect similar ideas for comparing the two competing models. When χ is "small", $\mathcal{H}_{0,\chi}$ can be seen as an approximation of $\bar{\mathcal{H}}_0^C$ because H_χ pointwise approximates the discontinuous map in (9). Moreover, if $\bar{\mathcal{H}}_0^C$ is not satisfied, there exists a sequence (χ_n), $\chi_n \to 0$ such that $\mathcal{H}_{0,\chi_n}$ is never satisfied (otherwise, the dominated convergence theorem applies). Thus, a test of $\mathcal{H}_{0,\chi}$ can be seen as an acceptable "proxy" test of $\bar{\mathcal{H}}_0^C$ in general, at least when χ is "small".

Generally speaking, when $P\big(f_{\alpha^\star}(X) = g_{\beta^\star}(X)\big) > 0$, too small values of χ will induce high asymptotic variances of $\sqrt{n}\,\hat{C}_n$ and a higher statistical uncertainty when building confidence intervals. At the opposite, too large values of χ could induce a loss of power against some alternatives. An interesting question would be to find an optimal data-driven value for χ to keep the asymptotic level under $\mathcal{H}_{0,\chi} \cap \bar{\mathcal{H}}_0^C$ without losing too much power. Such a complex question would be related to local power analyses and lies far beyond the scope of this paper.

Example 1. To illustrate the theory above, assume that $\Psi = \Phi$, which implies that $H_\chi(f_{\hat{\alpha}_n}, g_{\hat{\beta}_n}, X) = \Phi\big((f_{\hat{\alpha}_n}(X) - g_{\hat{\beta}_n}(X))/\chi\big) - \Phi\big((g_{\hat{\beta}_n}(X) - f_{\hat{\alpha}_n}(X))/\chi\big)$. Additionally, assume that the true DGP follows a bivariate normal distribution with independent standard normal margins. Introduce the (misspecified) bivariate density of $X = (X_1, X_2)$ as

$$\psi_{(a_1, a_2)}(x) := \frac{a_1 a_2}{4} \exp\big(-a_1|x_1|\big) \exp\big(-a_2|x_2|\big), \ a_1, a_2 > 0.$$

It can be easily checked that the pseudo-true values for a_1 and a_2 are equal to $a^* := \sqrt{2\pi}/2$. Define the two competing parametric families of densities as $f_\alpha = \psi_{(\alpha, a^*)}$ and $g_\beta = \psi_{(a^*, \beta)}$. Obviously, the optimal models are identical, i.e. $f_{\alpha^\star} = g_{\beta^\star}$ with $\alpha^\star = \beta^\star = a^*$, and

$$h(\alpha, \beta) = \mathbb{E}\big[\Phi\big((f_\alpha(X) - g_\beta(X))/\chi\big) - \Phi\big((g_\beta(X) - f_\alpha(X))/\chi\big)\big].$$

Interchanging differentiation and integration, we obtain

$$h_k(\alpha^\star, \beta^\star) = \frac{(-1)^{k-1} a^*}{2\sqrt{2\pi}\,\chi} \mathbb{E}\big[(1 - a^*|X_k|) \exp\big(-a^*|X_1| - a^*|X_2|\big)\big] \neq 0, \ k = 1, 2,$$

which shows that $\sqrt{n}\hat{C}_n$ is non-degenerate asymptotically normal under $\mathcal{H}_{0,\chi}$. A further empirical analysis of this example can be found in Section 7.

Remark 2. A similar smoothing trick may be applied in the quantile regression setting when it cannot be excluded that $P\big(f_{\alpha^\star}(X) = g_{\beta^\star}(X)\big) > 0$ and/or to ensure the existence of the partial derivatives h_1 and h_2 appearing in the asymptotic variance of our test statistic. Similar to the (modified) Clarke test, smoothing is not necessary if the competing models are strictly locally non-nested, as long as the law of (Y, Z) is sufficiently regular. Otherwise, consider for instance the smooth function

$$l_\sigma(t) = \sigma(1 - \tau)\ln\big(2\cosh(t/\sigma)\big) \mathbf{1}\{t \leq 0\} + \sigma\tau\ln\big(2\cosh(t/\sigma)\big) \mathbf{1}\{t > 0\},$$

where $\sigma > 0$ acts as a smoothing parameter. If σ is small, then $l_\sigma(t) \approx \rho_\tau(t)$, the usual non-smooth loss-function of quantile regression. Now, one could test

$$\mathbb{E}\big[\rho_\tau\big(Y - q_1(Z, \alpha^\star)\big) - \rho_\tau\big(Y - q_2(Z, \beta^\star)\big)\big] = 0,$$

with the surrogate null hypothesis

$$\mathcal{H}_0 : \ \mathbb{E}\big[l_\sigma\big(Y - q_1(Z, \alpha^\star)\big) - l_\sigma\big(Y - q_2(Z, \beta^\star)\big)\big] = 0,$$

which ensures the existence of the partial derivatives h_1 and h_2 under suitable moment conditions on (Y, X). Moreover, in the same spirit as the modified Clarke test, one could test

$$P\big(\rho_\tau\big(Y - q_1(Z, \alpha^\star)\big) \geq \rho_\tau\big(Y - q_2(Z, \beta^\star)\big)\big) = P\big(\rho_\tau\big(Y - q_1(Z, \alpha^\star)\big) \leq \rho_\tau\big(Y - q_2(Z, \beta^\star)\big)\big)$$

with the surrogate null hypothesis

$$\mathcal{H}_0 : \mathbb{E}\left[\Phi_\chi\left(\rho_\tau\left(Y - q_1(Z, \alpha^\star)\right) - \rho_\tau\left(Y - q_2(Z, \beta^\star)\right)\right) - \Phi_\chi\left(\rho_\tau\left(Y - q_2(Z, \beta^\star)\right) - \rho_\tau\left(Y - q_1(Z, \alpha^\star)\right)\right)\right] = 0,$$

setting $\Phi_\chi(t) = \Phi(t/\chi)$ with a smoothing parameter $\chi > 0$.

Remark 3. Note that the smoothing parameter χ in (11) is assumed to be fixed. In particular, we have not considered a sequence of tuning parameters that would tend to zero. Contrary to an inference problem, the consistency of our test statistic $\sqrt{n}\,\hat{C}_n$ under $\mathcal{H}_{0,\chi}$ does not require such a behavior. Nevertheless, it could be interesting to use a sample size dependent smoothing parameter χ_n to ensure that H_{χ_n} tends to the discontinuous map H given in (9). Such ideas have been used in Horowitz (1998) or Whang (2006) to construct sample size dependent smoothed versions of the empirical likelihood in a quantile regression setting. In our testing framework, applying the modified Clarke test with $\Psi = \Phi$ and a smoothing parameter χ_n such that $\sqrt{n}\chi_n \to \infty$, it can be proved that $\sqrt{n}\chi_n\hat{C}_n$ is asymptotically normal under the assumption that $f_{\alpha^\star} = g_{\beta^\star}$ and additional regularity conditions: make a limited expansion of Φ_{χ_n} around zero and note that $(f_{\hat{\alpha}_n} - g_{\hat{\alpha}_n})(X_i) \simeq (\hat{\alpha}_n - \alpha^\star)\partial_\alpha f_{\alpha^\star}(X_i) - (\hat{\beta}_n - \beta^\star)\partial_\beta g_{\beta^\star}(X_i)$. Nonetheless, such a result is of little practical use because it cannot be extended when $f_{\alpha^\star} \neq g_{\beta^\star}$. Indeed, in such a case, the rate of convergence of $\hat{C}_n$ depends on the unknown rate of convergence of $P(H_{\chi_n} - H)^2$, which is why we do not provide additional details towards this direction.

6. Bootstrap test statistic

Invoking bootstrap techniques, we propose an estimation procedure of the asymptotic law of the general statistic $\sqrt{n}\,\hat{C}_n$. For this purpose, we need to introduce a slightly modified mathematical framework to account for the additional randomness induced by the bootstrap weights.

Let $(\xi_{i,n})_{1\le i\le n; n\in\mathbb{N}}$ denote an exchangeable triangular array of non-negative random variables (bootstrap weights) on a probability space Ω_2 endowed with a probability measure $\mathbb{P}_W$. Assume that the $\xi_{i,n}$ satisfy the following usual conditions given in Section 3.6.2 of van der Vaart and Wellner (1996).

Assumption.

W1. $\sum_{i=1}^n \xi_{i,n} = n$;
W2. $\sup_n \int_0^\infty \sqrt{\mathbb{P}_W(|\xi_{1,n} - 1| > x)}\, dx < \infty$;
W3. $\mathbb{E}_{\mathbb{P}_W}\left[\max_{1\le i\le n} |\xi_{i,n} - 1|\right]/\sqrt{n} \longrightarrow 0$;
W4. For some constant $c > 0$, $n^{-1}\sum_{i=1}^n \left(\xi_{i,n} - 1\right)^2 \longrightarrow c^2$ in $\mathbb{P}_W$-probability.

Many usual bootstrap schemes can be obtained by choosing particular exchangeable bootstrap weights. For instance, if $(\xi_{1,n}, \ldots, \xi_{n,n})$ are multinomial with parameters n and probabilities $(1/n, \ldots, 1/n)$, we recover Efron's nonparametric bootstrap (resampling with replacement) with $c = 1$. In econometrics, this corresponds to the so-called "pairwise bootstrap", for which couples (Y_i, Z_i) are redrawn with replacement. Moreover, consider i.i.d. non-negative random variables $(\tilde{\xi}_i)_{i\in\mathbb{N}}$ with mean $0 < \mu < \infty$ and variance $0 < \tau^2 < \infty$, such that $\int_0^\infty \left(\mathbb{P}_W(|\tilde{\xi}_1| > x)\right)^{1/2} dx < \infty$. Then, conditions W1-W4 are satisfied by setting $\xi_{i,n} := n\tilde{\xi}_i \left(\sum_{1\le i\le n} \tilde{\xi}_i\right)^{-1}$ with $c = \tau/\mu$, which yields a version of a multiplier bootstrap scheme (Kosorok (2008), Section 10.1). In the particular case of exponentially distributed random variables $\tilde{\xi}_i$ this is called Bayesian bootstrap in the literature.

Define the exchangeable bootstrap empirical measure as $\tilde{\mathbb{P}}_n := n^{-1}\sum_{i=1}^n \xi_{i,n}\delta_{X_i}$. Furthermore, define the bootstrap empirical process as $\tilde{\mathbb{G}}_n := \sqrt{n}c^{-1}(\tilde{\mathbb{P}}_n - \mathbb{P}_n)$ and assume that the sequences $(X_i)_{i\in\mathbb{N}}$ and $(\xi_{i,n})_{1\le i\le n; n\in\mathbb{N}}$ originate from a probability space with product structure as defined in Section 3 of Bücher and Kojadinovic (2018). This means that the sequence $(X_i)_{i\in\mathbb{N}}$ only depends on the first coordinate of some probability space $(\Omega := \Omega_1 \times \Omega_2, \mathcal{A} := \mathcal{A}_1 \times \mathcal{A}_2, \mathbb{P}_{XW})$ equipped with a probability measure $\mathbb{P}_{XW} := \mathbb{P}_X \otimes \mathbb{P}_W$. Similarly, the triangular array $(\xi_{i,n})_{1\le i\le n; n\in\mathbb{N}}$ only depends on the second component of the latter space, which implies that $(X_i)_{i\in\mathbb{N}}$ and $(\xi_{i,n})_{1\le i\le n; n\in\mathbb{N}}$ are independent. We write $\tilde{\mathbb{G}}_n \underset{\xi}{\rightsquigarrow} \mathbb{G}$ to denote conditional weak convergence in the sense of the bounded Lipschitz metric, as defined on p.73 in van der Vaart and Wellner (1996).

It is well-known that the bootstrap empirical process $\tilde{\mathbb{G}}_n$ weakly tends to the same asymptotic law as the empirical process $\mathbb{G}_n$ (Theorem 3.6.13 in van der Vaart and Wellner (1996)). In order to prove the convergence of our bootstrap process, we need additional assumptions on the estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ and on the class of functions $\mathcal{H}_\gamma$. Let $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ denote some "bootstrap estimators" of $\alpha^\star$ and $\beta^\star$, i.e. estimators of the pseudo-true values calculated from the bootstrap sample $(\xi_{i,n}, X_i)_{i=1,\ldots,n}$. Moreover, set $\tilde{\phi}_n := \phi(\cdot, \tilde{\alpha}_n, \tilde{\beta}_n)$.

Assumption.

B5. The bootstrap estimators satisfy $\sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) = \sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P}_n)s_1 + o_{\mathbb{P}_{XW}}(1)$ and $\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) = \sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P}_n)s_2 + o_{\mathbb{P}_{XW}}(1)$.
B6. $\int (\tilde{\phi}_n - \phi_n)^2\, dP$ tends to zero in $\mathbb{P}_{XW}$-probability.

The exchangeable bootstrap version of $\hat{C}_n$ is defined as

$$\tilde{C}_n := n^{-1} \sum_{i=1}^{n} \xi_{i,n} \phi(X_i, \tilde{\alpha}_n, \tilde{\beta}_n) = n^{-1} \sum_{i=1}^{n} \xi_{i,n} H\big(f_{\tilde{\alpha}_n}(X_i), g_{\tilde{\beta}_n}(X_i), X_i\big).$$

The next theorem, whose proof is given in Appendix A.4, shows that the bootstrap statistic $\tilde{C}_n$ behaves similarly to $\hat{C}_n$.

Theorem 6.1. *Assume that Assumptions B1–B6 are satisfied, and that $\mathcal{H}_\gamma$ is P-Donsker (for the constant γ given in B4). Then,*

(i) $\sqrt{n}\big(\tilde{C}_n - \hat{C}_n\big)/c \underset{\xi}{\rightsquigarrow} \mathcal{N}(0, \sigma^2).$

(ii) Statement (i) is also satisfied unconditionally, replacing $\underset{\xi}{\rightsquigarrow}$ by $\rightsquigarrow$ w.r.t. $\mathbb{P}_{XW}$.

Under $\mathcal{H}_0$, the asymptotic law of $\sqrt{n}\,\hat{C}_n$ can be numerically approximated by independently drawing $M \gg 1$ vectors of weights, which provide M realizations $\tilde{C}_n^{(j)}$, $j \in \{1, \dots, M\}$, of $\tilde{C}_n$, given the initial sample. Therefore, this yields a sample of $\sqrt{n}\big(\tilde{C}_n - \hat{C}_n\big)/c$, which follows the same asymptotic law as $\sqrt{n}\,\hat{C}_n$. As usual, the empirical percentiles of such values allow building confidence intervals. The bootstrap percentile method only relies on weak convergence results (not convergence in L^2). To be specific, any τ-quantile of the limiting law of $\sqrt{n}\big(\tilde{C}_n - \hat{C}_n\big)/c$, $\tau \in (0, 1)$, may be estimated by $\hat{q}_\tau := \inf\{t \mid G_M(t) > \tau\}$, where

$$G_M(t) := M^{-1} \sum_{j=1}^{M} \mathbf{1}\big(\sqrt{n}\big(\tilde{C}_n^{(j)} - \hat{C}_n\big)/c \le t\big),$$

for any t. This directly yields asymptotic confidence intervals without any asymptotic variance estimation: under $\mathcal{H}_0$, $\sqrt{n}\,\hat{C}_n$ belongs to $[\hat{q}_{\tau/2}, \hat{q}_{1-\tau/2}]$ with approximate probability $1 - \tau$, when n is sufficiently large. By a similar procedure, we obtain bootstrapped p-values of the test statistic $\sqrt{n}\,\hat{C}_n$ under $\mathcal{H}_0$. As a particular example, the bootstrap version of $\hat{B}_n$ is given by

$$\tilde{B}_n := n\tilde{\mathbb{P}}_n \mathbf{1}\{\psi(\cdot, \tilde{\alpha}_n, \tilde{\beta}_n) > 0\} = \sum_{i=1}^{n} \xi_{i,n} \mathbf{1}\{\psi(X_i, \tilde{\alpha}_n, \tilde{\beta}_n) > 0\}.$$

In this case we can replace Assumption B6 by a more explicit sufficient assumption.

Assumption.

B6′. For any compact set $K \subset Range(X)$ ψ satisfies the following uniform continuity condition: for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\|(\alpha, \beta) - (\alpha^\star, \beta^\star)\|_1 \le \delta \Rightarrow \sup_{x \in K} |\psi(x, \alpha, \beta) - \psi(x, \alpha^\star, \beta^\star)| \le \epsilon.$$

If there exists $\gamma > 0$ such that $\psi(x, \alpha, \beta)$ is continuous for all $(x, \alpha, \beta) \in \mathbb{R}^d \times E_\gamma$, then Assumption B6′ is satisfied (uniform continuity on a compact subset). We can deduce the next corollary from Theorem 6.1, whose proof can be found in Appendix A.5.

Corollary 6.2. *Assume Assumptions B1, B2, B3′, B4, B5, B6′, and that $\mathcal{H}_\gamma$ is P-Donsker (for the constant γ given in B3′ and B4). Then,*

(i) $\big(\tilde{B}_n - \hat{B}_n\big)/(c\sqrt{n}) \underset{\xi}{\rightsquigarrow} \mathcal{N}(0, \sigma_\psi^2).$

(ii) Statement (i) is also satisfied unconditionally, replacing $\underset{\xi}{\rightsquigarrow}$ by $\rightsquigarrow$ w.r.t. $\mathbb{P}_{XW}$.

Define $C_n^{(j)} := \sqrt{n}(\tilde{C}_n^{(j)} - \hat{C}_n)/c$, where $\tilde{C}_n^{(j)}$ denotes the bootstrap version of $\hat{C}_n$ calculated from the jth bootstrap sample, $j = 1, \dots, M$. Lemma 3.1(b) in Bücher and Kojadinovic (2018) implies that $(\mathbb{G}_n, \tilde{\mathbb{G}}_n^{(1)}, \dots, \tilde{\mathbb{G}}_n^{(M)}) \rightsquigarrow (\mathbb{G}, \mathbb{G}^{(1)}, \dots, \mathbb{G}^{(M)})$ for every fixed integer M, where $\tilde{\mathbb{G}}_n^{(j)}$ denotes the bootstrap empirical process calculated from the jth bootstrap sample and $\mathbb{G}, \mathbb{G}^{(1)}, \dots, \mathbb{G}^{(M)}$ are independent and identically distributed. Therefore, $\big(C_n^{(1)}, \dots, C_n^{(M)}\big) \rightsquigarrow \big(N^{(1)}, \dots, N^{(M)}\big)$, where $N^{(1)}, \dots, N^{(M)}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. It is then usual practice to estimate σ^2 by its bootstrap empirical variance $\hat{\sigma}_M^2$, i.e. by the sample variance (or the empirical second-order moment) of $\big(C_n^{(j)}\big)_{1 \le j \le M}$. In this case, we obtain the "bootstrap" test statistic $\sqrt{n}\,\hat{C}_n/\hat{\sigma}_M$.

Unfortunately, the convergence of $\hat{\sigma}_M^2$ towards σ^2 in $\mathbb{P}_{XW}$-probability is not always guaranteed (e.g. see a counterexample in Ghosh et al. (1984)). Indeed, the convergence in law of some sequence of statistics does not imply their convergence in mean. Here, we provide the consistency of $\hat{\sigma}_M^2$ under the following uniform integrability assumption,

which guarantees the equivalence between convergence in law and in mean (Theorem 25.12 in Billingsley, 1995). Again, the proof can be found in Appendix A.6.

Assumption.

B7. The sequence $\left(\left(C_n^{(1)}\right)^2\right)_{n\in\mathbb{N}}$ is uniformly integrable, i.e.

$$\lim_{A\to\infty}\sup_n \mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\mathbf{1}\left\{\left(C_n^{(1)}\right)^2 > A\right\}\right] = 0.$$

Proposition 6.3. *Under $\mathcal{H}_0^C$ and under Assumptions B1-B7 we have*

$$\hat{\sigma}_M^2 := \frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 \xrightarrow{n,M\to\infty} \sigma^2 \text{ in } \mathbb{P}_{XW}\text{-probability.}$$

7. Examples and simulation study

In this section, we present two simulation studies to explore the performance of the general test statistic $\sqrt{n}\,\hat{C}_n$ from (2) with the respective null hypothesis $\mathcal{H}_0$ in (1). An additional analysis of the empirical levels/powers of the corrected Clarke test statistic (7) is provided in Brück (2019), which also includes a simulation study for the example in Appendix B.

7.1. Gaussian regression models: the original and corrected Clarke tests

Consider Example 1 from Shi (2015b), who studies two Gaussian regression models using the two-step Vuong test and its extensions. We do not aim to compare the performances of the Clarke test and its corrected version with the performances of Vuong-type tests and we only borrow the framework of this instructive example. The models $\mathcal{F} := \{f_\alpha \mid \alpha \in \Theta_\alpha\}$ and $\mathcal{G} := \{g_\beta \mid \beta \in \Theta_\beta\}$ are density functions deduced from the linear regression models

$$\mathcal{F}: Y = \alpha^{(0)} + \sum_{j=1}^{K_f}\alpha^{(j)}Z_{f,j} + v, \text{ and } \mathcal{G}: Y = \beta^{(0)} + \sum_{j=1}^{K_g}\beta^{(j)}Z_{g,j} + u,$$

respectively. The covariates $Z_{f,j}$ and $Z_{g,j}$ are jointly independent and normally distributed with expectations 0 and variances 1. The errors u and v are again normally distributed with unknown variances σ_f^2 and σ_g^2, respectively, and they do not depend on the covariates. Note that these unknown variances σ_f^2 and σ_g^2 also constitute parameters of the models $\mathcal{F}$ and $\mathcal{G}$, respectively: our vectors of parameters are then $\alpha := \left(\alpha^{(0)}, \alpha^{(1)}, \ldots, \alpha^{(K_f)}, \sigma_f\right)$ and $\beta := \left(\beta^{(0)}, \beta^{(1)}, \ldots, \beta^{(K_g)}, \sigma_g\right)$. The true DGP depends on two parameters a_1 and a_2 and is given by

$$Y = 1 + \frac{a_1}{\sqrt{K_f}}\sum_{j=1}^{K_f}Z_{f,j} + \frac{a_2}{\sqrt{K_g}}\sum_{j=1}^{K_g}Z_{g,j} + \epsilon, \tag{12}$$

where ϵ is again normally distributed with expectation 0 and variance 1. For $a := a_1 = a_2$, the null hypothesis of the Clarke ($\mathcal{H}_0^C$) test is satisfied for every K_f and K_g, and the pseudo-true values $\alpha^\star$ and $\beta^\star$ can easily be determined. In particular, we get $\alpha^\star = (1, a/\sqrt{K_f}, \ldots, a/\sqrt{K_f}, \sqrt{1+a^2})$ for model $\mathcal{F}$ as well as $\beta^\star = (1, a/\sqrt{K_g}, \ldots, a/\sqrt{K_g}, \sqrt{1+a^2})$ for model $\mathcal{G}$.

To investigate the accuracy of the asymptotic distributional approximation of the corrected Clarke test given by Corollary 4.2 and Theorem 6.1, in particular its empirical level, we draw 1000 Monte-Carlo samples from the true DGP with $a_1 = a_2 = 0.25$, $K_f = 10$ and $K_g = 1, 2, \ldots, 20$. Moreover, we illustrate that the asymptotic variance of the Clarke test statistic is not equal to 1/4. To estimate its correct asymptotic variance, we draw $M = 1000$ bootstrap samples (resampling with replacement) from each underlying Monte-Carlo sample. Even though Clarke's null hypothesis is satisfied for any K_g, this dimension has an influence on the accuracy of the limiting law for the Clarke test statistic normalized by the bootstrapped standard deviation. If the numbers of covariates in models $\mathcal{F}$ and $\mathcal{G}$ significantly differ, we have observed a rather poor approximation of the asymptotic normal distribution in small samples. Therefore, we will consider the sample sizes $n = 1000, 2500$ and 5000 in the sequel.

Panel A of Fig. 1 shows QQ-plots of the Clarke test statistic normalized by the originally proposed standard deviation 1/2 for the sample size $n = 2500$ and $K_g \in \{1, 5, 10, 20\}$ in model $\mathcal{G}$. We can clearly conclude that the asymptotic standard normal approximation fails for the latter test statistic since the points in the QQ-plots significantly deviate from the gray straight line with intercept 0 and slope 1. Moreover, the shape of the QQ-plots indicate that the true asymptotic variance is much larger than 1/4. Panel B of Fig. 1 shows similar QQ-plots for the Clarke test statistic normalized by the bootstrapped standard deviation. For this correctly normalized test statistic, we observe that the asymptotic approximation with a standard normal distribution acceptably holds. Further, the influence of K_g on the asymptotic normal approximation is continuously diminishing with increasing sample size.
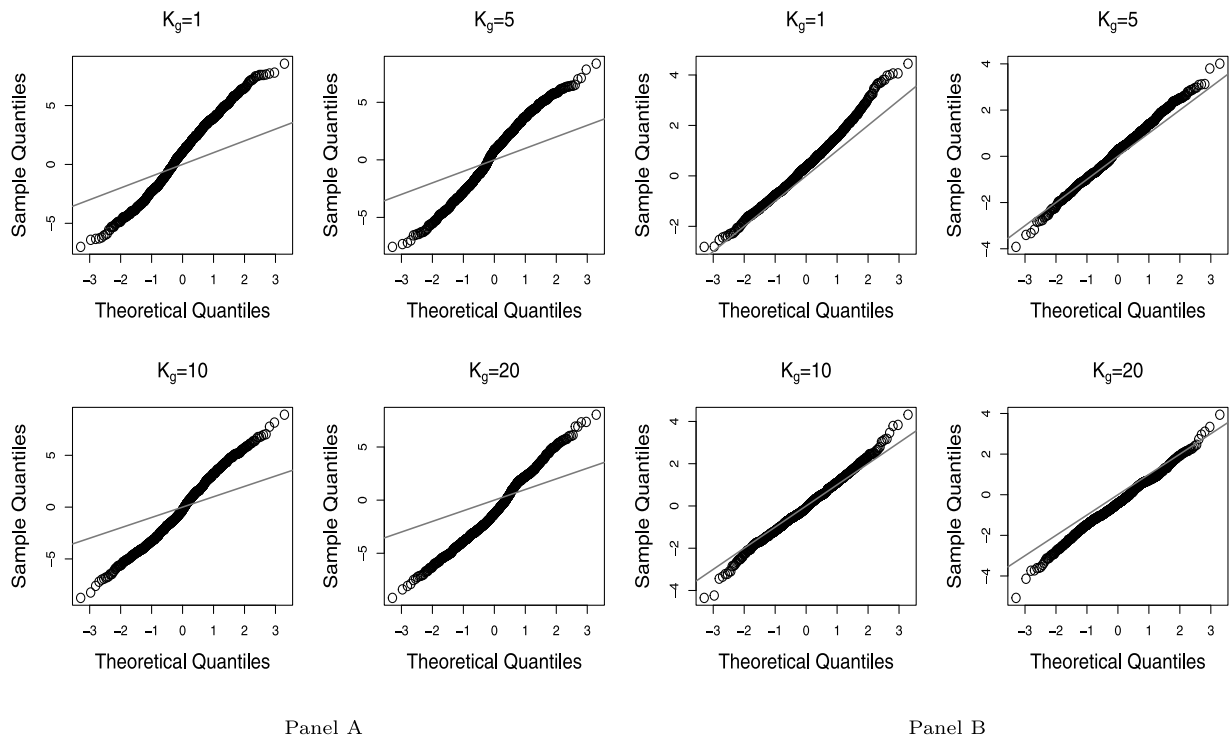
Panel A

Panel B

Fig. 1. Linear regression example: QQ-plots of the Clarke statistic normalized by 1/2 (Panel A) and of the correctly normalized Clarke statistic (Panel B) for sample size $n = 2500$ and different K_g, based on 1000 samples. The gray straight line has intercept 0 and slope 1.
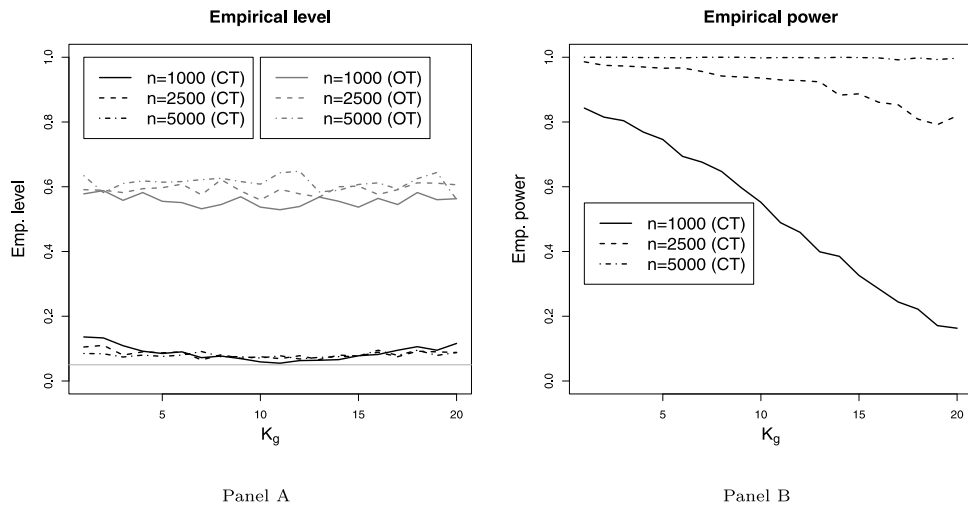

Panel A

Panel B

Fig. 2. Linear regression example: Empirical level (Panel A) and empirical power (Panel B) of the corrected Clarke test (CT) computed with bootstrapped p-values for $K_f = 10$ and different sample sizes (1000 samples). The empirical level of the original Clarke test (OT) is displayed in Panel A, where the solid horizontal line corresponds to the nominal level of 5%.

Panel A of Fig. 2 shows the empirical level of the original Clarke test (OT) and the corrected Clarke test (CT) for sample sizes $n = 1000, 2500$ and 5000 and for different number of covariates K_g. The empirical level of the corrected Clarke test is computed by bootstrapped p-values. The gray straight line indicates the chosen significance level of 5%. First, we observe that the original test completely fails to keep the nominal level. In contrast, the corrected Clarke test performs better, even if it slightly over-estimates the nominal level, especially for $n = 1000$.

To investigate the empirical power of the corrected Clarke test, we consider different a_1 and a_2, namely $a_1 = 0.25$ and $a_2 = 0.15$. Therefore, the null hypothesis $\mathcal{H}_0^C$ does not hold, which means that the error of one model tends to be

smaller than the error of the other model. For different sample sizes, we now determine the empirical powers for $K_f = 10$ and K_g running from 1 to 20. Panel B of Fig. 2 presents the empirical power of the corrected Clarke test computed with bootstrapped p-values. First and as expected, the latter test becomes more powerful with an increasing sample size. Second, the empirical power decreases by increasing the number of covariates K_g, due to a lack of estimation accuracy. The empirical power of the original Clarke test is useless and not shown here since it did not keep the nominal level at all.

7.2. Quantile regression models: a Clarke-type test

Let us keep the same DGP as specified in (12). Now, our aim is no longer to predict Y given some covariates Z but rather to predict the τ-quantile of Y given some covariates Z. To this purpose, we modify the previous example towards the two competing quantile regression models

$$\mathcal{F}: \ q_{f,\alpha}(Z) = \alpha^{(0)} + \sum_{j=1}^{K_f} \alpha^{(j)} Z_{f,j}, \ \text{and} \ \mathcal{G}: \ q_{g,\beta}(Z) = \beta^{(0)} + \sum_{j=1}^{K_g} \beta^{(j)} Z_{g,j},$$

where $q_{f,\alpha}(z)$ and $q_{g,\beta}(z)$ denote the τ-quantiles of Y given $Z = z$ given by the models $\mathcal{F}$ and $\mathcal{G}$ respectively. Our general testing methodology will allow to discriminate between $\mathcal{F}$ and $\mathcal{G}$. The common inference approach is to minimize the loss functions $L(Y - q_{f,\alpha}(Z)) := \tau(Y - q_{f,\alpha}(Z))\mathbf{1}\{Y > q_{f,\alpha}(Z)\} + (1 - \tau)(q_{f,\alpha}(Z) - Y)\mathbf{1}\{Y \leq q_{f,\alpha}(Z)\}$ and $L(Y - q_{g,\beta}(Z)) := \tau(Y - q_{g,\beta}(Z))\mathbf{1}\{Y > q_{g,\beta}(Z)\} + (1 - \tau)(q_{g,\beta}(Z) - Y)\mathbf{1}\{Y \leq q_{g,\beta}(Z)\}$, respectively. Thus, the pseudo-true value $\alpha^\star$ is given by

$$\underset{\alpha \in \mathbb{R}^{K_f+1}}{\operatorname{argmin}} \ \mathbb{E}\left[L\left(Y - \alpha^{(0)} - \sum_{j=1}^{K_f} \alpha^{(j)} Z_{f,j}\right)\right]$$

and similarly for $\beta^\star$. Simple calculations provide

$$\alpha^\star = \left(1 + \sqrt{1 + a_2^2}\,\Phi^{-1}(\tau), \frac{a_1}{\sqrt{K_f}}, \dots, \frac{a_1}{\sqrt{K_f}}\right), \ \text{and} \ \beta^\star = \left(1 + \sqrt{1 + a_1^2}\,\Phi^{-1}(\tau), \frac{a_2}{\sqrt{K_g}}, \dots, \frac{a_2}{\sqrt{K_g}}\right),$$

where Φ denotes the cumulative distribution function of a standard normal random variable. Now let us define $H(q_{f,\alpha}(Z), q_{g,\beta}(Z), Y) = \mathbf{1}\{L(Y - q_{f,\alpha}(Z)) > L(Y - q_{g,\beta}(Z))\} - 1/2$ which can be seen as a version of the Clarke test to discriminate between quantile regression models. Thus, the corresponding null hypothesis is given by

$$\mathcal{H}_0^{C,q}: \ P\left(L(Y - q_{f,\alpha^\star}(Z)) > L(Y - q_{g,\beta^\star}(Z))\right) = 1/2.$$

A violation of $\mathcal{H}_0^{C,q}$ would imply that one model tends to have a larger loss than the other model. Plugging in $\alpha^\star$ and $\beta^\star$ yields

$$L(Y - q_{f,\alpha^\star}(Z)) = \bar{f}\left(\frac{1}{\sqrt{K_g}}\sum_{j=1}^{K_g} Z_{g,j}; a_1; a_2; \epsilon\right), \ \text{and} \ L(Y - q_{g,\beta^\star}(Z)) = \bar{f}\left(\frac{1}{\sqrt{K_f}}\sum_{j=1}^{K_f} Z_{f,j}; a_2; a_1; \epsilon\right),$$

for some fixed rather complicated function $\bar{f}$ that we do not specify here. Therefore, if $a_1 = a_2$, $\mathcal{H}_0^{C,q}$ is clearly satisfied since, conditioned on ϵ, the remaining random variables in $\bar{f}$ are i.i.d. (apply the tower rule for conditional expectations). Under some regularity conditions, see Section 4.3 of Koenker (2005), we also obtain the (joint) asymptotic normality of the parameter estimators

$$\hat{\alpha}_n := \underset{\alpha \in \mathbb{R}^{K_f+1}}{\operatorname{argmin}} \ \frac{1}{n}\sum_{i=1}^n L(Y_i - q_{f,\alpha}(Z_i)) \ \text{and} \ \hat{\beta}_n := \underset{\beta \in \mathbb{R}^{K_f+1}}{\operatorname{argmin}} \ \frac{1}{n}\sum_{i=1}^n L(Y_i - q_{g,\beta}(Z_i)).$$

Now, we investigate the accuracy of the asymptotic distributional approximation provided in Theorems 3.1 and 6.1. To calculate the empirical level of the test statistic $T_n^{C,q} := \sqrt{n}\hat{C}_n/\hat{\sigma}_M$ corresponding to the null hypothesis $\mathcal{H}_0^{C,q}$ for the particular case $\tau = 0.5$, we draw 1000 Monte-Carlo samples from the true DGP with $a_1 = a_2 = 0.25$, $K_f = 10$ and $K_g = 1, 2, \dots, 20$. To estimate the asymptotic variance of $\sqrt{n}\hat{C}_n$, we draw $M = 1000$ bootstrap samples from each underlying Monte-Carlo sample. The null hypothesis $\mathcal{H}_0^{C,q}$ is satisfied for any K_g. However, its influence on the distributional approximation of the test statistic is again present for small sample sizes. As for the linear regression example, we therefore consider the sample sizes $n = 1000, 2500, 5000$ in the sequel.

Fig. 3 shows the QQ-plots of the test statistic $T_n^{C,q}$ for the sample size $n = 2500$ and the number of covariates $K_g = 1, 5, 10$, and 20. Similarly to the linear regression case, the influence of K_g on the quality of the asymptotic normal approximation is rather rapidly diminishing with the sample sizes. Panel A of Fig. 4 presents the empirical levels of the test based on $T_n^{C,q}$ for the considered sample sizes. Here, the proposed test keeps its level relatively conveniently.
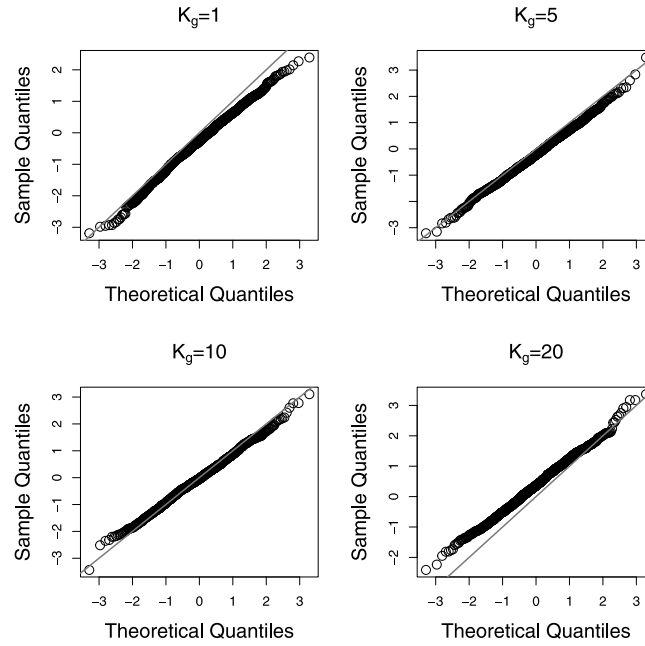
Fig. 3. Quantile regression: QQ-plots of the test statistic $T_n^{C,q}$ for $K_f = 10$ and sample size $n = 2500$ (based on 1000 samples). The gray straight line has intercept 0 and slope 1.
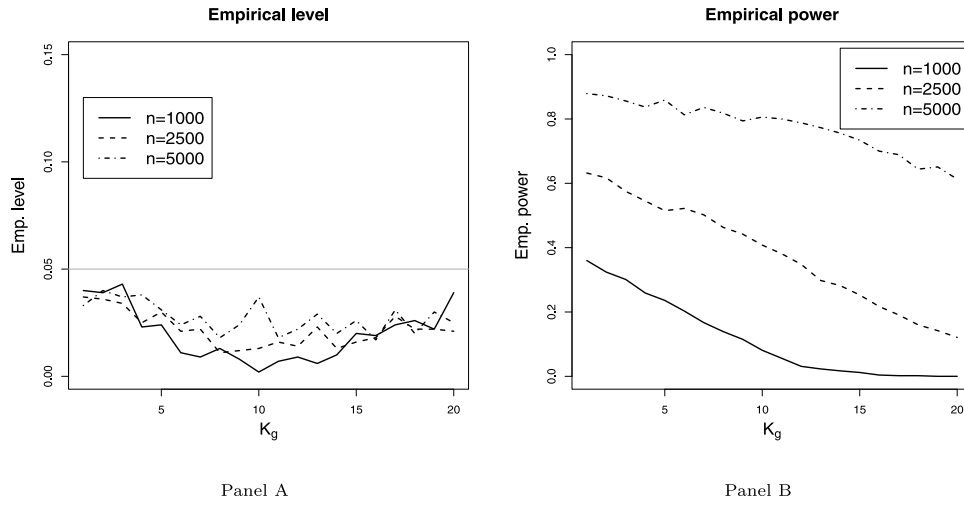

Fig. 4. Quantile regression: Empirical level (Panel A) and empirical power (Panel B) of the test based on $T_n^{C,q}$ for $K_f = 10$ (based on 1000 samples). The solid horizontal line in Panel A corresponds to the nominal level of 5%.

To investigate the empirical power of the latter test, we again set $K_f = 10$. Further, we consider slightly different a_1 and a_2, namely $a_1 = 0.25$ and $a_2 = 0.15$. Therefore, the null hypothesis $\mathcal{H}_0^{C,q}$ does not hold. For different sample sizes, we determine the empirical powers for K_g running from 1 to 20. Panel B of Fig. 4 presents the empirical power of the test based on $T_n^{C,q}$. As expected, the latter test becomes more powerful with an increasing sample size. Further, the empirical power decreases by increasing the number of covariates in the competing model $\mathcal{G}$, due to a rising amount of estimation noise.

7.3. Identical optimal models: a modified Clarke test for bivariate Laplace distributions

Consider Example 1 from Section 5. We will present the empirical levels and powers of the modified Clarke test $\hat{C}_{n,\chi}$ that is associated to the map H_χ in (10) for identical optimal models and sample sizes $n \in \{100, 250, 500\}$. Moreover,
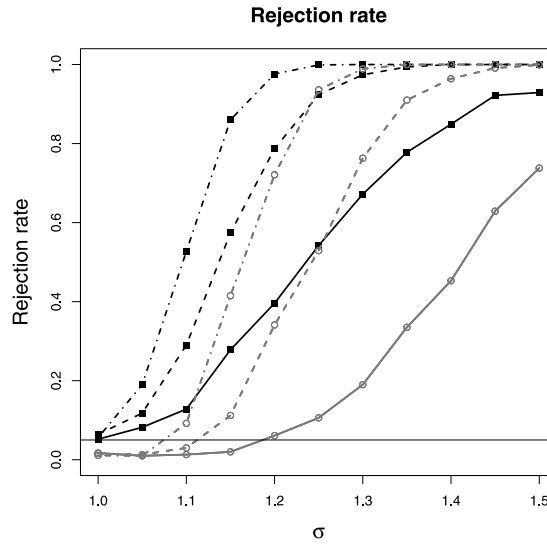
Rejection rate

Fig. 5. Rejection rates of the null hypothesis by the smoothed Clarke-type test (black lines) and the SW-test (gray lines) for sample sizes $n = 100$ (solid), $n = 250$ (dashed), $n = 500$ (dot-dashed) and $\sigma = 1 + 0.05 \cdot k$ with $k = 0, 1, \ldots, 10$. The solid horizontal line corresponds to the nominal level of 5%. The smoothing parameter $\chi = 2^{-4}$ and the optimal regularization parameter ε_n for the modified Clarke test and the SW-test are chosen, respectively.

we will compare the latter test with the SW-test by Schennach and Wilhelm (2017). The nominal level of these tests is set to 0.05. Note that both tests depend on a corresponding tuning parameter, which is theoretically investigated for the SW-test and stays an open problem for our test. We follow the data-driven procedure of Schennach and Wilhelm (2017) for the selection of their regularization parameter ε_n.

The models $\mathcal{F} := \{f_\alpha = \psi_{(\alpha, a^\star)} \mid \alpha > 0\}$ and $\mathcal{G} := \{g_\beta = \psi_{(a^\star, \beta)} \mid \beta > 0\}$ are the density functions introduced in Section 5. Recall that the two optimal models are identical for $\alpha = \beta = a^\star = \sqrt{2\pi}/2$ if the true DGP is a bivariate normal distribution with independent standard normal margins. Now, we simulate the first margin X_1 from the normal distribution with expectation 0 and standard deviation $\sigma = 1 + 0.05 \cdot k$ for $k = 0, 1, \ldots, 10$ while the second margin X_2 is drawn from the standard normal distribution. It can be analytically or numerically checked that, when $k = 0$ (i.e. $f_{\alpha^\star} = g_{\beta^\star}$), $\bar{\mathcal{H}}_0^C$ and $\mathcal{H}_{0,\chi}$ are both true, for any value of χ. At the opposite, when $k \geq 1$, $\bar{\mathcal{H}}_0^C$ and $\mathcal{H}_{0,\chi}$ are not satisfied, for almost any value of χ. Thus, the percentage of rejections of the null hypothesis for $k = 0$ (resp. for $k = 1, \ldots, 10$) corresponds to the empirical level (resp. the empirical power) of our "smoothed" Clarke test statistic $\hat{C}_{n,\chi}$ and should be close to 5% under $\bar{\mathcal{H}}_0^C$. Note that the chosen DGP for power analysis purposes favors the model $\mathcal{F}$ over the model $\mathcal{G}$ since the pseudo-maximum likelihood estimator of α can capture the changing data variability in the first margin. Fig. 5 shows the empirical levels ($\sigma = 1$) and the empirical powers ($\sigma > 1$) for $\hat{C}_{n,\chi}$ in addition to the SW-test, in our considered framework and with a reasonable value of χ (see below). The empirical level of the modified Clarke test $\hat{C}_{n,\chi}$ is slightly larger than the nominal level 0.05 while the SW-test is a bit liberal. It seems that the former test is slightly more powerful than the latter, for this example. However, this empirical observation should be dealt with caution and may be linked to the respective choices of the tuning parameters.

To empirically illustrate the dependence of $\hat{C}_{n,\chi}$ on the smoothing parameter χ, Table 1 presents the rejection fractions of the null hypothesis for $\chi = 2^{-k}$, $k = 10, \ldots, 0$. We still consider the standard deviations $\sigma = 1 + 0.05 \cdot k$ with $k = 0, 1, \ldots, 10$ and a single sample size $n = 500$. Too small values of χ do not keep the nominal level and result in a power loss. If χ increases, then our modified Clarke test starts to keep the nominal level and becomes more powerful. We observe that the value $\chi = 2^{-4} = 0.0625$ chosen in Fig. 5 or even larger values yield convenient level and power results, when $n \in \{100, 250, 500\}$. Note that, with values of χ larger than one, $\Phi((f_\alpha - g_\beta)/\chi) - \Phi((g_\beta - f_\alpha)/\chi)$ significantly deviates from the desired target $\mathbf{1}\{f_\alpha(x) > g_\beta(x)\} - \mathbf{1}\{f_\alpha(x) < g_\beta(x)\}$ but the performances of $\hat{C}_{n,\chi}$ remain very good even in such circumstances. However, this is a consequence of our particular choices for the DPG and the considered models and it cannot be generalized.

8. Empirical analysis

To compare the performances of the original and corrected Clarke tests in practical applications, we have considered stock price indices per countries. We have chosen the MSCI Barra (formerly Morgan Stanley Capital International) indices that yield standard benchmarks for financial markets and are freely available on the web. Here, we present our empirical analysis on the daily log-returns calculated from the MSCI indices of Austria, Germany, Ireland, Italy, Netherlands,

Table 1

Rejection rate of the null hypothesis for the different standard deviations σ and the smoothing parameters χ. The sample size n is equal to 500.

Parameter	$\chi = 2^{-k}$										
σ	$k = 10$	$k = 9$	$k = 8$	$k = 7$	$k = 6$	$k = 5$	$k = 4$	$k = 3$	$k = 2$	$k = 1$	$k = 0$
$\sigma = 1.00$	0.200	0.205	0.181	0.143	0.106	0.066	0.056	0.052	0.048	0.046	0.049
$\sigma = 1.05$	0.385	0.415	0.410	0.358	0.286	0.223	0.190	0.181	0.175	0.173	0.172
$\sigma = 1.10$	0.637	0.697	0.717	0.685	0.630	0.550	0.527	0.504	0.488	0.495	0.496
$\sigma = 1.15$	0.817	0.919	0.935	0.930	0.912	0.884	0.860	0.860	0.840	0.840	0.847
$\sigma = 1.20$	0.731	0.939	0.991	0.991	0.986	0.979	0.976	0.973	0.969	0.971	0.969
$\sigma = 1.25$	0.537	0.855	0.994	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000
$\sigma = 1.30$	0.294	0.654	0.969	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma = 1.35$	0.131	0.394	0.888	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma = 1.40$	0.073	0.217	0.745	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma = 1.45$	0.062	0.112	0.557	0.988	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma = 1.50$	0.099	0.046	0.320	0.960	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Singapore and Sweden from December 31, 1998 till March 12, 2018. For risk management purposes, a standard way of calculating risk measures as Value-at-Risk or Expected Shortfall is first to "filter" every series of returns by removing their conditional means. Afterwards, most often, the associated sequences of residuals could reasonably well be considered as i.i.d. By specifying marginal parametric laws and a copula, the full multivariate process is modeled and can be simulated. To follow this strategy, we first filter the considered seven return time series with an $ARMA(p, q) - GARCH(1, 1)$ model with normal innovations, $p \in \{0, 1\}$ and $q \in \{0, 1\}$. Such models are popular in financial econometrics as they are the simplest and most robust volatility models. They conveniently fit many data series (Hill et al. (2018), p.526) and are often sufficient to capture volatility clustering features (Brooks (2014), p.430). Furthermore, a single lag for conditional means and volatilities is often sufficient empirically (Angelidis et al. (2004), Olson and Wu (2017) p.63). For every time series, we have chosen the best orders (p, q) of each $ARMA(p, q)$ part by the BIC. The standardized residuals of the chosen $ARMA(p, q) - GARCH(1, 1)$ models represent the data on which we apply our tests. The null hypothesis of the augmented Dickey–Fuller test for the latter series is rejected at 5% nominal level and such series of residuals will be considered as stationary.

Now, we focus on our model residuals whose laws are most often fat-tailed, a key feature for risk measure calculation. Here, our initial model selection problem naturally appears. By construction and for the sake of consistency, we have to assume that the standardized residuals are i.i.d. with expectation 0 and variance 1. To satisfy the two latter constraints, we need to consider two parametric families with at least three free parameters. Model $\mathcal{F}$ assumes that the standardized residuals follow a Student distribution with location parameter zero, scale parameter $s > 0$ and ν degrees of freedom, $\nu \in (2, 20)$. To ensure the restriction on the variance, the scale parameter s is set to $s = (\nu - 2)/\nu$. Therefore, the degrees of freedom ν completely parameterize model $\mathcal{F}$. Model $\mathcal{G}$ assumes that the standardized residuals follow a mixture of two independent Gaussian distributions with zero expectation. Further parameters of this Gaussian mixture are the weight $\omega \in (0, 1)$, the variance of the first mixture component $\sigma_1^2 > 0$ and the variance of the second mixture component $\sigma_2^2 > 0$. We relate the weight ω to the first mixture component. Further, the restriction on the residual variance is satisfied if $\omega\sigma_1^2 + (1 - \omega)\sigma_2^2 = 1$. Thus, model $\mathcal{G}$ is parameterized by ω and σ_1^2. Identification and numerical maximization of the log-likelihood of model $\mathcal{G}$ require further constraints on ω and σ_1^2, namely $\omega \in (0.5, 0.9999)$ and $\sigma_1^2 \in (0.0001, 0.9999)$ to obtain a well-defined estimator for σ_2^2.

For the considered residuals, the pseudo-maximum likelihood estimators of ν in model $\mathcal{F}$ range between 7.230 (Ireland) and 8.820 (Italy). The latter estimators for model $\mathcal{G}$ clearly indicate a non-degenerate Gaussian mixture since the estimators of ω range between 0.504 (Germany) and 0.917 (Italy). The original Clarke test rejects the null hypothesis in favor of $\mathcal{F}$ with very low p-values ranging between $6.95 \cdot 10^{-36}$ (Singapore) and $6.08 \cdot 10^{-7}$ (Italy). Table 2 presents all p-values of the original Clarke test in the second column. It also displays the bootstrapped p-values for the corrected Clarke test, which does not reject the null hypothesis for the considered standardized residuals at the 5% nominal level. According to the last column of Table 2, the one-step Vuong test cannot differentiate between models $\mathcal{F}$ and $\mathcal{G}$ at the 5% level either. It is not surprising that the original Clarke test rejects the null hypotheses with very low p-values, since it is not able to keep the nominal level at all. Therefore, the use of the original Clarke test in applied science may lead to erroneous decisions. Finally, note that our two-step selection framework as detailed above is popular in empirical science to select ARMA-GARCH models and their parametric innovations. It may be preferred over simultaneous comparisons of ARMA-GARCH models with different specifications and innovations using formal statistical tests, since the latter way of working would require a corresponding asymptotic theory and related bootstrap procedures for stationary time series.

9. Discussion and conclusion

In this paper we revisit the intuition of the original Clarke test for model selection purposes, possibly under misspecification, which was introduced in Clarke (2007). We show that the originally proposed test statistic is not Binomial

Table 2

P-values for the original Clarke test, the corrected Clarke test with bootstrapped p-values and one-step Vuong test for Australia, Canada, Italy, Japan, Spain, Switzerland and UK.

Country	P-value of the original the Clarke test	Bootstrapped P-value of the corrected Clarke test	P-value of the one-step Vuong test
Austria	$5.99 \cdot 10^{-23}$	0.374	0.370
Germany	$3.19 \cdot 10^{-9}$	0.366	0.327
Ireland	$7.17 \cdot 10^{-12}$	0.345	0.124
Italy	$6.08 \cdot 10^{-7}$	0.099	0.094
Netherlands	$3.85 \cdot 10^{-8}$	0.390	0.328
Singapore	$6.95 \cdot 10^{-36}$	0.483	0.392
Sweden	$2.69 \cdot 10^{-9}$	0.298	0.103

distributed as claimed in Clarke (2007). To overcome this issue and to go beyond, we introduce a family of general test statistics for model selection, which are suitable for non-smooth criterion functions. They may potentially be applied to non-nested and some nested or overlapping competing models. Their wide range of applicability is demonstrated by several examples including linear, quantile and logistic regression. We derive the asymptotic normality of the general test statistic and provide its bootstrap consistency. Moreover, two consistent estimators of the asymptotic variance of the latter test statistic – one is based on numerical derivatives, the other is based on a bootstrap approach – are studied. As a byproduct, we prove the consistency of a general asymptotic variance estimator based on the empirical variance of bootstrapped test statistics under minimal assumptions. We obtain that a normalized version of the original Clarke test statistic follows an asymptotic normal distribution, whose asymptotic variance usually exceeds the asymptotic variance of the claimed Binomial distribution.

The finite sample properties of our general test statistic are demonstrated in a simulation study with the help of linear and quantile regression models. The results show that the original Clarke test statistic cannot keep its level under the null hypothesis, whereas our corrected version of this test yields acceptable results. Moreover, we shortly illustrate how a modified version of the Clarke test can be applied to models which are not strictly locally non-nested. Finally, we have applied the corrected Clarke test to stock market data using ARMA-GARCH models. It is illustrated that the corrected Clarke test statistic can revert the decisions made by the original Clarke test.

Further research may allow for increasing dimensions of α_n and β_n with increasing sample size or even for infinite dimensional parameters (α, β). This would yield a significant extension of Theorem 3.1.

Moreover, the problem of simultaneously testing multiple competing families seems to be feasible, but would require a lot of modifications. Additionally, a finite sample correction in the spirit of Shi (2015b), Liao and Shi (2020) or Schennach and Wilhelm (2017) may be investigated. Finally, a mathematical framework for the optimal choice of the shrinkage parameter $e(n)$ in Theorem 3.2 would prove valuable.

Acknowledgments

Appendix A. Proofs

A.1. Proof of Theorem 3.1

Proof of (i). Under the null, we can rewrite the normalized test statistic as

$$\sqrt{n}\,\hat{C}_n = \sqrt{n}(\mathbb{P}_n\phi_n - P\phi_n) + \sqrt{n}(P\phi_n - P\phi_\star)$$
$$= \mathbb{G}_n(\phi_n - \phi_\star) + \mathbb{G}_n\phi_\star + \sqrt{n}P(\phi_n - \phi_\star).$$

We use Corollary 2.3.12 in van der Vaart and Wellner (1996) to show that $\mathbb{G}_n(\phi_n - \phi_\star)$ is $o_{\mathbb{P}}(1)$. Consider some arbitrary $\epsilon > 0$ and $v > 0$. Since $\mathcal{H}_\gamma$ is Donsker, there exists $\delta > 0$ and $N \in \mathbb{N}$ such that the equicontinuity condition (2.1.8) from van der Vaart and Wellner (1996) is satisfied, i.e.

$$\mathbb{P}\left(\sup_{\rho_P(f-g)\leq\delta} |\mathbb{G}_n(f - g)| > \epsilon\right) \leq v,$$

for $n \geq N$. Obviously, we implicitly assume the maps f and g above belong to $\mathcal{H}_\gamma$. Moreover, Assumption B3 implies that $\rho_P(\phi_n, \phi_\star) := \sqrt{P(\phi_n - \phi_\star)^2 - (P(\phi_n - \phi_\star))^2} \to 0$ in probability. Therefore, we can impose $\mathbb{P}(\rho_P(\phi_n, \phi_\star) > \delta) \leq v$ when

$n \geq N$ in addition. This yields

$$
\begin{aligned}
&\mathbb{P}\left(|\mathbb{G}_n(\phi_n - \phi_\star)| > \epsilon\right)\\
&\quad\leq \mathbb{P}\left(|\mathbb{G}_n(\phi_n - \phi_\star)| > \epsilon , \rho_P(\phi_n, \phi_\star) > \delta\right) + \mathbb{P}\left(|\mathbb{G}_n(\phi_n - \phi_\star)| > \epsilon , \rho_P(\phi_n, \phi_\star) \leq \delta\right)\\
&\quad\leq \mathbb{P}\left(\rho_P(\phi_n, \phi_\star) > \delta\right) + \mathbb{P}\left(\sup_{\rho_P(f - g) \leq \delta} |\mathbb{G}_n(f - g)| > \epsilon\right) \leq 2\nu.
\end{aligned}
$$

Since ν was arbitrary, we get $\mathbb{G}_n(\phi_n - \phi_\star) = o_{\mathbb{P}}(1)$, which allows us to solely focus on the convergence of $\mathbb{G}_n\phi_\star + \sqrt{n}P(\phi_n - \phi_\star)$ in the remaining part of the proof. By a limited expansion of h and under $\mathcal{H}_0$, we have

$$
\begin{aligned}
P(\phi_n - \phi_\star) &= h(\alpha^\star, \beta^\star) + h_1(\alpha^\star, \beta^\star) \cdot (\hat\alpha_n - \alpha^\star) + h_2(\alpha^\star, \beta^\star) \cdot (\hat\beta_n - \beta^\star)\\
&\quad + o_{\mathbb{P}}\left(\|\hat\alpha_n - \alpha^\star\|\right) + o_{\mathbb{P}}\left(\|\hat\beta_n - \beta^\star\|\right)\\
&= h_1(\alpha^\star, \beta^\star) \cdot (\hat\alpha_n - \alpha^\star) + h_2(\alpha^\star, \beta^\star) \cdot (\hat\beta_n - \beta^\star) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

noting that $h(\alpha^\star, \beta^\star) = 0$. Thus, we get

$$
\begin{aligned}
\sqrt{n}\,\hat{C}_n &= \mathbb{G}_n\phi_\star + h_1(\alpha^\star, \beta^\star) \cdot \sqrt{n}(\hat\alpha_n - \alpha^\star) + h_2(\alpha^\star, \beta^\star) \cdot \sqrt{n}(\hat\beta_n - \beta^\star) + o_{\mathbb{P}}(1)\\
&= \mathbb{G}_n\left(\phi_\star + h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2\right) + o_{\mathbb{P}}(1),
\end{aligned}
$$

by Assumption B2. By the usual CLT, we obtain

$$
\mathbb{G}_n\left(\phi_\star + h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2\right) + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}(0, \sigma^2),
$$

proving the result.

Proof of (ii). By the first part of the proof, we know that

$$
\begin{aligned}
\sqrt{n}\,\hat{C}_n &= \mathbb{G}_n(\phi_n - \phi_\star) + \mathbb{G}_n\phi_\star + \sqrt{n}P(\phi_n - \phi_\star) + \sqrt{n}P\phi_\star\\
&= O_{\mathbb{P}}(1) + \sqrt{n}\,P\phi_\star.
\end{aligned}
$$

If $P\phi_\star < 0$, then this yields $\sqrt{n}\,\hat{C}_n \to -\infty$.

Proof of (iii). Use a similar argument when $P\phi_\star > 0$.

Remark 4. The proof of Theorem 3.1 has been obtained by a direct reasoning instead of applying some more general results that are available in the literature. Indeed, the high level of generality of many papers does not yield minimal assumptions or induces unnecessary technicalities. For example, it is difficult to apply Theorem 2 of Chen et al. (2003) that is well-suited to nonparametric first-stage estimators. To do that, it would be necessary to discuss pathwise derivatives, when our problem is fundamentally fully (two-stage) parametric. Moreover, we do not need to control the rate of convergence to zero of $\|f_{\hat\alpha_n} - f_{\alpha^\star}\|$ (for some norm in a functional space), as in their assumption (2.4). The same constraint arises in Ichimura and Lee (2010), that requires a control of the sup-norm between the latter maps, when the key quantity is $\|\hat\alpha_n - \alpha^\star\|$ for us. Another classical reference is Theorem 3.3 in Pakes and Pollard (1989). But checking its assumption (iv) is exactly redoing the main content of our proof. The same remark applies concerning Theorem 1 and Assumption N(d) in Andrews (1994). In the literature, we are not aware of a theoretical result in a pure two-stage parametric framework under weaker assumptions than in our Theorem 3.1.

A.2. Proof of Theorem 3.2

Define $\mathcal{Q}_\gamma := \{f + u_1 \cdot s_1 + u_2 \cdot s_2 \mid f \in \mathcal{H}_\gamma, (u_1, u_2) \in [h_1(\alpha^\star, \beta^\star) - \gamma, h_1(\alpha^\star, \beta^\star) + \gamma] \times [h_2(\alpha^\star, \beta^\star) - \gamma, h_2(\alpha^\star, \beta^\star) + \gamma]\}$. $\mathcal{Q}_\gamma$ is P-Glivenko–Cantelli, since it is a finite sum of P-Donsker classes. Additionally, $\mathcal{Q}_\gamma^2 := \{q^2 \mid q \in \mathcal{Q}_\gamma\}$ is also P-Glivenko–Cantelli in $\mathbb{P}$-probability by Lemma 2.10.14 in van der Vaart and Wellner (1996). Then, it is sufficient to show

$$
\mathbb{P}_n\left(\phi_n + \hat{h}_{1,n} \cdot s_1 + \hat{h}_{2,n} \cdot s_2\right)^k \to P\left(\phi_\star + h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2\right)^k, \tag{A.1}
$$

in $\mathbb{P}$-probability, when $k = 1, 2$. For $k = 1$, the latter result comes from the fact $\mathcal{Q}_\gamma$ is P-Glivenko–Cantelli. Let us formally prove (A.1) when $k = 2$. Set $\epsilon > 0$ and denote $q_n := \phi_n + \hat{h}_{1,n} \cdot s_1 + \hat{h}_{2,n} \cdot s_2$ and $q := \phi_\star + h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2$. Simple calculations provide

$$
\begin{aligned}
P(q_n^2 - q^2) &= P(\phi_n^2 - \phi_\star^2) + 2P(\phi_n - \phi_\star)\hat{h}_{1,n} \cdot s_1 + 2P\phi_\star(\hat{h}_{1,n} - h_1(\alpha^\star, \beta^\star)) \cdot s_1\\
&\quad + 2P(\phi_n - \phi_\star)\hat{h}_{2,n} \cdot s_2 + 2P\phi_\star(\hat{h}_{2,n} - h_2(\alpha^\star, \beta^\star)) \cdot s_2\\
&\quad + P\left((\hat{h}_{1,n} - h_1(\alpha^\star\beta^\star)) \cdot s_1\right)\left((\hat{h}_{1,n} + h_1(\alpha^\star\beta^\star)) \cdot s_1\right)
\end{aligned}
$$

$$+P\big((\hat{h}_{2,n} - h_2(\alpha^\star\beta^\star)) \cdot s_2\big)\big((\hat{h}_{2,n} + h_2(\alpha^\star\beta^\star)) \cdot s_2\big)$$

$$+2\sum_{i=1}^{d_\alpha}\sum_{j=1}^{d_\beta}\Big(\hat{h}_{1,n,i}\hat{h}_{2,n,j} - h_{1,i}(\alpha^\star, \beta^\star)h_{2,j}(\alpha^\star, \beta^\star)\Big)Ps_{1,i}s_{2,j},$$

which tends to zero in $\mathbb{P}$-probability. Indeed, by B3 or B3′, $P(\phi_n - \phi_\star)^2 \to 0$. Moreover, by Lemma 1 (see below), $\hat{h}_{1,n,i} \to h_{1,i}(\alpha^\star, \beta^\star)$, $\hat{h}_{2,n,i} \to h_{2,i}(\alpha^\star, \beta^\star)$ as well as $\hat{h}_{1,n,i}\hat{h}_{2,n,j} \to h_{1,i}(\alpha^\star, \beta^\star) h_{2,j}(\alpha^\star, \beta^\star)$ in $\mathbb{P}$-probability. Therefore, for every $\epsilon > 0$, we have

$$\mathbb{P}\left(|\mathbb{P}_n q_n^2 - Pq^2| > 2\epsilon\right) \leq \mathbb{P}\left(|(\mathbb{P}_n - P)q_n^2| > \epsilon\right) + \mathbb{P}\left(|P(q_n^2 - q^2)| > \epsilon\right)$$

$$\leq \mathbb{P}\left(\sup_{q^2 \in \mathcal{Q}_\gamma^2}|(\mathbb{P}_n - P)q^2| > \epsilon\right) + \mathbb{P}\left(\|\hat{h}_{1,n} - h(\alpha^\star, \beta^\star)\|_1 > \gamma\right)$$

$$+\mathbb{P}\left(\|\hat{h}_{2,n} - h_2(\alpha^\star, \beta^\star)\|_1 > \gamma\right) + \mathbb{P}\left(|\phi_n - \phi_\star| > \gamma\right) + \mathbb{P}\left(|P(q_n^2 - q^2)| > \epsilon\right),$$

that tends to zero with n, by the Glivenko–Cantelli property of $\mathcal{Q}_\gamma^2$ and because $\phi_n \to \phi_\star$, $\hat{h}_{1,n} \to h_1(\alpha^\star, \beta^\star)$, $\hat{h}_{2,n} \to h_2(\alpha^\star, \beta^\star)$ and $P(q_n^2 - q^2) \to 0$ in $\mathbb{P}$-probability.

Lemma 1. *Consider a positive function $e(n)$ such that $\lim_{n\to\infty} e(n) = 0$ and $\lim_{n\to\infty} \sqrt{n}\, e(n) > 0$. Define $h_n : E_\gamma \to [0, 1]$, $h_n(\alpha, \beta) = \int \phi(x, \alpha, \beta)\,d\mathbb{P}_n(x)$. Under the assumptions of Theorem 3.1 (or Corollary 4.2) and for every $i \in \{1, \dots, d_\alpha\}$ and every $j \in \{1, \dots, d_\beta\}$, we have*

$$\hat{h}_{1,n,i} := \frac{1}{2e(n)}\big(h_n(\hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) - h_n(\hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n)\big) \longrightarrow \frac{\partial h}{\partial \alpha_i}(\alpha^\star, \beta^\star) \text{ and}$$

$$\hat{h}_{2,n,j} := \frac{1}{2e(n)}\big(h_n(\hat{\alpha}_n, \hat{\beta}_n + e(n)u_j^\beta) - h_n(\hat{\alpha}_n, \hat{\beta}_n - e(n)u_j^\beta)\big) \longrightarrow \frac{\partial h}{\partial \beta_j}(\alpha^\star, \beta^\star)$$

in probability, when $n \to +\infty$.

Proof. By simple algebraic manipulations, we obtain

$$\hat{h}_{1,n,i} = \frac{1}{2\sqrt{n}e(n)}\left\{\mathbb{G}_n\left(\phi(\cdot, \hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) - \phi_\star\right) - \mathbb{G}_n\left(\phi(\cdot, \hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n) - \phi_\star\right)\right\}$$

$$+ \frac{1}{2e(n)}P\left(\phi(\cdot, \hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) - \phi(\cdot, \hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n)\right). \tag{A.2}$$

Using that $(\hat{\alpha}_n, \hat{\beta}_n) \to (\alpha^\star, \beta^\star)$ in probability and that $\lim_{n\to\infty} \sqrt{n}e(n) = \infty$, we get

$$\frac{1}{2\sqrt{n}e(n)}\left\{\mathbb{G}_n\left(\phi(\cdot, \hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) > 0 - \phi_\star\right) - \mathbb{G}_n\left(\phi(\cdot, \hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n) - \phi_\star\right)\right\} = o_{\mathbb{P}^*}(1).$$

Concerning the last term of the r.h.s. of (A.2), use the same Taylor expansion as in the proof of Theorem 3.1 to state

$$\frac{1}{2e(n)}P\left(\phi(\cdot, \hat{\alpha}_n + e(n)u_i^\alpha, \hat{\beta}_n) - \phi(\cdot, \hat{\alpha}_n - e(n)u_i^\alpha, \hat{\beta}_n)\right)$$

$$= \frac{1}{2e(n)}\Big(h_1(\alpha^\star, \beta^\star) \cdot \big(\hat{\alpha}_n + e(n)u_i^\alpha - \alpha^\star\big) - h_1(\alpha^\star, \beta^\star) \cdot \big(\hat{\alpha}_n - e(n)u_i^\alpha - \alpha^\star\big) + o_{\mathbb{P}}\left(e(n)\right)\Big),$$

which tends to $\partial h(\alpha^\star, \beta^\star)/\partial \alpha_i$ in $\mathbb{P}$-probability. The convergence of $\hat{h}_{2,n,j}$ follows analogously. □

A.3. Proof of Corollary 4.2

The proof is a direct consequence of Theorem 3.1. The only point is to check Assumption B3 in this particular case. By the continuous mapping theorem, $\psi(x, \hat{\alpha}_n, \hat{\beta}_n) \to \psi(x, \alpha^\star, \beta^\star)$ $\mathbb{P}$-almost surely for any $x \in \mathbb{R}^d$, since ψ is continuous on E_γ by Assumption B3′. Since $\psi(X, \alpha^\star, \beta^\star)$ has no probability mass at 0, by the definition of strictly locally non-nested models, the Dominated Convergence Theorem tells us that $\mathbf{1}\{\psi(X, \hat{\alpha}_n, \hat{\beta}_n) > 0\}$ tends to $\mathbf{1}\{\psi(X, \alpha^\star, \beta^\star) > 0\}$ in $L_2(P)$, which yields Assumption *B3*.

A.4. Proof of Theorem 6.1

Proof of (i). Use the same notations as in the proof of Theorem 3.1. We know that $\tilde{\mathbb{G}}_n \underset{\xi}{\rightsquigarrow} \mathbb{G}$ by Theorem 3.6.13 in van der Vaart and Wellner (1996) since $\mathcal{H}_\gamma$ is P-Donsker. Simple calculations provide

$$
\begin{aligned}
\frac{\sqrt{n}}{c}(\tilde{C}_n - \hat{C}_n) &= \sqrt{n}c^{-1}\big(\tilde{\mathbb{P}}_n\tilde{\phi}_n - \mathbb{P}_n\phi_n\big) \\
&= \sqrt{n}c^{-1}P\big(\tilde{\phi}_n - \phi_n\big) + \tilde{\mathbb{G}}_n\tilde{\phi}_n + c^{-1}\mathbb{G}_n\big(\tilde{\phi}_n - \phi_n\big) \\
&= \tilde{\mathbb{G}}_n\phi_\star + \sqrt{n}c^{-1}P\big(\tilde{\phi}_n - \phi_n\big) + \tilde{\mathbb{G}}_n\big(\tilde{\phi}_n - \phi_\star\big) + c^{-1}\mathbb{G}_n\big(\tilde{\phi}_n - \phi_n\big).
\end{aligned}
$$

By B2, B4 and B5, a Taylor expansion yields

$$
\begin{aligned}
\sqrt{n}c^{-1}P\big(\tilde{\phi}_n - \phi_n\big) &= \sqrt{n}c^{-1}\big(h(\tilde{\alpha}_n, \tilde{\beta}_n) - h(\hat{\alpha}_n, \hat{\beta}_n)\big) \\
&= \tilde{\mathbb{G}}_n\big(h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2\big) + o_{\mathbb{P}_{XW}^*}(1).
\end{aligned}
$$

Moreover, Lemma 2 (proved below) says that $c^{-1}\mathbb{G}_n\big(\tilde{\phi}_n - \phi_n\big)$ and $\tilde{\mathbb{G}}_n\big(\tilde{\phi}_n - \phi_\star\big)$ are also $o_{\mathbb{P}_{XW}^*}(1)$. Therefore,

$$
\frac{\sqrt{n}}{c}(\tilde{C}_n - \hat{C}_n) = \tilde{\mathbb{G}}_n\Big(\phi_\star + h_1(\alpha^\star, \beta^\star) \cdot s_1 + h_2(\alpha^\star, \beta^\star) \cdot s_2\Big) + o_{\mathbb{P}_{XW}^*}(1).
$$

Now, the result follows by the usual bootstrap convergence.

Proof of (ii). The proof immediately follows from Lemma 3.1 in Bücher and Kojadinovic (2018).

Remark 5. The application of Theorem 3.6.13 in van der Vaart and Wellner (1996) requires an additional measurability assumption on $\mathcal{H}_\gamma$. In most cases, this assumption can be easily verified. Moreover, if we only consider the exchangeable bootstrap schemes described in Kosorok (2008) page 19ff., the measurability assumption on $\mathcal{H}_\gamma$ can even be omitted.

Lemma 2. *Under the assumptions of Theorem 6.1*

(i') $(\tilde{\alpha}_n, \tilde{\beta}_n) \to (\alpha^\star, \beta^\star)$ *in* $\mathbb{P}_{XW}$-*probability*,
(ii') $\mathbb{G}_n\big(\phi(\cdot, \tilde{\alpha}_n, \tilde{\beta}_n) - \phi(\cdot, \hat{\alpha}_n, \hat{\beta}_n)\big) = o_{\mathbb{P}_{XW}^*}(1)$, *and*
(iii') $\tilde{\mathbb{G}}_n\big(\phi(\cdot, \tilde{\alpha}_n, \tilde{\beta}_n) - \phi(\cdot, \alpha^\star, \beta^\star)\big) = o_{\mathbb{P}_{XW}^*}(1)$.

Proof. **Proof of** (i'): Let us just prove that $\tilde{\alpha}_n$ tends to $\alpha^\star$ in $\mathbb{P}_{XW}$-probability, since the result for $\tilde{\beta}_n$ follows similarly. Using Assumption B2 and B5, we have

$$
\tilde{\alpha}_n - \alpha^\star = \tilde{\alpha}_n - \hat{\alpha}_n + \hat{\alpha}_n - \alpha^\star = (\tilde{\mathbb{P}}_n - \mathbb{P}_n)s_1 + o_{\mathbb{P}_{XW}^*}(1).
$$

Define new weights $\tilde{\xi}_{i,n} = n^{-1}\xi_{i,n}$, which remain exchangeable and non-negative. Our assumptions yield $\sum_{i=1}^n \tilde{\xi}_{i,n} = 1$ and $\max_{1 \le i \le n} \tilde{\xi}_{i,n} \to 0$ in $\mathbb{P}_W$-probability. Thus, apply Lemma 3.6.16 in van der Vaart and Wellner (1996) and the conclusion follows.
  Proof of (ii'): For arbitrary constants $\zeta, \pi > 0$, choose $\lambda > 0$ and N large enough such that

$$
\mathbb{P}_X^*\Big(\sup_{f,g \in \mathcal{H}_\gamma : \rho_P(f,g) \le \lambda} \mathbb{G}_n(f - g) > \pi\Big) \le \zeta \text{ for all } n \ge N,
$$

which is possible due to 2.1.8 in van der Vaart and Wellner (1996). Additionally, choose N large enough such that $\mathbb{P}_{XW}\big(\rho_P(\tilde{\phi}_n, \phi_n) > \lambda\big) \le \zeta$ for all $n \ge N$, which is possible due to Assumption B6. We deduce

$$
\begin{aligned}
&\mathbb{P}_{XW}\big(|\mathbb{G}_n(\tilde{\phi}_n - \phi_n)| > \pi\big) \\
&\le \mathbb{P}_X^*\Big(\sup_{f,g \in \mathcal{H}_\gamma : \rho_P(f,g) < \lambda} |\mathbb{G}_n(f - g)| > \pi\Big) + \mathbb{P}_{XW}\big(|\rho_P(\tilde{\phi}_n, \phi_n)| \ge \lambda\big) \le 2\zeta.
\end{aligned}
$$

Since ζ was arbitrary, this yields $\mathbb{P}_{XW}\big(|\mathbb{G}_n(\tilde{\phi}_n - \phi_n)| > \pi\big) \to 0$, i.e. $\mathbb{G}_n(\tilde{\phi}_n - \phi_n)$ is $o_{\mathbb{P}_{XW}^*}(1)$.
  Proof of (iii'): Note that by the Cauchy–Schwarz inequality, we have

$$
\mathbb{E}_P\big[(\tilde{\phi}_n - \phi_\star)^2\big]^{1/2} \le \mathbb{E}_P\big[(\tilde{\phi}_n - \phi_n)^2\big]^{1/2} + \mathbb{E}_P\big[(\phi_n - \phi_\star)^2\big]^{1/2}.
$$

Assumption B3 and B6 imply that $\rho_P(\tilde{\phi}_n, \phi_\star) \to 0$ in $\mathbb{P}_{XW}$-probability. For arbitrary $\pi, \zeta > 0$, choose $\lambda > 0$ and N large enough such that $\mathbb{P}_{XW}\big(\rho_P(\tilde{\phi}_n, \phi_\star) \ge \lambda\big) \le \zeta$ and

$$
\mathbb{P}_{XW}^*\Big(\sup_{f,g \in \mathcal{H}_\gamma : \rho_P(f,g) < \lambda} |\tilde{\mathbb{G}}_n(f - g)| > \pi\Big) \le \zeta,
$$

for all $n \geq N$, which is possible due to the equicontinuity of the bootstrap process $\tilde{\mathbb{G}}_n$ (see Theorem 3.6.13 in van der Vaart and Wellner (1996)). This yields

$$\mathbb{P}_{XW}\big(|\tilde{\mathbb{G}}_n(\tilde{\phi}_n - \phi_\star)| > \pi\big) \leq \mathbb{P}^*_{XW}\Big(\sup_{f,g \in \mathcal{H}_\gamma ; \rho_P(f,g) < \lambda} |\tilde{\mathbb{G}}_n(f - g)| > \pi\Big) + \mathbb{P}_{XW}\big(\rho_P(\tilde{\phi}_n, \phi_\star) \geq \lambda\big) \leq 2\zeta.$$

We deduce that $\tilde{\mathbb{G}}_n(\tilde{\phi}_n - \phi_\star) = o_{\mathbb{P}^*_{XW}}(1)$. □

A.5. Proof of Corollary 6.2

Define $\tilde{\phi}_n^C := \phi^C(X, \tilde{\alpha}_n, \tilde{\beta}_n)$. To prove the corollary, it is sufficient to prove that Assumption B6′ implies B6. Let us show that $\mathbb{E}_P\big[(\tilde{\phi}_n^C - \phi_\star^C)^2\big] \to 0$ in $\mathbb{P}_{XW}$-probability. Observe that

$$\mathbb{E}_P\big[(\tilde{\phi}_n^C - \phi_\star^C)^2\big] = P\big(\psi(X, \alpha^\star, \beta^\star) > 0, \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big) + P\big(\psi(X, \alpha^\star, \beta^\star) < 0, \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) > 0\big)$$

and note that the map $(\tilde{\alpha}_n, \tilde{\beta}_n) \to P(\psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) \in A)$ is measurable for any Borel set A. Choose $\pi \in (0,1)$ and a sufficiently small $\epsilon > 0$ such that $P(|\psi(X, \alpha^\star, \beta^\star)| \leq \epsilon) \leq \pi/2$. This is possible due to the strict local non-nestedness of the proposed models f and g. Indeed, the strict local non-nestedness assumption implies $0 = P(|\psi(X, \alpha^\star, \beta^\star)| = 0) = \lim_{n\to\infty} P(|\psi(X, \alpha^\star, \beta^\star)| \leq 1/n)$, by the continuity of measures. Additionally, due to B6′, we can choose a compact set $A_\pi \subset \mathbb{R}^d$ and $\delta > 0$, $P(X \in A_\pi) \geq 1 - \pi/4$, s.t. $\|(\alpha, \beta) - (\alpha^\star, \beta^\star)\|_1 \leq \delta$ and $x \in A_\pi$ implies $|\psi(x, \alpha, \beta) - \psi(x, \alpha^\star, \beta^\star)| < \epsilon$. Moreover, for an arbitrary $\nu > 0$ and the latter constant δ, choose N large enough such that

$$\mathbb{P}_{XW}\big(\|(\tilde{\alpha}_n, \tilde{\beta}_n) - (\alpha^\star, \beta^\star)\|_1 > \delta\big) \leq \nu,$$

for all $n \geq N$, which is stated in the proof of Theorem 6.1. Therefore, we get

$$\mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) > 0; \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big) > \pi\Big)$$
$$\leq \mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) > \epsilon; \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big) > \frac{\pi}{2}\Big) + \mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) \in (0, \epsilon]\big) > \frac{\pi}{2}\Big)$$
$$\leq \mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) > \epsilon; \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big) > \frac{\pi}{2}; \|(\tilde{\alpha}_n, \tilde{\beta}_n) - (\alpha^\star, \beta^\star)\|_1 \leq \delta\Big) + \nu + 0$$
$$\leq \mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) > \epsilon; \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0; X \in A_\pi\big) > \frac{\pi}{4}; \|(\tilde{\alpha}_n, \tilde{\beta}_n) - (\alpha^\star, \beta^\star)\|_1 \leq \delta\Big)$$
$$+ \mathbb{P}_{XW}\Big(P\big(X \notin A_\pi\big) > \frac{\pi}{4}\Big) + \nu. \tag{A.3}$$

By construction, $P(X \notin A_\pi) \leq \pi/4$ and the first term on r.h.s. of (A.3) is zero. As a consequence, $\mathbb{P}_{XW}\Big(P\big(\psi(X, \alpha^\star, \beta^\star) > 0; \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big) > \pi\Big) \leq \nu$. Since ν was arbitrary, $P\big(\psi(X, \alpha^\star, \beta^\star) > 0, \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) < 0\big)$ tends to zero in $\mathbb{P}_{XW}$-probability. Similarly, $P\big(\psi(X, \alpha^\star, \beta^\star) < 0, \psi(X, \tilde{\alpha}_n, \tilde{\beta}_n) > 0\big)$ tends to zero in $\mathbb{P}_{XW}$-probability. Combining such arguments, we have proven that $\rho_P^2(\tilde{\phi}_n^C, \phi_\star^C) = \mathbb{E}_P\big[(\tilde{\phi}_n^C - \phi_\star^C)^2\big]$ tends to zero in $\mathbb{P}_{XW}$-probability.

A.6. Proof of Proposition 6.3

It is sufficient to show that

$$\lim_{n,M \to \infty} \left| \mathbb{E}_{\mathbb{P}_{XW}}\left[f\Big(\frac{1}{Mc^2}\sum_{j=1}^M n\big(\tilde{C}_n^{(j)} - \hat{C}_n\big)^2\Big) - f\big(\sigma^2\big) \right] \right| = 0,$$

for every map $f : \mathbb{R} \to \mathbb{R}$ that is uniformly continuous and bounded. This would imply the claim, since convergence in probability to a constant is equivalent to weak convergence to a constant.

Theorem 6.1 and Bücher and Kojadinovic (2018, Lemma 2.2) imply that $\mathbb{E}_{\mathbb{P}_{XW}}\big[(C_n^{(1)})^2\big] \to \sigma^2$ and $\mathbb{E}_{\mathbb{P}_{XW}}\big[C_n^{(1)}\big] \to 0$, since $\big((C_n^{(1)})^2\big)_{n \in \mathbb{N}}$ is uniformly integrable by Assumption B7. Moreover, for any $\epsilon > 0$, we can choose a $\delta > 0$ small enough such that f is (ϵ, δ)-uniformly continuous, Lemma 3 implies that we can find $\overline{M}(\epsilon, \delta) \in \mathbb{N}$ and $\overline{n}(\epsilon, \delta) \in \mathbb{N}$ such that for any $n \geq \overline{n}$ and $M \geq \overline{M}$

$$\left| \mathbb{E}_{\mathbb{P}_{XW}}\left[f\Big(\frac{1}{M}\sum_{j=1}^M (C_n^{(j)})^2\Big) - f\big(\sigma^2\big) \right] \right|$$
$$\leq \left| \mathbb{E}_{\mathbb{P}_{XW}}\left[f\Big(\frac{1}{M}\sum_{j=1}^M (C_n^{(j)})^2\Big) - f\Big(\mathbb{E}_{\mathbb{P}_{XW}}\big[(C_n^{(1)})^2\big]\Big) \right] \right| + \left| f\Big(\mathbb{E}_{\mathbb{P}_{XW}}\big[(C_n^{(1)})^2\big]\Big) - f\big(\sigma^2\big) \right|$$

$$\leq \mathbb{E}_{\mathbb{P}_{XW}}\left[\left|f\left(\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2\right) - f\left(\mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\right]\right)\right| \mathbf{1}\left\{\left|\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 - \mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\right]\right| > \delta\right\}\right] + 2\epsilon$$

$$\leq (C+2)\epsilon,$$

since f is bounded by a constant $C/2 > 0$. This proves the claim, since $\epsilon > 0$ was arbitrary.

Lemma 3. *For all $\delta, \epsilon > 0$ there exist $\overline{M}(\epsilon, \delta), \overline{n}(\epsilon, \delta) \in \mathbb{N}$ such that for all $M \geq \overline{M}$ we have that*

$$\sup_{n \geq \overline{n}} \mathbb{P}_{XW}\left(\left|\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 - \mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\right]\right| > \delta\right) < \epsilon.$$

Proof. Set $\left(C_n^{(0)}\right) := \sqrt{n}\left(\hat{C}_n - P\phi^\star\right)$. Denote the empirical measure $M^{-1}\sum_{i=1}^{M}\delta_{C_n^{(j)}} =: P_M^{C_n}$, the law of $C_n^{(0)}$ as $P_{C_n^{(0)}}$, the law of $C_n^{(1)}$ as $P_{C_n^{(1)}}$ and let BL^K be the set of bounded Lipschitz functions with Lipschitz constant $\leq 2K$. W.l.o.g. choose $\epsilon, \delta \in (0, 1)$. Let $K(\epsilon, \delta)$ be large enough such that

$$\mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2 \mathbf{1}\{|C_n^{(1)}| \geq K\}\right] < \frac{\delta}{24}\epsilon$$

for all $n \in \mathbb{N}$, which is possible since $\left(\left(C_n^{(1)}\right)^2\right)_{n \in \mathbb{N}}$ is uniformly integrable. Moreover, depending on the chosen K, choose $\overline{n}(\epsilon, \delta, K)$ and $\overline{M}(\epsilon, \delta, K)$ large enough such that for all $n \geq \overline{n}$ and $M \geq \overline{M}$ we have

$$\sup_{f \in BL^K}\left|\int f(x)\mathrm{d}\left(P_{C_n^{(0)}} - P_{C_n^{(1)}}\right)(x)\right| < \frac{\delta}{4}, \text{ and}$$

$$\mathbb{P}_{XW}\left(\sup_{f \in BL^K}\left|\int f(x)\mathrm{d}\left(P_M^{C_n} - P_{C_n^{(0)}}\right)(x)\right| > \frac{\delta}{4}\right) < \frac{\epsilon}{3},$$

which is possible due to [Bücher and Kojadinovic (2018](#), Lemma 2.2 (a)+(d)), since $C_n^{(1)}$ and $C_n^{(0)}$ have the same weak limit $\mathcal{N}(0, \sigma^2)$. Note that

$$\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 - \mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\right] = \int x^2 \mathrm{d}P_M^{C_n}(x) - \int x^2 \mathrm{d}P_{C_n^{(1)}}(x)$$

$$= \int \min\{x^2, K^2\}\mathrm{d}\left(P_M^{C_n} - P_{C_n^{(0)}}\right)(x) + \int \min\{x^2, K^2\}\mathrm{d}\left(P_{C_n^{(0)}} - P_{C_n^{(1)}}\right)(x)$$

$$+ \int_{|x|>K} x^2 \mathrm{d}\left(P_M^{C_n} - P_{C_n^{(1)}}\right)(x) + \int_{|x|>K} K^2 \mathrm{d}\left(P_{C_n^{(1)}} - P_M^{C_n}\right)(x).$$

Thus, for $n \geq \overline{n}$ and $M \geq \overline{M}$, we have

$$\mathbb{P}_{XW}\left(\left|\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 - \mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2\right]\right| > \delta\right)$$

$$\leq \mathbb{P}_{XW}\left(\left|\int \min\{x^2, K^2\}\mathrm{d}\left(P_M^{C_n} - P_{C_n^{(0)}}\right)(x)\right| > \frac{\delta}{4}\right) + \mathbb{P}_{XW}\left(\left|\int \min\{x^2, K^2\}\mathrm{d}\left(P_{C_n^{(0)}} - P_{C_n^{(1)}}\right)(x)\right| > \frac{\delta}{4}\right)$$

$$+ \mathbb{P}_{XW}\left(\left|\int_{|x|>K} x^2 \mathrm{d}\left(P_M^{C_n} - P_{C_n^{(1)}}\right)(x)\right| > \frac{\delta}{4}\right) + \mathbb{P}_{XW}\left(\left|\int_{|x|>K} K^2 \mathrm{d}\left(P_{C_n^{(1)}} - P_M^{C_n}\right)(x)\right| > \frac{\delta}{4}\right)$$

$$\leq \mathbb{P}_{XW}\left(\sup_{f \in BL^K}\left|\int f(x)\mathrm{d}\left(P_M^{C_n} - P_{C_n^{(0)}}\right)(x)\right| > \frac{\delta}{4}\right) + \mathbb{P}_{XW}\left(\sup_{f \in BL^K}\left|\int f(x)\mathrm{d}\left(P_{C_n^{(0)}} - P_{C_n^{(1)}}\right)(x)\right| > \frac{\delta}{4}\right)$$

$$+ 2\,\mathbb{P}_{XW}\left(\left|\int_{|x|>K} x^2 \mathrm{d}P_M^{C_n}(x)\right| > \frac{\delta}{8}\right) + 2\,\mathbb{P}_{XW}\left(\left|\int_{|x|>K} x^2 \mathrm{d}P_{C_n^{(1)}}(x)\right| > \frac{\delta}{8}\right)$$

$$\leq \frac{\epsilon}{3} + 2\,\mathbb{P}_{XW}\left(\left|\int_{|x|>K} x^2 \mathrm{d}P_M^{C_n}(x)\right| > \frac{\delta}{8}\right),$$

when $n \geq \bar{n}$ and $M \geq \overline{M}$. For our chosen K, n we obtain

$$
2\,\mathbb{P}_{XW}\left(\left|\int_{x \geq K} x^2 \mathrm{d}P_M^{C_n}(x)\right| > \frac{\delta}{8}\right) \leq \frac{16}{\delta}\,\mathbb{E}_{\mathbb{P}_{XW}}\left[\int_{|x| \geq K} x^2 \mathrm{d}P_M^{C_n}(x)\right]
$$

$$
= \frac{16}{\delta}\,\mathbb{E}_{\mathbb{P}_{XW}}\left[\mathbb{E}_{\mathbb{P}_{XW}}\left[\frac{1}{M}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 \mathbf{1}\{|C_n^{(j)}| > K\}\,\Big|(X_1, \ldots, X_n)\right]\right]
$$

$$
\stackrel{\star}{=} \frac{16}{\delta}\,\mathbb{E}_{\mathbb{P}_{XW}}\left[\left(C_n^{(1)}\right)^2 \mathbf{1}\{|C_n^{(1)}| > K\}\right] < \frac{16}{\delta}\cdot\frac{\delta}{24}\epsilon = \frac{2}{3}\epsilon ,
$$

where $\star$ uses that $\left(C_n^{(j)}\right)_{j \in \mathbb{N}}$ are conditionally i.i.d. given $(X_1, \ldots, X_n)$. Since ϵ was arbitrary, the claim is proven. $\square$

It is interesting to note that our proof of the consistency of $\hat{\sigma}_M^2$ also holds in a more general setting: denote by $\left(\tilde{S}_n^{(j)}\right)_{j \in \mathbb{N}}$ a sequence of bootstrap versions of some statistic S_n under the constraint that the vectors of bootstrap weights are independently drawn and satisfy the assumptions stated in the beginning of Section 6.

Proposition A. *Assume there exists a random variable Z with finite second moment s.t. $S_n \rightsquigarrow Z$ and $\tilde{S}_n^{(1)} \underset{\xi}{\rightsquigarrow} Z$. If $\left(\left(\tilde{S}_n^{(1)}\right)^2\right)_{n \in \mathbb{N}}$ is uniformly integrable, then we have*

$$
\frac{1}{M}\sum_{j=1}^{M}\left(\tilde{S}_n^{(j)} - \frac{1}{M}\sum_{i=1}^{M}\tilde{S}_n^{(i)}\right)^2 \xrightarrow{n, M \to \infty} \mathrm{Var}(Z) \text{ in } \mathbb{P}_{XW}\text{-probability.}
$$

The proof of the general case follows exactly the same thoughts as the proof of Proposition 6.3. Simply replace $\mathbb{E}\left[\left(C_n^{(1)}\right)^2\right]$ by $\mathrm{Var}\left(C_n^{(1)}\right)$, and also prove that $\left|M^{-1}\sum_{j=1}^{M} C_n^{(j)} - \mathbb{E}\left[C_n^{(1)}\right]\right| \to 0$ in $\mathbb{P}_{XW}$-probability, which follows by the same steps as $\left|M^{-1}\sum_{j=1}^{M}\left(C_n^{(j)}\right)^2 - \mathbb{E}\left[\left(C_n^{(1)}\right)^2\right]\right| \to 0$ in $\mathbb{P}_{XW}$-probability.

To the best of our knowledge the consistency of such asymptotic variance estimation procedures has not been stated elsewhere in the literature with the same degree of generality.

Appendix B. The Clarke test statistic is not binomial distributed

We present a short counterexample to illustrate that $\hat{B}_n$ is generally not Binomial distributed when $\mathcal{H}_0^C$ is satisfied. In other words, the conclusion drawn on the distribution of $\hat{B}_n$ in Section 2.2 of Clarke (2007) is incorrect.

Let the true distribution P follow a univariate normal distribution with mean μ_0 and fixed variance 1. We compare the density of a normal distribution with fixed variance σ_f^2 and the density of a normal distribution with fixed variance σ_g^2 while estimating the mean for both of these families. This translates to the families of densities $f(x, \alpha) = (2\pi\sigma_f^2)^{-1/2}\exp\left(-(x-\alpha)^2/2\sigma_f^2\right)$ and $g(x, \beta) = (2\pi\sigma_g^2)^{-1/2}\exp\left(-(x-\beta)^2/2\sigma_g^2\right)$. Note that the two models are strictly locally non-nested according to Definition 4.1.

In the following, we will choose $\sigma_f \neq \sigma_g$ not equal to one, such that the null hypothesis of the Clarke test is satisfied. Before fixing σ_f and σ_g, we check that $\alpha^\star$ and $\beta^\star$ are both equal to μ_0. Indeed, in the case of $\alpha^\star$, we have

$$
\mathbb{E}\left[\log\bigl(f(X, \alpha)\bigr)\right] = \mathbb{E}\left[-\log(\sqrt{2\pi\sigma_f^2}) - \frac{(X-\alpha)^2}{2\sigma_f^2}\right]
$$

$$
= -\log(\sqrt{2\pi\sigma_f^2}) - \frac{\mathrm{Var}(X)}{2\sigma_f^2} - \frac{1}{2\sigma_f^2}(\mu_0 - \alpha)^2 \leq -\log(\sqrt{2\pi\sigma_f^2}) - \frac{\mathrm{Var}(X)}{2\sigma_f^2},
$$

where the last inequality is an equality iff we choose $\alpha = \mu_0$. Second, estimators of the pseudo-true values $\alpha^\star$ and $\beta^\star$ in the case of known variances are obtained by pseudo maximum likelihood inference: $\hat{\beta}_n = \hat{\alpha}_n = \bar{X}$, the usual empirical mean. Third, we choose σ_f and σ_g such that the null hypothesis $\mathcal{H}_0^C$ in (3) is satisfied. To this goal, calculate the probability of the set

$$
\left\{\log\left(\frac{f(X, \alpha^\star)}{g(X, \beta^\star)}\right) > 0\right\} = \left\{\frac{(X-\beta^\star)^2\sigma_f^2 - (X-\alpha^\star)^2\sigma_g^2}{2\sigma_f^2\sigma_g^2} > \log\left(\frac{\sigma_f}{\sigma_g}\right)\right\}.
$$

Replacing $\alpha^\star$ and $\beta^\star$ with μ_0, and assuming w.l.o.g. that $\sigma_g < \sigma_f$, we get

$$
\left\{\frac{(X-\beta^\star)^2\sigma_f^2 - (X-\alpha^\star)^2\sigma_g^2}{2\sigma_f^2\sigma_g^2} > \log\left(\frac{\sigma_f}{\sigma_g}\right)\right\} = \left\{(X-\mu_0)^2 > \log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}\right\}.
$$

Since $(X - \mu_0)$ is standard normally distributed, we obtain

$$P\left(\log\left(\frac{f(X, \alpha^\star)}{g(X, \beta^\star)}\right) > 0\right) = 2\Phi\left(-\sqrt{\log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}}\right).$$

Consider $\sigma_g =: \overline{\sigma}_g$ as fixed and set $(\sigma_f/\sigma_g)^2 = t$. Noting that the map $t \mapsto t\log t/(t-1)$, $t > 0$ attains all values in $\mathbb{R}_+$, we can find $\overline{\sigma}_f$ such that

$$-\left\{\log(\overline{\sigma}_f/\overline{\sigma}_g)\overline{\sigma}_f^2\overline{\sigma}_g^2/(\overline{\sigma}_f^2 - \overline{\sigma}_g^2)\right\}^{1/2} = \Phi^{-1}(1/4).$$

Then, with such values, $\mathcal{H}_0^C$ is satisfied. For instance, if $\bar{\sigma}_g = 1/2$, we get an approximated value of $\overline{\sigma}_f \approx 0.98$. From now on, consider that we have chosen $\sigma_f \neq 1$, $\sigma_g \neq 1$ and $\sigma_f > \sigma_g$ such that

$$P\left(\left\{(X - \mu_0)^2 > \log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}\right\}\right) = \frac{1}{2}.$$

Therefore, the null hypothesis of the Clarke test is satisfied for the chosen values of σ_f and σ_g.

In Clarke (2007), it is stated that the statistic $\hat{B}_n$ is Binomial distributed with parameter $p = 0.5$. We show that $\hat{B}_n$ is generally not even Binomial distributed for any $p \in [0, 1]$. For $n = 2$ and the considered example, the test statistic $\hat{B}_2$ takes the form:

$$\mathbf{1}\left\{\log\left(\frac{f(X_1, \hat{\alpha}_2)}{g(X_1, \hat{\beta}_2)}\right) > 0\right\} + \mathbf{1}\left\{\log\left(\frac{f(X_2, \hat{\alpha}_2)}{g(X_2, \hat{\beta}_2)}\right) > 0\right\},$$

which is equal to

$$\mathbf{1}\left\{(X_1 - \bar{X})^2 > \log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}\right\} + \mathbf{1}\left\{(X_2 - \bar{X})^2 > \log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}\right\}.$$

The equality $(X_1 - \bar{X})^2 = (X_2 - \bar{X})^2$ yields

$$\hat{B}_2 = 2 \times \mathbf{1}\left\{\left(\frac{X_1}{2} - \frac{X_2}{2}\right)^2 > \log\left(\frac{\sigma_f}{\sigma_g}\right)\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 - \sigma_g^2}\right\},$$

which takes values in $\{0, 2\}$ and is clearly not Binomial distributed.

References

Adrian, T., Brunnermeier, M., 2016. CoVaR. Am. Econ. Rev. 106 (7), 1705–1741.

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. J. Bank. Finance 32 (8), 1541–1551.

Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle. Springer New York, pp. 199–213.

Andrews, D.W.K., 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. Econometrica 62 (1), 43–72.

Angelidis, T., Benos, A., Degiannakis, S., 2004. The use of GARCH models in VaR estimation. Statist. Methodol. 1 (1-2), 105–128.

Barth, M.E., Gow, I.D., Taylor, D.J., 2012. Why do pro forma and street earnings not reflect changes in GAAP? Evidence from SFAS 123R. Rev. Account. Stud. 17 (3), 526–562.

Bonnal, L., Fougère, D., Sérandon, A., 1997. Evaluating the impact of French employment policies on individual labour market histories. Rev. Econ. Stud. 64 (4), 683–713.

Breckling, J., Chambers, R., 1988. M-quantiles. Biometrika 75 (4), 761–771.

Brooks, C., 2014. Introductory Econometrics For Finance. Cambridge University Press.

Brück, F., 2019. Clarke's Test for Non-Nested Model Comparison (Master's thesis). Technical University Munich, URL https://mediatum.ub.tum.de/node?id=1472565.

Bücher, A., Kojadinovic, I., 2018. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. J. Theor. Probab. 32 (3), 1145–1165.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference. Springer New York.

Caballero, R.J., Engel, E.M., 1999. Explaining investment dynamics in US manufacturing: A generalized (S, s) approach. Econometrica 67 (4), 783–826.

Cameron, S.V., Heckman, J.J., 1998. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. J. Polit. Economy 106 (2), 262–333.

Chen, X., Fan, Y., 2005. Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. Can. J. Statist. 33 (3), 389–414.

Chen, X., Hong, H., Shum, M., 2007. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. J. Econometrics 141 (1), 109–140.

Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71 (5), 1591–1608.

Clarke, K.A., 2007. A simple distribution-free test for nonnested model selection. Polit. Anal. 15 (3), 347–363.

Czado, C., Schepsmeier, U., Min, A., 2012. Maximum likelihood estimation of mixed C-vines with application to exchange rates. Statist. Model. 12 (3), 229–255.

Czajkowski, M., Buszko-Briggs, M., Hanley, N., 2009. Valuing changes in forest biodiversity. Ecol. Econ. 68 (12), 2910–2917.

Douven, I., Schupbach, J.N., 2015. Probabilistic alternatives to Bayesianism: The case of explanationism. Front. Psychol. 6, 459.

Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. 1 (2), 279–290.

Erhardt, V., Czado, C., 2012. Modeling dependent yearly claim totals including zero claims in private health insurance. Scand. Actuarial J. 2012 (2), 106–129.

Fafchamps, M., 1993. Sequential labor decisions under uncertainty: An estimable household model of West-African farmers. Econometrica 61 (5), 1173–1197.

Ghosh, M., Parr, W.C., Singh, K., Babu, G.J., 1984. A note on bootstrapping the sample median. Ann. Statist. 12 (3), 1130–1135.

Harrison, G.W., Martínez-Correa, J., Swarthout, J.T., 2015. Reduction of compound lotteries with objective probabilities: Theory and evidence. J. Econ. Behav. Organ. 119, 32–55.

Hey, J.D., Panaccione, L., 2011. Dynamic decision making: What do people do? J. Risk Uncertain. 42 (2), 85–123.

Hill, R.C., Griffiths, W.E., Lim, G.C., 2018. Principles of Econometrics. John Wiley & Sons.

Hofbauer, C.J., Whelan, S.F., Hirschler, M., Allacher, P., Horling, F.M., Lawo, J.-P., Oldenburg, J., Tiede, A., Male, C., Windyga, J., et al., 2015. Affinity of FVIII-specific antibodies reveals major differences between neutralizing and nonneutralizing antibodies in humans. Blood 125 (7), 1180–1188.

Hong, H., Preston, B., 2012. Bayesian averaging, prediction and nonnested model selection. J. Econometrics 167 (2), 358–369.

Horowitz, J.L., 1998. Bootstrap methods for median regression models. Econometrica 66 (6), 1327–1351.

Hsu, Y.-C., Shi, X., 2017. Model selection tests for conditional moment restriction models. Econometrics J. 20 (1), 52–85.

Ichimura, H., Lee, S., 2010. Characterization of the asymptotic distribution of semiparametric M-estimators. J. Econometrics 159 (2), 252–266.

Imai, K., Tingley, D., 2012. A statistical method for empirical testing of competing theories. Am. J. Polit. Sci. 56 (1), 218–236.

Karaivanov, A., Townsend, R.M., 2014. Dynamic financial constraints: Distinguishing mechanism design from exogenously incomplete regimes. Econometrica 82 (3), 887–959.

Kendall, C., Nannicini, T., Trebbi, F., 2015. How do voters respond to information? Evidence from a randomized campaign. Am. Econ. Rev. 105 (1), 322–353.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.

Koliai, L., 2016. Extreme risk modeling: an EVT-pair-copulas approach for financial stress tests. J. Bank. Finance 70, 1–22.

Kosorok, M.R., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Statist. 22 (1), 79–86.

Lavergne, P., Vuong, Q.H., 1996. Nonparametric selection of regressors: The nonnested case. Econometrica 64 (1), 207–219.

Leshno, M., Levy, H., 2002. Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. Manage. Sci. 48 (8), 1074–1085.

Li, T., 2009. Simulation based selection of competing structural econometric models. J. Econometrics 148 (2), 114–123.

Liao, Z., Shi, X., 2020. A nondegenerate Vuong test and post selection confidence intervals for semi/nonparametric models. Quant. Econ. 11 (3), 983–1017.

Markwat, T., Kole, E., Van Dijk, D., 2009. Contagion as a domino effect in global stock markets. J. Bank. Finance 33 (11), 1996–2012.

Martin, L.W., Vanberg, G., 2014. Parties and policymaking in multiparty governments: The legislative median, ministerial autonomy, and the coalition compromise. Am. J. Polit. Sci. 58 (4), 979–996.

Monfort, A., 1996. A reappraisal of misspecified econometric models. Econom. Theory 12 (4), 597–619.

Murphy, K.M., Topel, R.H., 1985. Estimation and inference in two-step econometric models. J. Bus. Econ. Statist. 3 (4), 370–379.

New, B., Grainger, J., 2011. On letter frequency effects. Acta Psychol. 138 (2), 322–328.

Newey, W.K., 1984. A method of moments interpretation of sequential estimators. Econ. Lett. 14 (2-3), 201–206.

Newey, W.K., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55 (4), 819–847.

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70 (3), 971–1005.

Olson, D.L., Wu, D., 2017. Predictive Data Mining Models. Springer.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.

Panagiotelis, A., Czado, C., Joe, H., 2012. Pair copula constructions for multivariate discrete data. J. Amer. Statist. Assoc. 107 (499), 1063–1072.

Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. Econometrics J. 5 (1), 1–39.

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46 (6), 1273–1291.

Schennach, S.M., Wilhelm, D., 2017. A simple parametric model selection test. J. Amer. Statist. Assoc. 112 (520), 1663–1674.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

Shi, X., 2015a. Model selection tests for moment inequality models. J. Econometrics 187 (1), 1–17.

Shi, X., 2015b. A nondegenerate Vuong test. Quant. Econ. 6 (1), 85–121.

Van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge University Press.

van der Vaart, A.W., Wellner, J.A., 1996. Weak convergence and empirical processes. Springer New York.

Verbeeten, F.H., Vijn, P., 2010. Are brand-equity measures associated with business-unit financial performance? Empirical evidence from the Netherlands. J. Account. Audit. Finance 25 (4), 645–671.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57 (2), 307–333.

Whang, Y.-J., 2006. Smoothed empirical likelihood methods for quantile regression models. Econom. Theory 22 (2), 173–205.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50 (1), 1–25.

White, H., 1996. Estimation, Inference and Specification Analysis. Cambridge University Press.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Zhou, X., Lin, H., Zhang, S., Ren, J., Wang, Z., Zhang, Y., Wang, M., Zhang, Q., 2016. Effects of climatic factors on plasma lipid levels: A 5-year longitudinal study in a large chinese population. J. Clin. Lipidol. 10 (5), 1119–1128.