FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Juraj Bodík

Detection of causality in time series using extreme values

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	doc. RNDr. Zbyněk Pawlas, Ph.D.
Consultant of the master thesis:	RNDr. Milan Paluš, DrSc.
Study programme:	Mathematics
Study branch:	Probability, Mathematical Statistics and Econometrics

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I want to thank all my professors and teachers, that were trying to keep high standards of teaching despite the pandemic. Foremost, I would like to thank doc. Zbyněk Pawlas, who was always willing to help and answer my questions. I am very grateful that he was also my supervisor of this thesis, that he spent much of his time with me, for his patience with my English and all of his advice on every part of the thesis.

Next, I want to thank doc. Milan Paluš, who introduced me to the fascinating topic of causal inference. Moreover, he offered me great resources and advice about possible applications essential for my master thesis completion.

Finally, I want to thank my parents, friends, new Swiss colleagues, the cooking crew, and the Trojsten community. Due to them, I enjoyed this pandemic time much more than most of the world.

Title: Detection of causality in time series using extreme values

Author: Juraj Bodík

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zbyněk Pawlas, Ph.D.

Consultant: RNDr. Milan Paluš, DrSc.

Abstract: This thesis is dealing with the following problem: Let us have two stationary time series with heavy-tailed marginal distributions. We want to detect whether they have a causal relation, i.e. if a change in one of them causes a change in the other. The question of distinguishing between causality and correlation is essential in many different science fields. Usual methods for causality detection are not well suited if the causal mechanisms only manifest themselves in extremes. In this thesis, we propose a new method that can help us in such a nontraditional case distinguish between correlation and causality. We define the so-called causal tail coefficient for time series, which, under some assumptions, correctly detects the asymmetrical causal relations between different time series. We will rigorously prove this claim, and we also propose a method on how to statistically estimate the causal tail coefficient from a finite number of data. The advantage is that this method works even if nonlinear relations and common ancestors are present. Moreover, we will mention how our method can help detect a time delay between the two time series. Finally, we will show some simulations and an actual application of how our method performs in practice.

V tejto práci riešime nasledovný problém: Máme dve stacionárne časové rady, ktorých marginálne distribúcie majú ťažké chvosty. My chceme zistiť, či majú kauzálny vzťah, teda či zmena v jednej z nich spôsobí zmenu v druhej. Otázka, či náhodné premenné majú kauzálny súvis alebo sú iba korelované, je dôležitá v mnohých oblastiach vedy. Bežné metódy na detekciu kauzalít nefungujú dobre, ak sa vzájomné vzťahy prejavujú výhradne pri extrémnych hodnotách. V tejto práci navrhneme nový spôsob, ako v takomto netradičnom prípade rozlišovať medzi koreláciou a kauzalitou. Definujeme si tzv. kauzálny chvostový koeficient pre časové rady, ktorý za istých predpokladov detekuje asymetrické kauzálne vzťahy medzi dvoma časovými radami. Toto tvrdenie rigorózne dokážeme a ešte navrhneme spôsob akým kauzálny chvostový koeficient štatisticky odhadneme iba z konečného množstva dát. Výhodou je, že táto metóda funguje aj pri nelineárnych vzťahoch medzi časovými radami a aj za prítomnosti spoločnej príčiny. Navyše, spomenieme spôsob akým táto metóda môže pomôcť pri zisťovaní časového posunu medzi dvoma časovými radami. Konečne, na simuláciách a aplikácii ukážeme, ako táto metóda funguje v praxi.

Keywords: Granger causality, Causal inference, Nonlinear time series, VAR process, Extremal value theory, Heavy tails

Contents

Introduction

0.0.1 Background

The ultimate goal of causal inference is to understand relations between random variables and to predict future values. It can be used in almost every scientific field. In hydro-meteorology, to predict the temperature or the amount of rainfall and understand what causes the sudden changes. In medicine, to understand the spread of epileptic seizures in different regions of a brain. In economy, to better predict the financial stocks and exchange rates.

In medicine, to test whether a vaccine or some drug causes a health improvement. In biology, to predict which DNA sequence causes some mutations. In the economy, to better predict the financial stocks.

A neutral definition is notoriously hard to provide, since every aspect of causation has received substantial debate. The task of causal inference divides into two major classes: Causal inference over random variables and a causal inference over time series. Probably the first person that provided the mathematical definition of causality for time series was Wiener [1956]. His definition can be simply explained in words – one time series causes the second one, if the knowledge of its previous values can help in the prediction. The introduction of the concept of causality into the practice, namely into analyses of data observed in consecutive time instants, time series, is due to Granger [1969], Nobel prize winner in economy. To this day, there are several different definitions, each used in a different field. Maybe the most well-known approach comes from an information theory, which uses entropy and mutual information to determine some properties of dynamical systems and complex systems.

0.0.2 Heavy tails

Despite the fact that many economic and environmental data are known to follow a heavy-tailed distribution, there is almost no literature that deals with the causal inference for such a case. Usual methods use some regression or entropy estimation, which relies on the assumption of a finite expected value and variance. To our knowledge, all of them require linear relations (i.e. assuming a simple VAR model). Therefore, such a problem is still an open one. This thesis provides a method that works even for some nonlinear class with heavy-tailed marginal distributions.

0.0.3 Example

We give an example of a typical time series, with which we will deal in our thesis. Let $(X,Y)^\top = ((X_t,Y_t)^\top, t \in \mathbb{Z})$ be a bivariate strictly stationary time series, defined by the following recurrent relations

$$X_t = \frac{1}{2}X_{t-1} + \varepsilon_t^X,$$
$$Y_t = \frac{1}{2}Y_{t-1} + \sqrt{X_{t-5}} + \varepsilon_t^Y,$$

where $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} Pareto(1,1)$ [1]. A sample realization of such a model is in Figure 1. Here, X causes Y (in the Granger sense), simply because the knowledge of X can help in the prediction of the future values of Y (note that it is not true for the other direction).

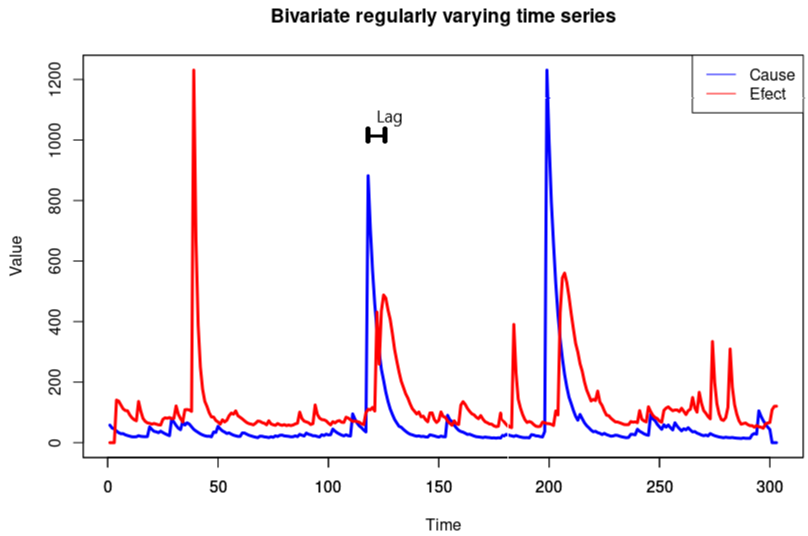

Figure 1: The figure represents a sample realisation of the $(X, Y)^\top$ time series from Subsection 0.0.3 (X is the blue one and Y is red the one). We can see that large values of the blue time series cause large values of the red time series, but not the other way. Here, it is easy to see that the blue is causing the red. The lag represents the time delay between the time series, which is equal to 5 in this case.

0.0.4 Main idea

Consider we have data such as in Figure 1; we want to detect a causal relationship between those time series. There is (at least in this realisation) an evident asymmetry between the two time series in the extremes. If the blue one is extremely large, then the red one will also be extremely large (see the second and third "jump"). However, if the red one is extremely large, the blue one is not necessarily extremely large (see the first red "jump"). Therefore, an extreme of X causes an extreme of Y and not the other way around in an intuitive sense.

We will put this simple idea into the mathematical language. The main problem is the *time lag* (or *time delay*). An extreme event of X does not mean an immediate extreme event of Y – it takes some time for the information from X to influence Y (in this artificial example, it takes exactly 5 time units).

Therefore, we propose the following coefficient,

$$\Gamma^{time}_{X,Y}(q) := \lim_{u \to 1^-} \mathbb{E}\left[\max\{F_Y(Y_0), \ldots, F_Y(Y_q)\} \mid F_X(X_0) > u\right],$$

[1] $\varepsilon_t^X, \varepsilon_t^Y$ are iid (independent and identically distributed), following a Pareto distribution with parameters equal to 1. Distribution function of a $Pareto(a, b)$ random variable is in the form $F(x) = 1 - (\frac{a}{x})^b$ for $x \geq a$, zero otherwise. When $a = b = 1$ it is often called Standard Pareto distribution.

where F_X, F_Y are stationary distributions of X_t and Y_t, respectively. It mathematically represents how large Y will be in the next q steps if X is extremely large (in their respective scales). In our example, $q = 5$ and if X_0 is extremely large, then Y_5 will surely be also extremely large (large blue implies large red), but not the other way around. This implies that the following should hold: $\Gamma^{time}_{X,Y}(q) = 1$, but $\Gamma^{time}_{Y,X}(q) < 1$. The main part of the thesis consists of determining the assumptions under which this is really true.

A similar idea was used in Gnecco et al. [2020], which was not dealing with time series. This is a copula-based coefficient, and a similar notion also provides a conditional tail expectation coefficient from the economy.

0.0.5 Thesis organization

The thesis is organised as follows. Chapter 1 gives some preliminaries about time series, their specific models such as VAR and nonlinear models. It also covers an extremal value theory, and formal definition of causation. Chapter 2 contains the main results, together with a model example of the method. Chapter 3 gives some extensions of the proposed method, provides its properties and discusses what will happen under violating the assumptions. Moreover, it also discusses the time lag estimation problem mentioned above. Chapter 4 deals with the problem of estimation of the method. It discusses some properties of a proposed estimator and uses simulations on artificial data sets. In the end, the method is applied to a real data set concerning geomagnetic storms. We will confirm results presented by another article, which uses conditional mutual information to determine the cause of this phenomena. Finally, Chapter 5 and Chapter 6 consist of proofs; Chapter 5 deals with auxiliary propositions, and Chapter 6 with direct proofs of theorems from the previous chapters. The thesis finishes with a Bibliography.

1. Preliminaries

This chapter will give a short review of some results from multivariate nonlinear time series analysis, extremal value theory and causal inference.

1.1 Time series

This section will define stationary stochastic processes, univariate autoregressive $AR(q)$ processes and vector autoregressive $VAR(q)$ processes and nonlinear autoregressive processes.

If the reader is not familiar with this topic, we recommend the references Prášková [2017] for the univariate time series, and Lütkepohl [2007] for multivariate time series and VAR processes. For an introduction to nonlinear time series, we recommend Tsay and Chen [2018].

1.1.1 Some basics

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω and P is a probability measure. A (discrete) univariate stochastic process (or univariate time series) is a real-valued function $X : \mathbb{Z} \times \Omega \to \mathbb{R}$, where for each fixed $t \in \mathbb{Z}$ is $X(t, \omega)$ a random variable, i.e. measurable w.r.t. $\mathcal{F}$. The random variable corresponding to a fixed t is usually denoted by X_t, and we will use the notation $X = (X_t, t \in \mathbb{Z})$. The underlying probability space will usually not be mentioned, and it will be understood that all random variables are defined on the same probability space. A multivariate (or d-dimensional) stochastic process is a function $Z : \mathbb{Z} \times \Omega \to \mathbb{R}^d$, such that for each fixed $t \in \mathbb{Z}$ is $Z(t, \omega)$ a d-dimensional random vector. We will usually work with $Z = (X, Y)^\top$, bivariate time series, where X (resp. Y) represents the first (resp. the second) component of Z.

A stochastic process Z is strictly stationary if the joint distributions of n consecutive variables are time-invariant. We will not work with other stationarity types – by a stationary process, we will always mean strict stationarity.

Let $(X_t, t \in \mathbb{Z})$ be a univariate stochastic process. We say that the sum $\sum_{t=1}^{\infty} X_t$ is summable (in probability, resp. almost surely (a.s. for short)), if $S_n = \sum_{t=1}^{n} X_t$ converges to some random variable (in probability, resp. a.s.) as $n \to \infty$. It holds in general that if $(X_t, t \in \mathbb{Z})$ are independent, then the sum $\sum_{t=1}^{\infty} X_t$ is summable in probability if and only if it is summable almost surely (see e.g. Theorem 6.5.7 in Gut [2013]) .

1.1.2 Linear models of univariate time series

Let $(\varepsilon_i, i \in \mathbb{Z})$ be iid random variables, and $q \in \mathbb{N}$ be a constant.

Definition 1.1. *Let α_i be a sequence of real constants. The stochastic process $(X_t, t \in \mathbb{Z})$ defined by*

$$X_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i},$$

is called causal linear process, if such process exists.

Let α_i fulfill $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$. If $(X_t, t \in \mathbb{Z})$ are iid random variables with finite variance, then $\sum_{i=1}^{\infty} \alpha_i X_i$ is a.s. summable (see e.g. Theorem 6.5.1 in Gut [2013]).

Definition 1.2. *Let $\alpha_1, \ldots, \alpha_q$ be a sequence of real constants. The stochastic process $(X_t, t \in \mathbb{Z})$ defined by*

$$X_t = \sum_{i=1}^{q} \alpha_i X_{t-i} + \varepsilon_t,$$

is called autoregressive stochastic process of order q, notation $AR(q)$, whenever such process exists.

Theorem 1.1. *Let α_i be a sequence of real constants, such that all the roots of the polynomial $f(x) = 1 - \alpha_1 x - \cdots - \alpha_q x^q$ lie outside the unit circle in $\mathbb{C}$, then the autoregressive process $(X_t, t \in \mathbb{Z})$ is a causal linear process, i.e.,*

$$X_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i},$$

where β_i are defined as the (unique) elements of the power series $\frac{1}{f(x)} = \sum_{i=0}^{\infty} \beta_i x^i$.

Proof. See e.g. Theorem 35 in Prášková [2017]. □

1.1.3 VAR(q) models

Let $(\varepsilon_i, i \in \mathbb{Z})$ be iid d-dimensional random vectors.

Definition 1.3. *Let $A_1, \ldots, A_q$ be fixed real $d \times d$ matrices. A stochastic process $(Z_t, t \in \mathbb{Z})$, defined by*

$$Z_t = A_1 Z_{t-1} + \cdots + A_q Z_{t-q} + \varepsilon_t,$$

is called the vector autoregressive model of order q, notation VAR(q), whenever such process exists. We understand the sum of vectors component-wise.

Remark. We should specify that $A_q \neq \mathbf{0}$ so that the order of the VAR process is uniquely defined. We will not do it for the latter convenience, and we admit that the $VAR(q)$ process is also a $VAR(q+h)$ process. We will refer to the maximal q such that $A_q \neq \mathbf{0}$ as the *minimal order*.

Definition 1.4. *A stochastic process $(Z_t, t \in \mathbb{Z})$ which follows $VAR(q)$ is called stable, if*

$$det(I_d - A_1 z - \cdots - A_q z^q) \neq 0, \quad \forall |z| \leq 1,$$

where I_d denotes unit d-dimensional matrix.

Theorem 1.2 (Causal representation). *If a stochastic process $(Z_t, t \in \mathbb{Z})$ which follows $VAR(q)$ is stable, then there exist matrices $B_i \in \mathbb{R}^{d \times d}$ such that*

$$Z_t = \sum_{i=0}^{\infty} B_i \varepsilon_{t-i}.$$

Proof. See e.g. page 25 in Lütkepohl [2007]. □

Theorem 1.3. *If a $VAR(q)$ process is stable, then it is also stationary. The other direction is not valid in general.*

Proof. Page 25 in Lütkepohl [2007]. $\square$

The stability condition is in literature often referred to as stationarity condition.

1.1.4 Nonlinear time series

The most general form of nonlinearity in time series can be achieved by considering that $Z_t = f(Z_{t-1}, \ldots, Z_{t-q}, \varepsilon_t)$ holds for some "reasonable" function f. Nonlinear autoregressive model (NAR) is a special case when $Z_t = f(Z_{t-1}, \ldots, Z_{t-q}) + \varepsilon_t$. We will consider only the case when function f is additive.

Definition 1.5. *Let $(\varepsilon_i, i \in \mathbb{Z})$ be iid d-dimensional random vectors. Let $f_1, f_2 \ldots, f_q$ be measurable real functions from $\mathbb{R}^n$ to $\mathbb{R}^n$. A stochastic process $(Z_t, t \in \mathbb{Z})$, defined by*

$$Z_t = f_1(Z_{t-1}) + \cdots + f_q(Z_{t-q}) + \varepsilon_t,$$

is called the nonlinear additive autoregressive model (NAAR) of order q, whenever such process exists.

We will now state two theorems which discuss the existence and stationarity of the process $Z_t = f(Z_{t-1}) + \varepsilon_t$ for $d = 1$.

Theorem 1.4 (sufficient condition)**.** *Let $(\varepsilon_i, i \in \mathbb{Z})$ be iid random variables. Let f be continuous function satisfying $\lim_{|x|\to\infty} \frac{|f(x)|}{|x|} < 1$. Let $\mathbb{E}\,|\varepsilon_t| < \infty$. Then the process $Z_t = f(Z_{t-1}) + \varepsilon_t$ is ergodic (and stationary).*

Proof. See Corollary 2.2 in Bhattacharya and Lee [1995]. Other reference is Theorem 2.2 in Andel [1989], that states a slightly stronger condition. $\square$

Lemma 1.1 (Necessary condition)**.** *Let $(Z_t, t \in \mathbb{Z})$ be a stationary univariate process. Let it satisfy*

$$Z_t = f(Z_{t-1}) + \varepsilon_t, \forall t \in \mathbb{Z},$$

where ε_t are iid, unbounded, non-negative and f is measurable. Then f must satisfy $\lim_{x\to\infty} \frac{f(x)}{x} \leq 1$, if the limit exists.

Proof. If $\lim_{x\to\infty} \frac{f(x)}{x} > 1$, then there exists x_0 such that $\forall x \geq x_0 : \frac{f(x)}{x} > 1$. With probability one, there exists $t \in \mathbb{Z}$ such that $\varepsilon_t > x_0$. Then, $x_0 \overset{\text{a.s.}}{<} Z_t \overset{\text{a.s.}}{<} Z_{t+1} \overset{\text{a.s.}}{<} \ldots$ because ε_t is non-negative. This is a contradiction with stationarity. $\square$

We will work only with such NAAR processes, for which f_i are all additive in the sense that $f_i(x_1, \ldots, x_d) = (\sum_{j=1}^d f_i^{j,1}(x_j), \ldots, \sum_{j=1}^d f_i^{j,d}(x_j))^\top$ for some measurable real functions $f_i^{j,k}$. For $d = 2$ and $q = 1$, we obtain the model as in Definition 1.14.

1.2 Extremal value theory

If the reader is not familiar with the concept of heavy-tailed distributions, we recommend Foss et al. [2013], or shorter article Mikosch [1999]. There is no unique definition of a random variable with a heavy-tailed distribution. Generally, heavy-tailed distributions are those whose tails decay to zero slower than at an exponential rate (or slower than normal distribution). For the following chapters, we will need the so-called max-sum equivalence theorem (or more precisely, "principle of the single big jump"), which can be conveniently defined for distributions with regularly varying tails.

Definition 1.6. *A positive, measurable function f is called regularly varying with index $\theta \in \mathbb{R}$ if it is defined on some neighbourhood $[x_0, \infty)$ of infinity, and*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\theta, \quad \forall t > 0.$$

If $\theta = 0$, we call f slowly varying function. We call θ the (heavy-tailed) tail index.

Lemma 1.2. *For every regularly varying function, there exists a slowly varying function L such that $f(x) = x^\theta L(x)$.*

Proof. Remark 1.1.3. in Mikosch [1999]. □

Notation. *We will use notation $f(x) \sim g(x) \iff \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.*

Definition 1.7. *A random variable X with distribution function F_X is called heavy-tailed (or regularly varying) with tail index $\theta > 0$, if the function $1 - F_X(x)$ is regularly varying function with tail index $-\theta$. Notation $X \sim RV(\theta)$.*

Definition 1.8. *A random variable X satisfies the tail balance condition, if $\lim_{u\to\infty} \frac{P(X>u)}{P(|X|>u)} = p \in [0, 1]$. We say that random variables X, Y have compatible (upper) tails, if $P(X > u) \sim P(Y > u)$.*

Note. *In some literature, compatible tails condition is often given in a slightly more general setup, where they require $\exists p > 0$ such that $\lim_{u\to\infty} \frac{P(X>u)}{P(Y>u)} = p$. For simplicity of notation, we will use only the case where $p = 1$.*

Example 1.1. Typical examples of heavy-tailed distributions with positive tail index are Cauchy distribution, Pareto distributions or Burr distributions.

Lemma 1.3. *Suppose $X \sim RV(\theta)$. Then $\mathbb{E}\,|X|^\beta < \infty$ for $\beta < \theta$, and $\mathbb{E}\,|X|^\beta = \infty$ for $\beta > \theta$.*

Proof. Proposition 1.3.2. in Mikosch [1999]. □

Theorem 1.5 (Max-sum equivalence)**.** *Let $X, Y \sim RV(\theta)$ be independent (not necessary with compatible tails). Then $X + Y$ is heavy-tailed with the same tail index θ and*

$$P(X + Y > x) \sim P(X > x) + P(Y > x) \sim P(\max(X, Y) > x).$$

Remark. Theorem 1.5 holds even if Y is not $RV(\theta)$, but only $\max(X, Y) \sim RV(\theta)$. It can be easily generalized for more than 2 random variables. The proof can be found in Section 1.3.1 in Mikosch [1999].

The following result is also a particular case of so-called Breiman's lemma.

Consequence. *For $X \sim RV(\theta)$ and $\alpha > 0$ holds $P(\alpha X > u) \sim \alpha^\theta P(X > u)$ for $u \to \infty$.*

Proof. Corollary 1.3.8 in Mikosch [1999]. $\square$

1.2.1 Causal linear process with heavy-tailed noise

Literature can be also found in Resnick [1987], Chapter 4. The main result that we want to use is the following.

Definition 1.9. *Let $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$. Let α_i be a sequence of non-negative constants. We will speak about a sum-equivalence if the following relation holds $P(\sum_{i=0}^{\infty} \alpha_i \varepsilon_i > u) \sim [\sum_{i=0}^{\infty} \alpha_i^\theta] P(\varepsilon_1 > u)$ (provided that the first sum is a.s. summable and the second sum is finite).*

We now give two theorems which assures the sum-equivalence.

Theorem 1.6 (Sufficient condition)**.** *Let $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$. If $\theta > 1$ we assume that $\mathbb{E}(\varepsilon_i) = 0$. Let α_i be a sequence of non-negative constants, such that one of the following conditions holds:*

1. *For $\theta > 2$ is $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$.*
2. *For $\theta \leq 2$ there $\exists \delta > 0$ such that $\sum_{i=0}^{\infty} \alpha_i^{\theta-\delta} < \infty$.*

Then it holds that

1. *$\sum_{i=0}^{\infty} \alpha_i \varepsilon_i$ is summable a.s.*
2. *Process defined by $X_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}$ is stationary.*
3. *$P(\sum_{i=0}^{\infty} \alpha_i \varepsilon_i > u) \sim [\sum_{i=0}^{\infty} \alpha_i^\theta] P(\varepsilon_1 > u)$.*

Proof. Lemma A.3 in Mikosch and Samorodnitsky [2000]. $\square$

Theorem 1.7. *Let $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$, with $P(|\varepsilon_i| > u) = L(u)u^{-\theta}$ for some slowly varying L for which the tail balance condition hold. Let either $L(u_2) \leq cL(u_1)$ hold for all $u_2 > u_1 > u_0$ for some constants u_0, c, or $L(u_1 u_2) \leq cL(u_1)L(u_2)$ for all $u_2, u_1 > u_0$ for some constants $u_0, c > 0$. Let α_i be a sequence of non-negative constants, such that $\sum_{i=0}^{\infty} \alpha_i^\theta < \infty$. Then, it holds that*

$$P(\sum_{i=0}^{\infty} \alpha_i \varepsilon_i > u) \sim [\sum_{i=0}^{\infty} \alpha_i^\theta] P(\varepsilon_1 > u),$$

provided that $\sum_{i=0}^{\infty} \alpha_i \varepsilon_i$ is a.s. summable.

Proof. Lemma A.4 in Mikosch and Samorodnitsky [2000]. $\square$

1.2.2 Nonlinear time series with heavy-tailed noise

Most of the theory from nonlinear time series relies on the existence of variance of the noise variables. There are a few publications that deal with the heavy-tailed case. We will refer to Yang and Hongzhi [2005]. The following statements are consequence of Theorem 2.3. from this reference.

Theorem 1.8. *Let* $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$ *with positive density everywhere. Let* $(X_t, t \in \mathbb{Z})$ *follow NAR model, specified by* $X_t = f(X_{t-1}, \dots, X_{t-q}) + \varepsilon_t$ *for some measurable* f*.*

- *Let* $\mathbf{X_t} = (X_t, \dots, X_{t-q+1})^\top, u_t = (\varepsilon_t, 0, \dots, 0)^\top$ *and* $\Phi(\mathbf{X_t}) = (f(\mathbf{X_t}), X_t, \dots, X_{t-q+2})^\top$[1]*.*
- *Let there exist* $\rho \in [0, 1), K > 0$ *and some norm* $|| \cdot ||$ *on* $\mathbb{R}^q$*, such that for all* $x \in \mathbb{R}^q$ *holds*

$$||\Phi(x)|| \le \rho||x|| + K.$$

Then, NAR model $(X_t, t \in \mathbb{Z})$ *is geometrically ergodic (and stationary) and* $X_t \sim RV(\theta)$*.*

Consequence. *Let* $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$ *with positive density everywhere. Let* $(X_t, t \in \mathbb{Z})$ *follow NAAR model, specified by* $X_t = f_1(X_{t-1}) + \dots + f_q(X_{t-q}) + \varepsilon_t$*, where* f_i *are some measurable continuous functions satisfying* $\lim_{|x|\to\infty} \frac{|f_i(x)|}{|x|} < 1$*. Then,* $(X_t, t \in \mathbb{Z})$ *is ergodic (and stationary) process with* $X_t \sim RV(\theta)$*.*

Proof. This condition for the specific function $f(x_1, \dots, x_q) = f_1(x_1) + \dots + f_q(x_q)$ is clearly stronger than the condition from the previous theorem. □

Conjecture. *Let* $(X, Y)^\top$ *follow* $NAAR(q)$ *model, specified by*

$$\begin{aligned} X_t &= f_1(X_{t-1}) + f_2(Y_{t-q}) + \varepsilon_t^X, \\ Y_t &= g_1(Y_{t-1}) + g_2(X_{t-q}) + \varepsilon_t^Y. \end{aligned}$$

We require for functions f_1, f_2, g_1, g_2 *to be continuous and* $\lim_{|x|\to\infty} \frac{|h(x)|}{|x|} < 1$ *for all* $h = f_1, f_2, g_1, g_2$*. Moreover, let* $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$ *have positive density everywhere. Then,* $(X, Y)^\top$ *is stationary with* $X_t, Y_t \sim RV(\theta)$*.*

Proof and remark. The second part of the claim follows from Proposition 5.3. We do not prove the stationarity property; therefore, we will always assume it, even though we believe it holds automatically.

1.3 Causal inference

If the reader has not encountered a mathematical definition of causality, we highly recommend Peters et al. [2017]. If the reader wants to read a short introduction into Granger causality and causalities in time series, we recommend Palachy [2019].

[1]Then, the first equation can be equivalently rewritten as $\mathbf{X_t} = \Phi(\mathbf{X_{t-1}}) + u_t$.

1.3.1 Structural causal model

Let X, Y be real random variables.

Definition 1.10. *A bivariate structural causal model (SCM) with graph $X \to Y$ consists of two assignments,*

$$\begin{aligned} X &:= \varepsilon_X, \\ Y &:= f(X, \varepsilon_Y), \end{aligned}$$

where $\varepsilon_X, \varepsilon_Y$ are independent and f is some measurable function [2]*. In this case, we say that X causes Y, or that X is the cause and Y is the effect.*

Lemma 1.4. *For every joint distribution $P_{X,Y}$, there exists an SCM, where $Y = f(X, \varepsilon_Y)$ where f is some measurable function and X is independent of ε_Y.*

Proof. Proposition 4.1 in Palachy [2019]. □

This result can be applied to $X \to Y$ and also to $Y \to X$. Therefore, without any other assumptions, we cannot detect any information about the causal direction only from observational distribution (this is not true in the multivariate SCM, though).

Definition 1.11. *Linear models with non-Gaussian additive noise (LiNGAMs) are a specific cases of SCM, where $f(X, \varepsilon_Y) = \alpha X + \varepsilon_Y$, where $\alpha \in \mathbb{R}$ is a constant and noise variables are not normally distributed.*

Lemma 1.5. *If we assume LiNGAMs model, the causal direction is identifiable. That is, if $P_{X,Y}$ admits the linear model $Y = \alpha X + \varepsilon_Y$ where X, ε_Y are independent, then there exist $\beta \in \mathbb{R}$ and ε_X independent of Y such that $X = \beta Y + \varepsilon_X$ if and only if both X, ε_Y are Gaussian.*

Proof. Theorem 4.2 in Palachy [2019]. □

Definition 1.12. *Nonlinear additive noise models (ANMs) are specific cases of SCM, where $f(X, \varepsilon_Y) = f_Y(X) + \varepsilon_Y$.*

ANMs models with some (quite general) conditions on f and the distributions of noise are also identifiable. For example, if noises are Gaussian, then only linear functions generate non-identifiable models.

1.3.2 Causality in time series

Now we will deal with structural causal models in more general form, when we potentially have dependent stochastic processes. There is a large number of different notions of causality. Generally, the process X causes Y if the knowledge of X can improve the prediction of Y.

In VAR processes, this definition is equivalent to the following definition, which will be used in the thesis.

[2] To prevent some trivial cases, we should specify that f depends on its first argument on some subset of the support of X with non-zero measure.

Definition 1.13. *Let* $(X, Y)^\top = ((X_t, Y_t)^\top, t \in \mathbb{Z})$ *follow stable* $VAR(q)$ *model, specified by*

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_q X_{t-q} + \gamma_1 Y_{t-1} + \cdots + \gamma_q Y_{t-q} + \varepsilon_t^X,$$
$$Y_t = \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q} + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \varepsilon_t^Y.$$

Then, we say that X *(Granger) causes* Y *if there exists* $\delta_i \neq 0$.

In the notion of Theorem 1.2, we can rewrite

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}^X + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}^Y,$$
$$Y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}^Y + \sum_{i=0}^{\infty} d_i \varepsilon_{t-i}^X.$$

Then, X causes Y if and only if $d_i \neq 0$ for some i.

Definition 1.14. *Let* $(X, Y)^\top = ((X_t, Y_t)^\top, t \in \mathbb{Z})$ *follow additive* $NAAR$ *model, i.e. for each* $t \in \mathbb{Z}$ *holds*

$$X_t = f_1(X_{t-1}) + f_2(Y_{t-1}) + \varepsilon_t^X,$$
$$Y_t = g_1(Y_{t-1}) + g_2(X_{t-1}) + \varepsilon_t^Y.$$

Then, we say that X *(Granger) causes* Y *if* g_2 *is a non-constant function on the support of* X_{t-1} *(a.s.).*

1.3.3 Testing causalities

There are many different tests and statistics for the causality detection. Perhaps the most popular one is the so-called Granger test. For $VAR(q)$ model as in Definition 1.13, it uses linear regression to test the hypothesis if the sub-model $Y_t = \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q} + \varepsilon_t^Y$ of model $Y_t = \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q} + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \varepsilon_t^Y$ is sufficient (sum of squares is significantly smaller than in the latter model). This was actually the very first definition of causality according the famous article Granger [1969].

Another test can be done using transfer entropy. A basic approach from an information theory for causality detection is using entropy and conditional mutual information techniques. We recommend Paluš et al. [2007] for further information.

2. Causal tail coefficient for time series

Let $(X,Y)^\top = ((X_t,Y_t)^\top, t\in\mathbb{Z})$ be a bivariate time series. The main definition of this thesis is the causal tail coefficient for time series $\Gamma^{time}_{X,Y}$, which gives a numerical value of the causal influence from X to Y. Causal tail coefficient for a pair of random variables in SCM was first introduced in Gnecco et al. [2020]. We deal with time series as follows.

Definition 2.1. *Let $(X,Y)^\top = ((X_t,Y_t)^\top, t\in\mathbb{Z})$ be a bivariate (strictly) stationary time series. Causal tail coefficient for time series with lag q is defined as*

$$\Gamma^{time}_{X,Y}(q) := \lim_{u\to 1^-} \mathbb{E}\left[\max\{F_Y(Y_0),\dots,F_Y(Y_q)\} \mid F_X(X_0) > u\right], \tag{2.1}$$

where F_X, F_Y are the distribution functions of X_0, Y_0, respectively (if the limit exists). The coefficient $\Gamma^{time}_{X,Y}(q)$ without the zero term $F_Y(Y_0)$ will be denoted by

$$\Gamma^{time}_{X,Y}(q;-0) := \lim_{u\to 1^-} \mathbb{E}\left[\max\{F_Y(Y_1),\dots,F_Y(Y_q)\} \mid F_X(X_0) > u\right]. \tag{2.2}$$

Lemma 2.1 (Obvious observations)**.** *Always $\Gamma^{time}_{X,Y}(q)\in[0,1]$, and $\Gamma^{time}_{X,Y}(q;-0)\le \Gamma^{time}_{X,Y}(q)\le\Gamma^{time}_{X,Y}(q+1)$. Moreover, $\Gamma^{time}_{X,Y}(q)$ is invariant under linear transformations of our time series.*

Remark. The previous definition mathematically expresses very natural questions: How large Y will be if X is large? Does an extreme in X always cause an extreme in Y? If X_0 is extremely large, will there be an Y_i in the next q steps, which is also extremely large?

We will show that, under some assumptions, $\Gamma^{time}_{X,Y}(q) = 1$ if and only if X causes Y. Hence, if $\Gamma^{time}_{X,Y}(q) = 1$ and $\Gamma^{time}_{Y,X}(q) < 1$, we found an asymmetry between time series X, Y and we can detect a causal relationship.

First, we need to establish some assumptions for the time series.

2.1 Models

Definition 2.2 (Heavy-tailed VAR model)**.** *Let $(X,Y)^\top$ follow stable $VAR(q)$ model, specified by*

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \dots + \alpha_q X_{t-q} + \gamma_1 Y_{t-1} + \dots + \gamma_q Y_{t-q} + \varepsilon^X_t, \\ Y_t &= \beta_1 Y_{t-1} + \dots + \beta_q Y_{t-q} + \delta_1 X_{t-1} + \dots + \delta_q X_{t-q} + \varepsilon^Y_t. \end{aligned}$$

We will denote its causal representation by

$$\begin{aligned} X_t &= \sum_{i=0}^{\infty} a_i\varepsilon^X_{t-i} + \sum_{i=0}^{\infty} c_i\varepsilon^Y_{t-i}, \\ Y_t &= \sum_{i=0}^{\infty} b_i\varepsilon^Y_{t-i} + \sum_{i=0}^{\infty} d_i\varepsilon^X_{t-i}. \end{aligned}$$

Assumptions:

- $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$,
- $\alpha_i, \beta_i, \gamma_i, \delta_i \geq 0$ *(this assumption is not necessary and will be discussed in Chapter 3),*
- $\exists \delta > 0$ *such that* $\sum_{i=0}^{\infty} a_i^{\theta-\delta} < \infty, \sum_{i=0}^{\infty} b_i^{\theta-\delta} < \infty, \sum_{i=0}^{\infty} c_i^{\theta-\delta} < \infty,$ $\sum_{i=0}^{\infty} d_i^{\theta-\delta} < \infty$ *(or the assumptions from Theorem 1.7 hold).*

Then, we will say that $(X, Y)^\top$ *follows Heavy-tailed VAR model.*

Definition 2.3 (Heavy-tailed NAR model)**.** *Let* $(X, Y)^\top$ *follow stationary* $NAR(q)$ *model, specified by*

$$
\begin{aligned}
X_t &= f_1(X_{t-1}) + f_2(Y_{t-q}) + \varepsilon_t^X, \\
Y_t &= g_1(Y_{t-1}) + g_2(X_{t-q}) + \varepsilon_t^Y.
\end{aligned}
$$

We require for functions f_1, f_2, g_1, g_2 *to be either constant functions, or they are continuous non-negative and satisfy* $\lim_{x\to\infty} h(x) = \infty$ *and* $\lim_{x\to\infty} \frac{h(x)}{x} < 1$ *for* $h = f_1, f_2, g_1, g_2$.

Moreover, let $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$ *be non-negative. Then, we will say that* $(X, Y)^\top$ *follows Heavy-tailed NAR model.*

2.2 Causal direction

Theorem 2.1. *Let* $(X, Y)^\top$ *be a time series which follows either Heavy-tailed VAR or Heavy-tailed NAR model. If* X *causes* Y*, then* $\Gamma_{X,Y}^{time}(q) = 1$.

Remark. We assumed that we know the exact (correct) order q. But, for every $p \geq q$ we also have $\Gamma_{X,Y}^{time}(p) \geq \Gamma_{X,Y}^{time}(q) = 1$. The choice of appropriate q will be discussed in Chapter 3.

Remark. Note that we did not use the heavy-tailed condition in the proof.

Theorem 2.2. *Let* $(X, Y)^\top$ *be a time series which follows either Heavy-tailed VAR or Heavy-tailed NAR model. If* Y *is not causing* X *then* $\Gamma_{Y,X}^{time}(p) < 1$ *for all* $p \in \mathbb{N}$.

Remark. The primary step of the proof stems from Proposition 5.2. The idea is that large sums of independent, regularly varying random variables tend to be driven by only a single large value. So if Y_0 is large, it can be because some ε_i^Y is large, which does not affect X_k.

Example 2.1. Let $(X, Y)^\top$ follow the following stable VAR model:

$$
\begin{aligned}
X_t &= 0.5X_{t-1} + \varepsilon_t^X, \\
Y_t &= 0.5Y_{t-1} + 0.5X_{t-1} + \varepsilon_t^Y,
\end{aligned}
$$

where $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Pareto(1, 1)$, i.e. with tail index $\theta = 1$.

Its causal representation is

$$
\begin{aligned}
X_t &= \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{t-i}^X, \\
Y_t &= \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{t-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{t-i}^X.
\end{aligned}
$$

In this case is lag $q = 1$, and it is sufficient to take only

$$
\Gamma_{X,Y}^{time}(1; -0) = \lim_{u \to 1^-} \mathbb{E}\left[F_Y(Y_1) \mid F_X(X_0) > u\right]
$$

(see also Section 3.3 for discussion). Let us give some vague computation of this coefficient. From Theorem 2.1 is $\Gamma_{X,Y}^{time}(1) = 1$. For the other direction, rewrite

$$
\lim_{u \to 1^-} \mathbb{E}\left[F_X(X_1) \mid F_Y(Y_0) > u\right] = \lim_{u \to \infty} \mathbb{E}\left[F_X(X_1) \mid \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^X > u\right].
$$

First, note the following (first follows from the independence, second from Lemma 5.2):

$$
\begin{aligned}
&\lim_{u \to \infty} \mathbb{E}\left[F_X(X_1) \mid \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Y > u\right] = \mathbb{E}\left[F_X(X_1)\right] = 1/2, \\
&\lim_{u \to \infty} \mathbb{E}\left[F_X(X_1) \mid \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^X > u\right] = 1.
\end{aligned}
$$

Next, we know that $P(X_1 < K \mid \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^X > u) = \frac{P(X_1 < K)}{2}$ for every $K \in \mathbb{R}$, which holds due to Proposition 5.2 [1]. Simply put, with probability 1/2 has $X_1 | \{F_Y(Y_0) > u\}$ the same distribution as non-conditional X_1, and with complementary probability it diverges to ∞ (as $u \to \infty$) . Together is $\lim_{u \to 1^-} \mathbb{E}\left[F_X(X_1) \mid F_Y(Y_0) > u\right] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$.

What if we do not know the exact lag q? If we put for example $q = 2$, we obtain that $\Gamma_{Y,X}^{time}(2) = \lim_{u \to 1^-} \mathbb{E}\left[\max\{F_X(X_0), F_X(X_1), F_X(X_2)\} \mid F_Y(Y_0) > u\right]$ will be slightly larger than $\frac{3}{4}$. More precisely, it will be equal to $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \mathbb{E}\left[\max\{F_X(X_0), F_X(X_1), F_X(X_2)\}\right]$. Using computer software and many simulations, the true value is somewhere near 0.80.

Remark. Note that we keep the "zero term" $F_Y(Y_0)$ in our definition. Our models do not allow instantaneous effects, i.e. cases where X_0 causes Y_0. In real data sets, such situations can happen if the data have considerable time differences between their measurements. Therefore, it is convenient to leave them in our definition.

[1] Note the identities $\sum_{i=0}^{\infty} \frac{i}{2^i} = 2 = \sum_{i=0}^{\infty} \frac{1}{2^i}$.

3. Properties and extensions

3.1 Modifications

Up to now, we assumed that large X causes large Y. In other words, we assumed that all coefficients in our Heavy-tailed VAR model are non-negative. In this section, we will discuss the extension to possibly negative coefficients and non-direct proportional dependencies.

3.1.1 Modification for non-direct proportion

The most straightforward modification can be used when large X causes small Y. For example, this is the usual case in some markets, where one shop's profit causes others' loss. In this case, it is enough to use $-X$ instead of X, as in the following example.

Example 3.1. Let $(X, Y)^\top$ follow the following VAR model:

$$\begin{aligned} X_t &= 0.7X_{t-1} + \varepsilon_t^X, \\ Y_t &= 0.6Y_{t-1} - 0.5X_{t-1} + \varepsilon_t^Y, \end{aligned}$$

where $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} \textit{Cauchy}$ [1]. Then, by transforming $X_t^* = -X_t$, we obtain

$$\begin{aligned} X_t^* &= 0.7X_{t-1}^* - \varepsilon_t^X, \\ Y_t &= 0.6Y_{t-1} + 0.5X_{t-1}^* + \varepsilon_t^Y, \end{aligned}$$

what fits into our previous theory. The same holds if we choose to transform Y instead of X.

This modification can not be used in the general case.

3.1.2 Modification with absolute value

Consider that our time series are centred around zero (if $\mathbb{E}(X_1), \mathbb{E}(Y_1)$ exist, they are zero) and have full support on $\mathbb{R}$. A simple transformation of our series can achieve these conditions. The idea to extend the causal tail coefficient for time series is to use the absolute values of $|X|, |Y|$ instead of X, Y. However, the general VAR series can have very complicated relations.

Example 3.2. Let $(X, Y)^\top$ follow the following VAR model:

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_t^X, \\ Y_t &= X_{t-1} - 0.5X_{t-2} + \varepsilon_t^Y. \end{aligned}$$

Then, its causal representation is

$$\begin{aligned} X_t &= \sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{t-i}^X, \\ Y_t &= \varepsilon_t^Y + \varepsilon_{t-1}^X. \end{aligned}$$

[1] Cauchy distribution is a distribution with density function $f(x) = \frac{1}{\pi}\frac{1}{x^2+1}$, which is regularly varying with the tail index $\theta = 1$.

Detecting some extremal causal relations can be very difficult, because even though X causes Y, extreme of X_{t-1} does not imply that Y_t will be also extreme (if X_{t-2} was large, then X_{t-1} will also be large, but Y_t not). Therefore, we will restrict our time series in such a way that this implication holds.

Definition 3.1 (Extremal causal condition)**.** *Let $(X,Y)^\top$ be a time series such that X causes Y. Let $(X,Y)^\top$ follow a stable $VAR(q)$ model, with its causal representation in the form*

$$
\begin{aligned}
X_t &= \sum_{i=0}^{\infty} a_i \varepsilon^X_{t-i} + \sum_{i=0}^{\infty} c_i \varepsilon^Y_{t-i},\\
Y_t &= \sum_{i=0}^{\infty} b_i \varepsilon^Y_{t-i} + \sum_{i=0}^{\infty} d_i \varepsilon^X_{t-i}.
\end{aligned}
$$

We say that it satisfies an extremal causal condition, if there exists $p \leq q$ such that the following implication holds:

$$\forall i \in \mathbb{N} \cup \{0\} : a_i \neq 0 \implies d_{i+p} \neq 0.$$

Lemma 3.1. *The extremal causal condition holds in Heavy-tailed VAR model (i.e. where the coefficients are non-negative), where X causes Y.*

Proof. In the notion of the definition of Heavy-tailed VAR model and Theorem 2.2, if $\delta_p > 0$, then

$$\sum_{i=0}^{\infty} d_i \varepsilon^X_{p-i} + \sum_{i=0}^{\infty} b_i \varepsilon^Y_{p-i} = Y_p = \delta_p X_0 + \cdots = \delta_p (\sum_{i=0}^{\infty} a_i \varepsilon^X_{-i} + \sum_{i=0}^{\infty} c_i \varepsilon^Y_{-i}) + \ldots .$$

Therefore, if $a_i > 0$, then $d_{i+p} \geq \delta_p a_i > 0$. □

Remark. The extremal causal condition implies that for every $k \geq p$, $|Y_k|$ will be arbitrarily large, provided that $|X_0|$ is large enough.

Theorem 3.1. *Let $(X,Y)^\top$ be a time series which follows Heavy-tailed VAR model, with possibly negative coefficients, satisfying the extremal causal condition. Moreover, let $\varepsilon^X_t, \varepsilon^Y_t$ have full support on $\mathbb{R}$, are iid satisfying tail balance condition. If X causes Y, but Y does not cause X, then $\Gamma^{time}_{|X|,|Y|}(q) = 1$, and $\Gamma^{time}_{|Y|,|X|}(q) < 1$.*

Remark. To detect the causal direction in time series such as in the previous Example 3.2, we could change our coefficient (2.1) in the following way

$$\Gamma^{time}_{X,Y}(q) := \lim_{u \to 1^-} \mathbb{E}\left[\max\{F_Y(Y_0), \ldots, F_Y(Y_q)\} \mid F_X(X_0) > u, F_X(X_{-1}) < u\right].$$

This would assure that the "extremal shock" happened at $t = 0$, not earlier. Such a coefficient does not have good properties and is still not applicable for a general class of VAR coefficients.

3.1.3 Derivatives

Sometimes, it is more convenient to interpret a causal influence on a local scale, where the "jump" will be the largest and not the global value will be the largest. Let us define the time series of differences obtained from $(X, Y)^\top$ by

$$\begin{aligned} X_t^* &:= X_t - X_{t-1}, \\ Y_t^* &:= Y_t - Y_{t-1}. \end{aligned}$$

It is easy to see that if $(X, Y)^\top$ is a stable VAR process, then also $(X^*, Y^*)^\top$ is a stable VAR process.

Example 3.3. Let $(X, Y)^\top$ follow the following $VAR(1)$ model:

$$\begin{aligned} X_t &= 0.999 X_{t-1} + \varepsilon_t^X, \\ Y_t &= 0.5 X_{t-1} + \varepsilon_t^Y, \end{aligned}$$

where $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Cauchy$. It is not hard to compute that $\Gamma_{Y,X}^{time}(1) \approx 1$. But on the other hand, $\Gamma_{|Y^*|,|X^*|}^{time}(1) \approx \frac{3}{4}$.

As we can see from Example 3.3, using $(X^*, Y^*)^\top$ instead of $(X, Y)^\top$ can sometimes lower the causal tail coefficient for time series, and it can be easier to distinguish it from 1 (i.e. we need fewer data to obtain the same p-value). Of course, it is not always true. The rule of thumb is to use this approach if we have one process with long-range dependence and the other with short-range dependence.

3.2 Common cause

Reichenbach's common cause principle states that for every two random variables X, Y holds exactly one of the following: They are independent, X causes Y, Y causes X, or there exists Z causing both X and Y. The problem is to distinguish between true causality and dependence due to a common cause.

Theorem 3.2. *Let $(X, Y, Z)^\top$ follow a 3 dimensional stable $VAR(q)$ model, with iid regularly varying noise variables. Let Z be a common cause of both X and Y. If X does not cause Y, then $\Gamma_{X,Y}^{time}(q) < 1$.*

Therefore, we can distinguish between true causality and common cause. We do not observe all relevant data in practice, but the Theorem 3.2 holds even if we do not observe the common cause. However, the common cause still needs to fulfill the condition that noise is regularly varying with not greater tail index than tail indexes of X and Y. We can not check this assumption if we do not observe all the relevant data.

Example 3.4. Let $(X, Y, Z)^\top$ follow a 3 dimensional stable $VAR(q)$ model, specified by

$$\begin{aligned} Z_t &= 0.5 Z_{t-1} + \varepsilon_t^Z, \\ X_t &= 0.5 X_{t-1} + 0.5 Z_{t-1} + \varepsilon_t^X, \\ Y_t &= 0.5 Y_{t-1} + 0.5 Z_{t-1} + \varepsilon_t^Y, \end{aligned}$$

where $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Pareto(2,2)$ and $\varepsilon_t^Z \overset{\text{iid}}{\sim} Pareto(1,1)$ (i.e. ε_t^Z has heavier tail than ε_t^X). Then, $\Gamma_{X,Y}^{time}(1) = \Gamma_{Y,X}^{time}(1) = 1$ even though X does not cause Y[2].

3.3 Estimating the lag q

3.3.1 Choosing lag q for the causal tail coefficient for time series

We assumed that we know the exact order q of our time series in all previous sections. What should we do if we do not know it? If we choose q too small, we do not obtain correct causal relations. On the other hand, we can choose q very large, and all the theoretical results will be still valid. The only problem is that if we choose large q, then $\Gamma_{Y,X}^{time}(q)$ will be close to 1, which makes it harder to statistically distinguish from 1 (i.e. we need more data to obtain the same p-value). In the following, we propose a possible choice for q. Section 4.5 also discusses this problem.

3.3.2 Extremogram

In classical time series, the well-accepted object for describing a serial dependence between different time series is an auto-correlation function. In general, auto-correlation function (or cross-correlation functions) do not behave properly under heavy-tailed marginals. This problem partially solves an extremogram. Reference is Davis and Mikosch [2009]. Some vague definition is that an extremogram is

$$\gamma_{A,B}(q) = \lim_{n\to\infty} n \cdot cov(I_{\{a_n^{-1}X_0 \in A\}}, I_{\{a_n^{-1}X_q \in B\}}),$$

for an appropriate scaling sequence $a_n \to \infty$, and A, B are Borel sets bounded away from 0. Therefore, it can be seen as a limiting correlogram. If we estimate the values for a wide range of $q \in \mathbb{Z}$ (and appropriate sets A, B), we obtain a tool similar to the auto-correlation function, but adjusted for extremes. Choosing the largest value out of all possible q, we have *some* choice for the lag. More importantly, it can be a powerful graphical tool, where we can somehow visualize the extremal dependence.

3.4 Time series synchronization

Consider the following problem: For time series $(X, Y)^\top$, where X causes Y, we want to estimate how long it takes for information from X to affect Y. If we do an intervention on X, *when* will it affect Y? A typical example from the economy can be the following. Let us have two time series representing prices of milk and prices of cheese in time. One day, the government raises taxes for the prices of milk by 10%. When can we anticipate an increase in the prices of cheese?

[2] We do not provide a rigorous proof of this equality, but such a proof follows the similar steps as the proof of Theorem 2.1, using the fact that $P(\varepsilon_t^Z + \varepsilon_t^X > u) \sim P(\varepsilon_t^Z > u)$ and causal representation $X_0 = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{-i}^X + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{-i}^Z$ and $Y_1 = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{1-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{1-i}^Z$.

Definition 3.2 (Minimal lag)**.** *Let $(X,Y)^\top$ follow a stable $VAR(q)$ model, specified by*

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \cdots + \alpha_q X_{t-q} + \gamma_1 Y_{t-1} + \cdots + \gamma_q Y_{t-q} + \varepsilon_t^X, \\ Y_t &= \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q} + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \varepsilon_t^Y. \end{aligned}$$

We call $p \in \mathbb{N}$ the minimal lag, if $\gamma_1 = \cdots = \gamma_{p-1} = \delta_1 = \cdots = \delta_{p-1} = 0$ and either $\delta_p \neq 0$ or $\gamma_p \neq 0$. If such p does not exist, we define the minimal lag as $+\infty$.

We propose a simple method for estimating the minimal lag.

Lemma 3.2. *Let $(X,Y)^\top$ follow Heavy-tailed VAR model, where X causes Y. Let p be the minimal lag. Then, $\Gamma^{time}_{X,Y}(r) < 1$ for all $r < p$, and $\Gamma^{time}_{X,Y}(r) = 1$ for all $r \geq p$.*

Therefore, a statement that the minimal lag is equal to p is equivalent to a statement that $\Gamma^{time}_{X,Y}(p)$ is the first coefficient which is equal to 1!

3.5 A note on other approaches

Using a causal tail coefficient for time series is undoubtedly not the only approach for detecting a causal direction in a heavy-tailed time series. Many similar concepts are present in the literature. In economy, the so-called *conditional tail expectation* is often used, which is only a slight modification of our causal tail coefficient (see Necir et al. [2010]). The conditional tail expectation is a function of an extremogram (or, more precisely, a quotient of two exponent measures). It may be interesting for a future research to show some connections between an extremogram and the causal tail coefficient.

Another interesting approach for causal detection can be using an extremal index (see e.g. Moloney et al. [2019]). Extremal index is a constant $\theta \in (0,1]$ associated with most stationary time series. For univariate series, it can be interpreted as an inverse of an average cluster size of extremes. For instance, if extremes appear in the size of 2, then $\theta = \frac{1}{2}$. For bivariate time series, there can be an asymmetry between cluster size of cause and effect extremes – if the cause is extreme, the cluster will be larger because it implies that the effect will also be extreme. It may be interesting for future research to examine such asymmetries using spatial statistics.

4. Estimations and simulations

All methods proposed in this chapter are programmed in R language (R Core Team [2020]). They can be found in the supplementary package [1].

4.1 Non-parametric estimator

In this section we discuss a possible estimator of a causal tail coefficient for a time series with lag $q \in \mathbb{N}$,

$$\Gamma^{time}_{X,Y}(q) = \lim_{u \to 1^-} \mathbb{E}\,[\max\{F_Y(Y_0), \dots, F_Y(Y_q)\} \mid F_X(X_0) > u],$$

based on a finite sample $(X_1, Y_1)^\top, \cdots, (X_n, Y_n)^\top$.

We propose a very natural estimator, which computes the estimate using only those values of $Y_i, \dots, Y_{i+q}$ where X_i is larger than some threshold.

Definition 4.1. *We define*

$$\hat{\Gamma}^{time}_{X,Y}(q) := \frac{1}{k} \sum_{i: X_i \geq \tau_k^X} \max\{\hat{F}_Y(Y_i), \dots, \hat{F}_Y(Y_{i+q})\},$$

where $\tau_k^X = X_{(n-k+1)}$ *is the k-th largest value of* $X_1, \dots, X_n$*, and* $\hat{F}_Y(Y_i) = \frac{1}{n}\sum_{j=1}^n 1\{Y_j \leq Y_i\}$.

Remark. This coefficient can possibly depend on random variables $Y_{n+1}, \dots, Y_{n+q}$. If we want to be fully rigorous, we should assume that we observe $n + q$ data, or define that $Y_{n+i} = Y_n$, which is a negligible modification for large n.

The number k represents number of extremes which we will take into account. In the following, k will depend on n, so to be more precise, we will write k_n instead of k. The basic condition in extremal value theory is that

$$k_n \to \infty, \frac{k_n}{n} \to 0, \text{ as } n \to \infty. \tag{4.1}$$

In the following methods we take $k_n = \sqrt{n}$. This choice is briefly discussed in Section 4.4.

The next theorem shows that such a statistic is "reasonable", by showing that it is (under much more general setting than just assuming VAR model) asymptotically unbiased.

Theorem 4.1. *Let* $(X, Y)^\top = ((X_t, Y_t)^\top, t \in \mathbb{Z})$ *be a stationary bivariate time series, whose marginal distributions are absolutely continuous with support on some neighbourhood of infinity. Let* $\Gamma^{time}_{X,Y}(q)$ *exists. Let* k_n *satisfy (4.1) and*

$$\frac{n}{k_n} P\Big(\frac{n}{k_n} \sup_{x \in \mathbb{R}} |\hat{F}_X(x) - F(x)| > \delta\Big) \overset{n \to \infty}{\to} 0, \ \ \forall \delta > 0. \tag{4.2}$$

Then, $\mathbb{E}\,\hat{\Gamma}^{time}_{X,Y}(q) \overset{n \to \infty}{\to} \Gamma^{time}_{X,Y}(q)$ [2].

[1] Or on the webpage https://github.com/jurobodik/Master_thesis.

[2] $\hat{\Gamma}^{time}_{X,Y}(q)$ depends on n, although it does not have such an index.

Remark. The condition (4.2) is satisfied for iid random variables when $\frac{k_n^2}{n} \to \infty$ (this follows from Dvoretzky–Kiefer–Wolfowitz inequality). For linear time series, the condition differs for different linear coefficients.

4.2 Some insight using simulations

We will simulate how the estimates $\hat{\Gamma}^{time}_{X,Y}$ work for a series of models. First, we (using Monte Carlo principle) estimate the distribution of $\hat{\Gamma}^{time}_{X,Y}$ and $\hat{\Gamma}^{time}_{Y,X}$ for the following model.

Definition 4.2. *Let* $(X, Y)^\top$ *follow*

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_t^X, \\ Y_t &= 0.5Y_{t-1} + \delta X_{t-2} + \varepsilon_t^Y, \end{aligned}$$

where $\varepsilon_t^X, \varepsilon_t^Y$ *are iid and* $(X, Y)^\top$ *is stable.*

Consider $\delta = 0.5$ and $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Cauchy$. We simulate such a time series with length $n = 10000$, and compute both $\hat{\Gamma}^{time}_{X,Y}$ and $\hat{\Gamma}^{time}_{Y,X}$. We repeat this 1000 times and the resulting numbers are drawn in Figure 4.1.

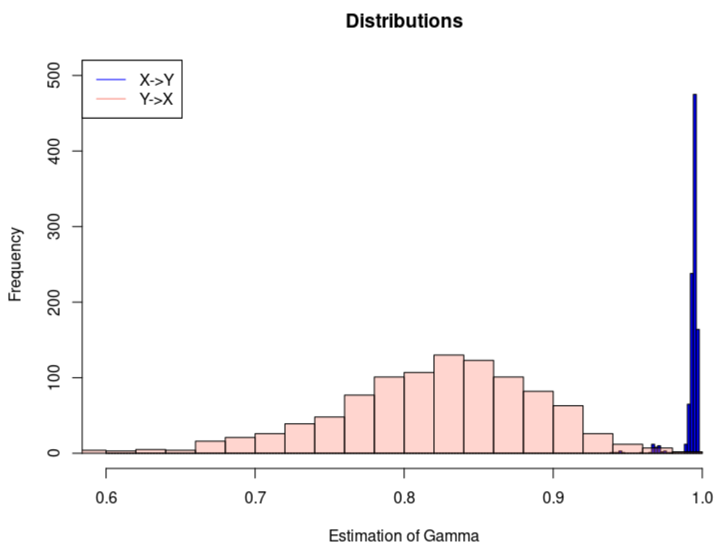

Figure 4.1: The histogram represent approximate distributions of $\hat{\Gamma}^{time}_{X,Y}$ (blue) and $\hat{\Gamma}^{time}_{Y,X}$ (red) from a model with a correct causal direction $X \to Y$ and a number of data $n = 10000$.

In the following, we will perform simulations for $(X, Y)^\top$ following the model given by Definition 4.2. We consider the following three choices for the parameter δ and three choices for the sample size n: $\delta = 0.1$, 0.5 and 0.9, $n = 100$, 1000 and 10000. The random variables ε_t^X and ε_t^Y are generated either from the standard normal or the standard Pareto distribution. For each δ, each noise distribution and each n, we compute the estimators $\hat{\Gamma}^{time}_{X,Y} := \hat{\Gamma}^{time}_{X,Y}(2)$. The procedure is repeated 200 times, and the means and quantiles of the estimators are computed.

Table 4.1 shows the results; each cell corresponds to the model with given δ, noise distribution and number n of data-points.

Example 4.1. The notation $\hat{\Gamma}^{time}_{X,Y} = 0.5 \pm 0.1$ means that out of all 200 simulated series from Definition 4.2, $\hat{\Gamma}^{time}_{X,Y}$ was on average equal to 0.5 and exactly 190 of those simulations were $\hat{\Gamma}^{time}_{X,Y} \leq 0.6$. We write $\pm$, because the 5% quantiles were in all cases symmetrical, i.e. cca 190 of those simulations fulfilled also $\hat{\Gamma}^{time}_{X,Y} \geq 0.4$.

Errors with standard Pareto distributions

	$n = 100$	$n = 1000$	$n = 10000$
$\delta = 0.1$	$\hat{\Gamma}^{time}_{X,Y} = 0.83 \pm 0.14$ $\hat{\Gamma}^{time}_{Y,X} = 0.66 \pm 0.23$	$\hat{\Gamma}^{time}_{X,Y} = 0.94 \pm 0.04$ $\hat{\Gamma}^{time}_{Y,X} = 0.66 \pm 0.16$	$\hat{\Gamma}^{time}_{X,Y} = 0.98 \pm 0.01$ $\hat{\Gamma}^{time}_{Y,X} = 0.65 \pm 0.12$
$\delta = 0.5$	$\hat{\Gamma}^{time}_{X,Y} = 0.91 \pm 0.07$ $\hat{\Gamma}^{time}_{Y,X} = 0.71 \pm 0.18$	$\hat{\Gamma}^{time}_{X,Y} = 0.98 \pm 0.01$ $\hat{\Gamma}^{time}_{Y,X} = 0.75 \pm 0.19$	$\hat{\Gamma}^{time}_{X,Y} = 0.994 \pm 0.002$ $\hat{\Gamma}^{time}_{Y,X} = 0.79 \pm 0.11$
$\delta = 0.9$	$\hat{\Gamma}^{time}_{X,Y} = 0.93 \pm 0.05$ $\hat{\Gamma}^{time}_{Y,X} = 0.75 \pm 0.17$	$\hat{\Gamma}^{time}_{X,Y} = 0.98 \pm 0.01$ $\hat{\Gamma}^{time}_{Y,X} = 0.8 \pm 0.15$	$\hat{\Gamma}^{time}_{X,Y} = 0.996 \pm 0.001$ $\hat{\Gamma}^{time}_{Y,X} = 0.84 \pm 0.1$

Errors with standard Gaussian distributions

	$n = 100$	$n = 1000$	$n = 10000$
$\delta = 0.1$	$\hat{\Gamma}^{time}_{X,Y} = 0.68 \pm 0.14$ $\hat{\Gamma}^{time}_{Y,X} = 0.63 \pm 0.19$	$\hat{\Gamma}^{time}_{X,Y} = 0.68 \pm 0.1$ $\hat{\Gamma}^{time}_{Y,X} = 0.63 \pm 0.13$	$\hat{\Gamma}^{time}_{X,Y} = 0.69 \pm 0.07$ $\hat{\Gamma}^{time}_{Y,X} = 0.62 \pm 0.08$
$\delta = 0.5$	$\hat{\Gamma}^{time}_{X,Y} = 0.83 \pm 0.11$ $\hat{\Gamma}^{time}_{Y,X} = 0.64 \pm 0.2$	$\hat{\Gamma}^{time}_{X,Y} = 0.86 \pm 0.06$ $\hat{\Gamma}^{time}_{Y,X} = 0.65 \pm 0.13$	$\hat{\Gamma}^{time}_{X,Y} = 0.90 \pm 0.03$ $\hat{\Gamma}^{time}_{Y,X} = 0.66 \pm 0.06$
$\delta = 0.9$	$\hat{\Gamma}^{time}_{X,Y} = 0.88 \pm 0.07$ $\hat{\Gamma}^{time}_{Y,X} = 0.64 \pm 0.19$	$\hat{\Gamma}^{time}_{X,Y} = 0.93 \pm 0.03$ $\hat{\Gamma}^{time}_{Y,X} = 0.65 \pm 0.13$	$\hat{\Gamma}^{time}_{X,Y} = 0.96 \pm 0.01$ $\hat{\Gamma}^{time}_{Y,X} = 0.66 \pm 0.09$

X with Pareto error, Y with Gaussian error

	$n = 100$	$n = 1000$	$n = 10000$
$\delta = 0.5$	$\hat{\Gamma}^{time}_{X,Y} = 0.96 \pm 0.02$ $\hat{\Gamma}^{time}_{Y,X} = 0.80 \pm 0.1$	$\hat{\Gamma}^{time}_{X,Y} = 0.98 \pm 0.0013$ $\hat{\Gamma}^{time}_{Y,X} = 0.92 \pm 0.04$	$\hat{\Gamma}^{time}_{X,Y} = 0.997 \pm 0.001$ $\hat{\Gamma}^{time}_{Y,X} = 0.98 \pm 0.011$

X with Gaussian error, Y with Pareto error

	$n = 100$	$n = 1000$	$n = 10000$
$\delta = 0.5$	$\hat{\Gamma}^{time}_{X,Y} = 0.65 \pm 0.15$ $\hat{\Gamma}^{time}_{Y,X} = 0.62 \pm 0.20$	$\hat{\Gamma}^{time}_{X,Y} = 0.67 \pm 0.1$ $\hat{\Gamma}^{time}_{Y,X} = 0.63 \pm 0.13$	$\hat{\Gamma}^{time}_{X,Y} = 0.68 \pm 0.05$ $\hat{\Gamma}^{time}_{Y,X} = 0.63 \pm 0.08$

Table 4.1: Consider 200 simulated time series following a model from Definition 4.2. Each cell represents a different coefficient δ, a different number of data-points n and a different noise distribution. Each value $\hat{\Gamma}^{time}_{X,Y} = \cdot \pm \cdot$ represents a mean of all 200 estimated coefficients $\hat{\Gamma}^{time}_{X,Y}$, and a difference between this mean and 95% quantile out of all 200 simulations. In every case, X causes Y.

Conclusion. *The method works surprisingly well under violating the heavy-tails assumption. If we consider Gaussian noise, the correct theoretical value is* $\Gamma^{time}_{X,Y} = \Gamma^{time}_{Y,X} = 1$. *But if we estimate these values, for finite* $u < 1$ *is the value of*

$\mathbb{E}\left[\max\{F_Y(Y_0),\ldots,F_Y(Y_q)\} \mid F_X(X_0) > u\right]$ *still larger than in the other direction. This results in seemingly correct causal directions.*

On the other hand, if the cause has heavier tails than the effect, our method suggests $\Gamma^{time}_{X,Y} = \Gamma^{time}_{Y,X} = 1$. *In this case, for large* n*, both estimates are very close to 1, and this results in the wrong causal directions. Therefore, the main problems are caused by a different tail behaviour.*

4.3 Testing

We want to develop a method that tells us the causal direction between two time series. One (quite trivial) option is to put a threshold, e.g. we say that X causes Y if and only if $\hat{\Gamma}^{time}_{X,Y}(q) \geq \tau$ where $\tau = 0.9$ or 0.95. The choice of τ should depend on number of data n – we can not expect for $n = 100$ to $\hat{\Gamma}^{time}_{X,Y}(q)$ be large. On the other hand, choosing small τ can lead to wrong conclusions.

Ideally, we would want to test the hypothesis $\Gamma^{time}_{X,Y}(q) = 1$ against the alternative $\Gamma^{time}_{X,Y}(q) < 1$. To do that, we would need to know (at least asymptotically) the distribution of $\hat{\Gamma}^{time}_{X,Y}(q)$. This is beyond the scope of this work (although the simulations suggest that it may follow some normal distribution).

Another method to estimate the confidence intervals is to use the block (sometimes called stationary) bootstrap technique[3].

Definition 4.3. *Let* $(X, Y, Z)^\top$ *follow*

$$\begin{aligned}
Z_t &= 0.5Z_{t-1} + \varepsilon^Z_t,\\
X_t &= 0.5X_{t-1} + 0.5Z_{t-2} + \varepsilon^X_t,\\
Y_t &= 0.5Y_{t-1} + 0.5Z_{t-1} + (X_{t-3})^{\frac{3}{4}} + 5\varepsilon^Y_t,
\end{aligned}$$

where $\varepsilon^X_t, \varepsilon^Y_t, \varepsilon^Z_t \overset{iid}{\sim} Pareto(1, 1)$. *In this case, process* Z *represents (not observed) common cause. A sample realization can be found in Figure 4.2.*

We understand that Z from Definition 4.3 is an unobserved common cause. Granger testing fails for such a case because Z creates spurious $Y \to X$ direction, although the truth is the opposite (also, Granger testing can not deal with nonlinearities and has problems with heavy-tailness). However, we obtain correct causal directions using $\hat{\Gamma}^{time}_{X,Y}(q)$. E.g. for $n = 1000$ we obtain very similar values of $\hat{\Gamma}^{time}_{X,Y}(q)$ and $\hat{\Gamma}^{time}_{Y,X}(q)$ such as in Figure 4.1. We will try to use the bootstrap method in the rest of the section to make this more formal.

In the rest of the section, we will perform a small simulations. Consider two models of time series. The first is a simple model given by Definition 4.2 (with $\delta = 0.5$ and $Pareto(1, 1)$ noise) as in the beginning of Section 4.2. The second model follows Definition 4.3. We will discuss three different methods for detecting the causal directions. First is the bootstrap method mentioned above. The second method is the classical Granger test[4]. We consider the significance level $\alpha = 0.05$.

[3]More precisely so-called Reverse Bootstrap Percentile Interval (see e.g. Section 5.4 in Hesterberg [2014]). Although in each resample is always $\hat{\Gamma}^{time}_{X,Y}(q) < 1$, this bootstrap modification can overcome this problem.

[4]using "grangertest" function from "lmtest" package Zeileis and Hothorn [2002].

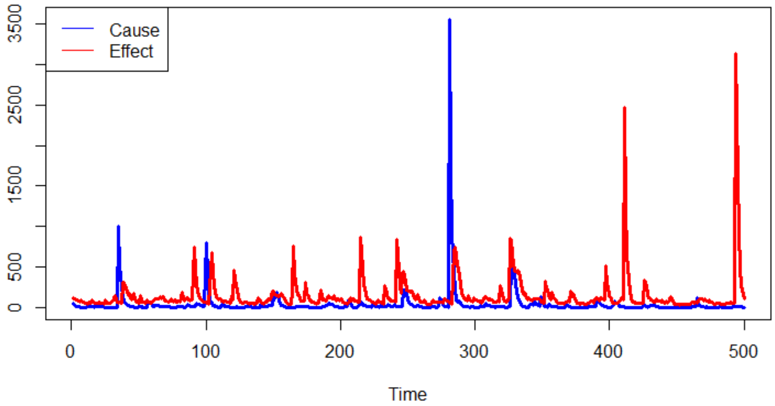

Figure 4.2: A sample realization of $(X,Y)^\top$ from Definition 4.3. Here, X is the blue line and Y the red one.

	First model				***First model***	
	Bootstrap method		Granger test		$\hat{\Gamma}^{time}_{X,Y} > \tau = 0.9$	
	$n = 500$	$n = 5000$	$n = 500$	$n = 5000$	$n = 500$	$n = 5000$
Sensitivity	65%	60%	100%	100%	*99%*	*100%*
Specificity	86%	96%	95%	96%	*79%*	*100%*
	Second model				***Second model***	
	Bootstrap method		Granger test		$\hat{\Gamma}^{time}_{X,Y} > \tau = 0.9$	
	$n = 500$	$n = 5000$	$n = 500$	$n = 5000$	$n = 500$	n=5000
Sensitivity	75%	73%	68%	52%	*90%*	*100%*
Specificity	65%	90%	33%	43%	*60%*	*100%*

Table 4.2: We consider two time series models, one is simple VAR model, the second is more complex nonlinear model with a common cause. We used three methods for the causal inference. The resulting percentage shows how many times the result was correct (in each direction, sensitivity corresponds to the $X \to Y$ direction, specificity to the other one). For the first and second tests, the results represent the percentage of cases when the corresponding p-value was less than $\alpha = 0.05$. For the third method (it is not a formal test), sensitivity 99% represents that in 99% of cases was $\hat{\Gamma}_{X,Y} > \tau$. Specificity 79% represents that in 79% of cases was $\hat{\Gamma}_{Y,X} \leq \tau$.

We will also use the method (not a formal test!), when we estimate $\hat{\Gamma}^{time}_{X,Y}(3)$ and conclude that X causes Y if and only if $\hat{\Gamma}^{time}_{X,Y}(3) \geq \tau$ with the choice $\tau = 0.9$.

We perform 100 simulations of the time series mentioned above, with the number of data $n = 500, 5000$. Finally, we compute the number of correctly inferring the causal directions $X \to Y$ and $Y \not\to X$ using these three methods. Table 4.2 shows the results in percentage. Here, the specificity represents the percentage of correct conclusions "X causes Y". The sensitivity represents the percentage of correct conclusions "Y does not cause X".

The results suggest that the Granger test behaves well under a simple model but can not handle a more complex model. The bootstrap method does not behave properly either. It is a common problem of bootstrap that we obtain smaller confidence intervals, and it is anti-conservative. Even for $n = 5000$, we still rejected the hypothesis $\Gamma^{time}_{X,Y}(3) = 1$ in 40% of cases even though the hypothesis holds. Therefore, using the bootstrap method may not be an appropriate approach here.

4.4 Choice of a threshold

A common problem in the extremal value theory is a choice of a threshold. In our case, it is the choice of the parameter k from Definition 4.1. There is a bias-variance trade-off; the smaller k, the smaller the bias (and the larger the variance) and vice versa. There is no universal method how to choose the threshold. It very much depends on the extremes of our series.

To give an idea of the behaviour, consider the time series model given by Definition 4.2 with the number of data $n = 1000$. Figure 4.3 represents the estimators $\hat{\Gamma}^{time}_{X,Y}(2)$ and $\hat{\Gamma}^{time}_{Y,X}(2)$ using different k. The thick line represents the mean of 100 realizations; the thin lines are 5% and 95% quantiles. The variance of $\hat{\Gamma}^{time}_{Y,X}(2)$ for small k is very large. On the other hand, the larger k, the more is $\hat{\Gamma}^{time}_{X,Y}(2)$ negatively biased.

Concluding from this example (and a few others), it seems that $k = \sqrt{n}$ may be a reasonable choice. We want to emphasize that it may not be optimal and does not behave well in general.

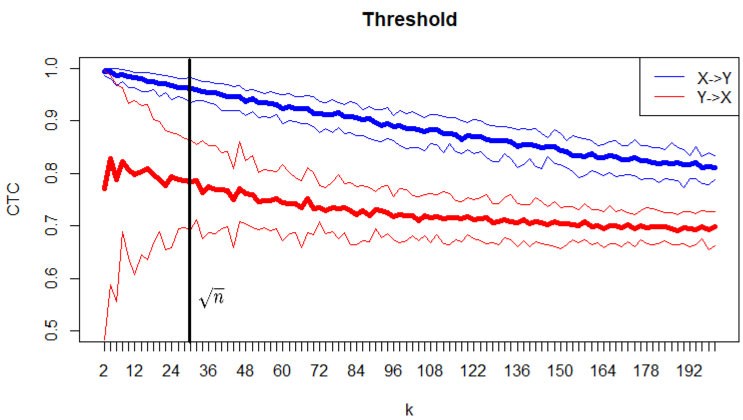

Figure 4.3: The figure represents the estimators $\hat{\Gamma}^{time}_{X,Y}(2)$ (blue) and $\hat{\Gamma}^{time}_{Y,X}(2)$ (red) with different choices of a parameter k for a specific model of time series with $n = 1000$. The thick line represents the mean of 100 realizations, the thin lines are 5% and 95% quantiles.

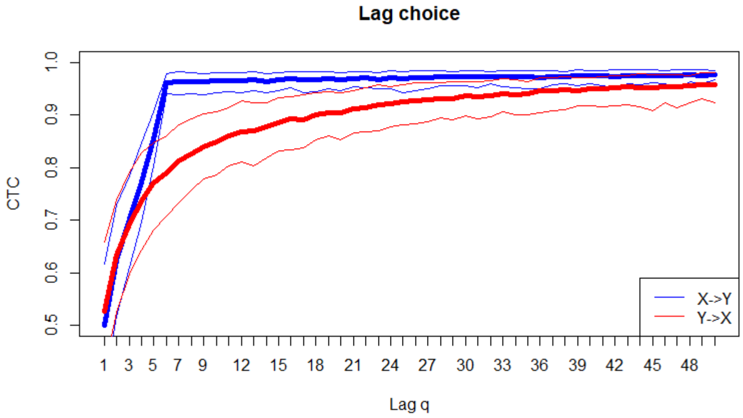

Figure 4.4: The figure represents the estimators $\hat{\Gamma}^{time}_{X,Y}(q)$ (blue) and $\hat{\Gamma}^{time}_{Y,X}(q)$ (red) with different choices of the lag q for a specific model of time series with $n = 1000$. The thick line represents the mean of 100 realizations, the thin lines are 5% and 95% quantiles.

4.5 Choice of the lag

How does $\Gamma^{time}_{X,Y}(q)$ behave for different lags q? To give an example, consider a model for $(X, Y)^\top$where

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_t^X, \\ Y_t &= 0.5Y_{t-1} + 0.5X_{t-6} + \varepsilon_t^Y, \end{aligned}$$

and $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Cauchy$. Notice that the minimal lag is $q = 6$.

Similarly as in Subsection 4.4, we did many simulations from this model with $n = 1000$ and computed $\hat{\Gamma}^{time}_{X,Y}(q)$ and $\hat{\Gamma}^{time}_{Y,X}(q)$ for different q. The mean, 5% and 95% quantiles of those estimates are drawn in Figure 4.4.

Usually (at least on the artificial datasets where X causes Y), the coefficient $\hat{\Gamma}^{time}_{X,Y}(q)$ in the causal direction rises much faster than in the other direction, until it reaches the "correct" lag. Then, this coefficient is very close to 1, and it stays there even for all larger q (just as the theory suggests). On the other hand, $\hat{\Gamma}^{time}_{Y,X}(q)$ rises slower and slowly but steadily converges to 1.

4.6 Application

In the following, we will be dealing with a problem from space weather studies. The term "space weather" refers to the variable conditions on the Sun and in space that can influence the performance of technology we use on Earth. Extreme space weather could potentially cause damage to critical infrastructure – especially the electric grid. In order to protect people and systems that might be at risk from space weather effects, we need to understand the causes of space weather [5].

Geomagnetic storms and substorms are indicators of geomagnetic activity. Visually, a substorm is seen as a sudden brightening and increased movement of

[5]Text taken from a webpage https://www.ready.gov/space-weather, accessed 18.5.2021.

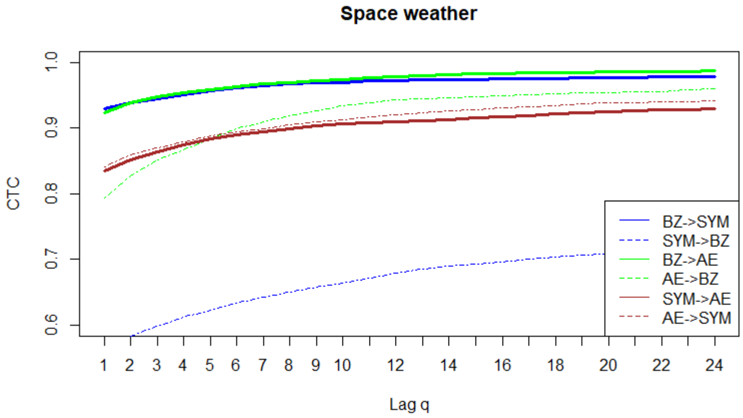

Figure 4.5: The figure refers to the real dataset from Section 4.6. It represents all values of $\hat{\Gamma}_{\cdot,\cdot}(q)$ for a range of lag $q \in [1, 24]$, with all pairs of time series SYM (magnetic storm index), AE (substorm index) and BZ (interplanetary magnetic field). We can see an asymmetry in the causal influence between the time series BZ-SYM and BZ-AE.

auroral arcs. It can cause magnetic field disturbances in the auroral zones up to a magnitude of 1000 nT (Tesla units). The basis of this geomagnetic activity begins in the Sun; specifically, there is a significant correlation with the solar wind (stream of negatively charged particles from the Sun) and also with an interplanetary magnetic field (a component of solar magnetic field dragged away from the Sun by the solar wind).

One of the fundamental problems in this area is determining and predicting some specific characteristics – magnetic storm index (SYM) and a substorm index (AE). It may seem that AE is a driving factor (cause) of SYM because usually, the accumulation of successive substorms precedes the occurrence of magnetic storms. However, a recent article Manshour et al. [2021] induces that it is not the case. A vertical component of an interplanetary magnetic field (BZ) seems to be a common cause of both of these indexes. We will apply our method to check this result and determine if the causal influence manifests itself in the extremes.

Our data consist of three time series (SYM, AE, BZ) with about 100000 measurements (every 5 minutes for the entire year 2000). Data (together with a commented R code) are available in the supplementary package or online[6]. Their plot can be found in Figure 4.6. From the nature of the data, we will compare extremes when SYM is extremely small, AE extremely large, BZ is extremely small (i.e. taking −SYM, +AE, −BZ and comparing only maximums). We also know that an appropriate lag will be smaller than $q = 24$ (2 hours).

First things first, we will discuss if the assumptions for our method are fulfilled. We estimate the tail indexes of our data[7]. Resulting numbers are the following: SYM has the estimated tail index 0.23 $(0.1, 0.37)$, AE has 0.18 $(0.1, 0.25)$ and BZ has 0.16 $(0.05, 0.31)$. Therefore, the assumption of the regular variation with the same tail index seems reasonable. Moreover, all confidence intervals do not

[6]NASA webpage https://cdaweb.gsfc.nasa.gov, accessed 18.05.2021.

[7]Using "HTailIndex" function in "ExtremeRisks" package in R with variable k=500, reference Padoan and Stupfler [2020].

include the zero value (although BZ is quite close), and therefore our time series can be considered heavy-tailed. The time series also seem stationary, therefore our assumptions seem reasonable for this application.

Finally, we will compute the causal tail coefficient with different lags (results were similar if we consider k other than $\sqrt{n}$). The resulting numbers can be found in Figure 4.5. They suggest that there is a strong asymmetry between BZ and SYM ($\hat{\Gamma}_{BZ,SYM} \approx 1$, and $\hat{\Gamma}_{SYM,BZ} \ll 1$), the asymmetry between BZ and AE ($\hat{\Gamma}_{BZ,AE} \approx 1$, and $\hat{\Gamma}_{SYM,AE} < 1$) and no asymmetry between SYM and AE ($\hat{\Gamma}_{AE,SYM} < 1$ is only negligibly larger than $\hat{\Gamma}_{AE,SYM}$).

Such results correspond to the hypothesis about BZ being a common cause of SYM and AE, with no causal relation between them. Note that our method can deal with a common cause (at least in theory). Although classical methods suggest that AE causes SYM, there are some extremal events where AE is extreme, but SYM is not.

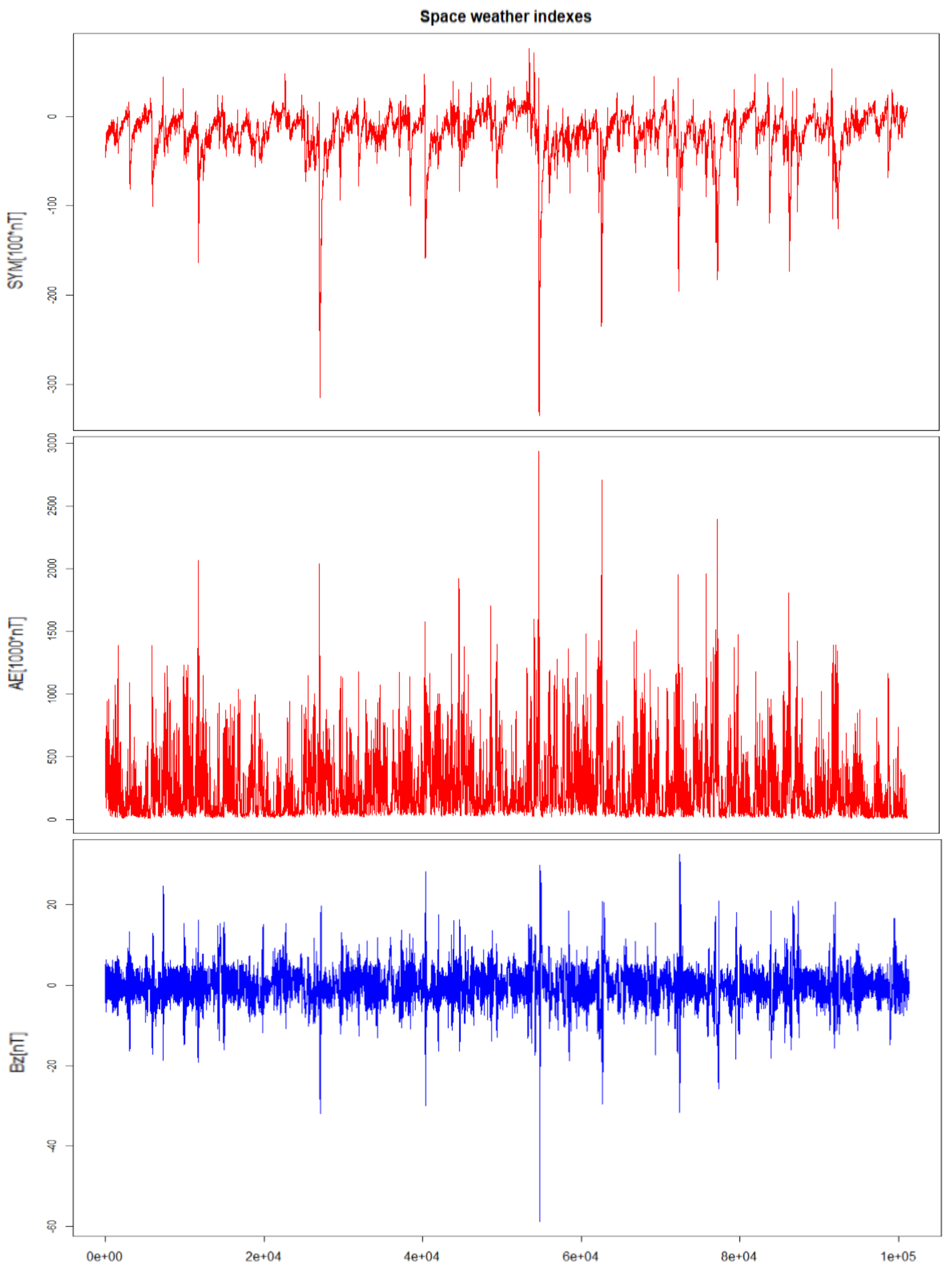

Figure 4.6: Space weather. The first plot represents SYM (magnetic storm index), the second one AE (substorm index) and the last one BZ (interplanetary magnetic field). Data were measured in 5 minutes intervals for a year 2000 by NASA.

5. Auxiliary propositions

5.1 Proposition 5.1

Proposition 5.1. *Let $X, Y, (\varepsilon_i, i \in \mathbb{N})$ be independent continuous random variables with support on some neighbourhood of infinity. Let $a_i, b_i \geq 0$, $i \in \mathbb{N}$ and $M_1, M_2 \in \mathbb{R}$ be constants. Then*

$$P(X + Y > M_1 \mid a_1 X + a_2 Y > M_2) \geq P(X + Y > M_1),$$

or more generally,

$$P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1 \mid \sum_{i=1}^{\infty} b_i \varepsilon_i > M_2) \geq P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1),$$

provided that the sums are a.s. summable, non-trivial.

Proof. First, we will prove that for all $n \in \mathbb{N}$ holds the following:

$$P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1 \mid \sum_{i=1}^{n} b_i \varepsilon_i > M_2) \geq P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1).$$

We will use the following fact: Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, then for any non-decreasing functions $f, g : \mathbb{R}^n \to \mathbb{R}$ holds

$$cov(f(\varepsilon), g(\varepsilon)) \geq 0.$$

This is a well-known result from the theory of associated random variables, see e.g. Theorem 2.1 in Esary et al. [1967]. Take $f(x_1, \ldots, x_n) = 1\{\sum_{i=1}^{n} a_i x_i > M_1\}, g(x_1, \ldots, x_n) = 1\{\sum_{i=1}^{n} b_i x_i > M_2\}$. They are non-decreasing because $a_i, b_i \geq 0$. Therefore, we obtain

$$\begin{aligned} 0 &\leq cov(f(\varepsilon), g(\varepsilon)) \\ &= P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1, \sum_{i=1}^{n} b_i \varepsilon_i > M_2) - P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1) P(\sum_{i=1}^{n} b_i \varepsilon_i > M_2). \end{aligned}$$

Dividing by $P(\sum_{i=1}^{n} b_i \varepsilon_i > M_2)$ (which is positive), we obtain the desirable inequality.

Next, we will give an argument for the infinite case. Denote $X_n := \sum_{i=1}^{n} a_i \varepsilon_i$, $X := \sum_{i=1}^{\infty} a_i \varepsilon_i$,$Y_n := \sum_{i=1}^{n} b_i \varepsilon_i$, $X := \sum_{i=1}^{\infty} b_i \varepsilon_i$. We know that $X_n \overset{a.s.}{\to} X$ and $Y_n \overset{a.s.}{\to} Y$ from the assumptions. Therefore,

$$\begin{aligned} P(X_n > M_1 \mid Y_n > M_2) &= \frac{P(X_n > M_1, Y_n > M_2)}{P(Y_n > M_2)} \overset{n\to\infty}{\to} \frac{P(X > M_1, Y > M_2)}{P(Y > M_2)} \\ &= P(X > M_1 \mid Y > M_2). \end{aligned}$$

But we know that

$$P(X_n > M_1 \mid Y_n > M_2) \geq P(X_n > M_1) \overset{n\to\infty}{\to} P(X > M_1).$$

Combining these results, our proposition follows. ☐

5.2 Proposition 5.2

Proposition 5.2.

- *Let* $(\varepsilon_i^X, \varepsilon_i^Y, i \in \mathbb{Z})$ *be iid* $RV(\theta)$ *continuous random variables.*
- *Let* $a_i, b_i, c_i \geq 0$ *be constants, such that for some* $\delta > 0$ *is* $\sum_{i=0}^{\infty} a_i^{\theta-\delta} < \infty, \sum_{i=0}^{\infty} b_i^{\theta-\delta} < \infty, \sum_{i=0}^{\infty} c_i^{\theta-\delta} < \infty$ *(i.e. all* $\sum_{i=0}^{\infty} a_i\varepsilon_i^X, \sum_{i=0}^{\infty} b_i\varepsilon_i^X, \sum_{i=0}^{\infty} c_i\varepsilon_i^Y$ *are a.s. summable).*
- *Denote* $A = \sum_{i=0}^{\infty} a_i^{\theta}, B = \sum_{i=0}^{\infty} b_i^{\theta}, C = \sum_{i=0}^{\infty} c_i^{\theta}$*, for which it holds that* $A, B, C \in (0, \infty)$*.*
- *Let* $\Phi = \{i \in \mathbb{N} \cup \{0\} : b_i > 0 = a_i\}$*.*

Then

$$\lim_{u\to\infty} P(\sum_{i=0}^{\infty} a_i\varepsilon_i^X < M \mid \sum_{i=0}^{\infty} b_i\varepsilon_i^X + \sum_{i=0}^{\infty} c_i\varepsilon_i^Y > u) = P(\sum_{i=0}^{\infty} a_i\varepsilon_i^X < M)\frac{C + \sum_{i\in\Phi} b_i^{\theta}}{C + B},$$

which holds for all $M \in \mathbb{R}$*.*

We will consider only those $M \in \mathbb{R}$ such that $P(\sum_{i=0}^{\infty} a_i\varepsilon_i^X < M) > 0$, otherwise the statement is trivial. We will prove this proposition using the following series of lemmas.

Lemma 5.1. *Let* $X, Y \sim RV(\theta)$ *be independent. Then*

$$\lim_{u\to\infty} P(X < M \mid X + Y > u) = P(X < M) \lim_{u\to\infty} \frac{P(Y > u)}{P(Y > u) + P(X > u)},$$

for every $M \in \mathbb{R}$*.*

Lemma 5.2. *Under the conditions from Proposition 5.2,*

$$\lim_{u\to\infty} P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0; i\notin\Phi}^{n} b_i\varepsilon_i^X > u) = 0,$$

which holds for all $n \in \mathbb{N}$*.*

Lemma 5.3. *Let* $Z \sim RV(\theta)$ *be independent of* $(\varepsilon_i^X, i \in \mathbb{Z})$*. Under the conditions from Proposition 5.2*

$$\begin{aligned}
&\lim_{u\to\infty} P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0; i\notin\Phi}^{n} b_i\varepsilon_i^X + Z > u) \\
&= P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M) \lim_{u\to\infty} \frac{P(Z > u)}{P(Z > u) + P(\sum_{i=0; i\notin\Phi}^{n} b_i\varepsilon_i^X > u)},
\end{aligned}$$

which holds for all $n \in \mathbb{N}$*.*

Proof of Lemma 5.1. Using the Bayes theorem, we obtain

$$\lim_{u\to\infty} P(X < M \mid X + Y > u) = \lim_{u\to\infty} P(X + Y > u \mid X < M)\frac{P(X < M)}{P(X + Y > u)}.$$

For the denominator we use the sum-equivalence $P(X+Y>u) \sim P(X>u)+P(Y>u)$. Therefore, it is enough to show that $P(X+Y>u \mid X<M) \sim P(Y>u)$.

Now, let W be a random variable independent of Y with a distribution satisfying $P(W \leq t) = P(X \leq t \mid X < M)$ for all $t \in \mathbb{R}$. Then, $P(X+Y>u \mid X<M) = P(W+Y>u)$. We obviously have $\lim_{u\to\infty} \frac{P(W>u)}{P(Y>u)} = 0$ and we can use e.g. Theorem 2.1. from Bingham et al. [2006] to obtain $\lim_{u\to\infty} \frac{P(X+Y>u|X<M)}{P(Y>u)} = \lim_{u\to\infty} \frac{P(Y+W>u)}{P(Y>u)} = 1$. Therefore $\lim_{u\to\infty} P(X+Y>u \mid X<M) = \lim_{u\to\infty} P(Y>u)$, what we wanted to prove. □

Proof of Lemma 5.2. WLOG $\Phi = \emptyset$, otherwise we have only lower n. Denote $\omega = \min_{i\leq n} a_i$, it holds that $\omega > 0$. In this proof only, we will denote $B = \sum_{i=0}^n b_i$, and $A = \sum_{i=0}^n a_i$. The following events relation hold:

$$\begin{aligned}\{\sum_{i=0}^n a_i\varepsilon_i^X < M; \sum_{i=0}^n b_i\varepsilon_i^X > u\} &\subset \{\exists j \leq n : \varepsilon_j^X > \frac{u}{B}, \sum_{i=0}^n a_i\varepsilon_i^X < M\} \\ &\subset \{\exists i,j \leq n : \varepsilon_j^X > \frac{u}{B}, \varepsilon_i^X < \frac{M-\frac{\omega u}{B}}{A}\}.\end{aligned}$$

(Simply put, there needs to be one large and one small $\varepsilon_\cdot^X$). Therefore, we can rewrite

$$\begin{aligned}&\lim_{u\to\infty} P(\sum_{i=0}^n a_i\varepsilon_i^X < M \mid \sum_{i=0;i\notin\Phi}^n b_i\varepsilon_i^X > u) \\ &= \lim_{u\to\infty} \frac{P(\sum_{i=0}^n a_i\varepsilon_i^X < M; \sum_{i=0}^n b_i\varepsilon_i^X > u)}{P(\sum_{i=0}^n b_i\varepsilon_i^X > u)} \\ &\leq \lim_{u\to\infty} \frac{P(\exists i,j \leq n : \varepsilon_i^X < \frac{M-\frac{\omega u}{B}}{A}, \varepsilon_j^X > \frac{u}{B}\})}{P(\sum_{i=0}^n b_i\varepsilon_i^X > u)} \\ &\leq \lim_{u\to\infty} \frac{n(n+1)P(\varepsilon_1^X < \frac{M-\frac{\omega u}{B}}{A}, \varepsilon_2^X > \frac{u}{B})}{P(\sum_{i=0}^n b_i\varepsilon_i^X > u)} \\ &= n(n+1)\lim_{u\to\infty} \frac{P(\varepsilon_1^X < \frac{M-\frac{\omega u}{B}}{A})P(\varepsilon_2^X > \frac{u}{B})}{\sum_{i=0}^n b_i^\theta \cdot P(\varepsilon_2^X > u)} \\ &= n(n+1)\lim_{u\to\infty} P(\varepsilon_1^X < \frac{M-\frac{\omega u}{B}}{A})\frac{B^\theta}{\sum_{i=0}^n b_i^\theta} = 0.\end{aligned}$$

□

Proof of Lemma 5.3. WLOG $\Phi = \emptyset$, otherwise we have only lower n. Proof is very similiar to the proof of Lemma 5.1. In this proof only, we will denote $B = \sum_{i=0}^n b_i$, and $A = \sum_{i=0}^n a_i$.

Let W be a random variable independent of Z with a distribution satisfying $P(W \leq t) = P(\sum_{i=0}^n b_i\varepsilon_i^X \leq t \mid \sum_{i=0}^n a_i\varepsilon_i^X < M)$ for all $t \in \mathbb{R}$. Then, $P(X+Y>u \mid X<M) = P(W+Y>u)$. Using Bayes theorem, we have

$$
\begin{aligned}
&\lim_{u\to\infty} P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0}^{n} b_i\varepsilon_i^X + Z > u) \\
&= \lim_{u\to\infty} P(\sum_{i=0}^{n} b_i\varepsilon_i^X + Z > u \mid \sum_{i=0}^{n} a_i\varepsilon_i^X < M) \frac{P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X + Z > u)} \\
&= P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M) \lim_{u\to\infty} \frac{P(W + Z > u)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X + Z > u)} \\
&= P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M) \lim_{u\to\infty} \frac{P(W > u) + P(Z > u)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u) + P(Z > u)}.
\end{aligned}
$$

In the last equality, we used the fact that W does not have a heavier tail than Z and therefore we can use the sum-equivalence.

All we need to prove is that $\lim_{u\to\infty} \frac{P(W>u)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u)} = 0$. Again, using the Bayes theorem, we obtain

$$
\begin{aligned}
\lim_{u\to\infty} \frac{P(W > u)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u)} &= \lim_{u\to\infty} \frac{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u \mid \sum_{i=0}^{n} a_i\varepsilon_i^X < M)}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u)} \\
&= \lim_{u\to\infty} \frac{P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0}^{n} b_i\varepsilon_i^X > u) \frac{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u)}{P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M)}}{P(\sum_{i=0}^{n} b_i\varepsilon_i^X > u)} \\
&= \lim_{u\to\infty} \frac{1}{P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M)} P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0}^{n} b_i\varepsilon_i^X > u).
\end{aligned}
$$

The rest follows from Lemma 5.2. □

Proof of Proposition 5.2. Let $\delta > 0$, define $\zeta = 1 - \sqrt{1-\delta} > 0$[1] and choose large $n_0 \in \mathbb{N}$ such that the following hold:

- $P(|\sum_{i=n_0+1}^{\infty} a_i\varepsilon_i^X| > \delta) < \delta$,
- $\frac{\sum_{i=0}^{n_0} b_i^\theta + C}{\sum_{i=0}^{\infty} b_i^\theta + C} > 1 - \zeta$,
- $P(|\sum_{i=n_0+1; i\notin\Phi}^{\infty} b_i\varepsilon_i^X| < \delta) > 1 - \zeta$.

Denote

- $E = \sum_{i=0}^{n_0} a_i\varepsilon_i^X, F = \sum_{i=n_0+1}^{\infty} a_i\varepsilon_i^X$,
- $G = \sum_{i=0; i\notin\Phi}^{n_0} b_i\varepsilon_i^X, H = \sum_{i=n_0+1; i\notin\Phi}^{\infty} b_i\varepsilon_i^X$,
- $Z = \sum_{i=0}^{\infty} c_i\varepsilon_i^Y + \sum_{i\in\Phi} b_i\varepsilon_i^X$.

Then, E, F, Z and also G, H, Z are pair-wise independent. With our notation, we *want* to prove that

$$
\begin{aligned}
&\lim_{u\to\infty} P(E + F < M \mid G + H + Z > u) \\
&\stackrel{?}{=} P(E + F < M) \lim_{u\to\infty} \frac{P(Z > u)}{P(G + H > u) + P(Z > u)}.
\end{aligned}
$$

[1] $1 - \sqrt{1-\delta}$ is a solution of $1 - (1-\zeta)(1-\zeta) = \delta$. When $\delta \to 0$ then also $\zeta \to 0$.

It is enough, because due to the sum-equivalence holds

$$P(Z > u) \sim [\sum_{i=0}^{\infty} c_i^\theta + \sum_{i\in\Phi} b_i^\theta] P(\varepsilon_1^X > u),$$

and the denominator $P(G + H + Z > u) \sim [\sum_{i=0}^{\infty} c_i^\theta + \sum_{i=0}^{\infty} b_i^\theta] P(\varepsilon_1^X > u)$.

First, due to Lemma 5.1

$$\begin{aligned}
&\lim_{u\to\infty} P(H > \delta \mid H + (G + Z) > u) = 1 - \lim_{u\to\infty} P(H \le \delta \mid H + (G + Z) > u) \\
&= 1 - P(H \le \delta) \lim_{u\to\infty} \frac{P(G + Z > u)}{P(G + Z + H > u)} = 1 - P(H \le \delta) \frac{\sum_{i=0}^{n_0} b_i^\theta + C}{\sum_{i=0}^{\infty} b_i^\theta + C} \\
&< 1 - (1 - \zeta)(1 - \zeta) = \delta.
\end{aligned}$$

Second, using previous results and independence of F and G, Z we obtain

$$\begin{aligned}
&\lim_{u\to\infty} P(F > \delta \mid G + H + Z > u) \\
&= \lim_{u\to\infty} \frac{P(F > \delta, G + H + Z > u, H > \delta)}{P(G + H + Z > u)} + \frac{P(F > \delta, G + H + Z > u, H \le \delta)}{P(G + H + Z > u)} \\
&\le \lim_{u\to\infty} \frac{P(G + H + Z > u, H > \delta)}{P(G + H + Z > u)} + \frac{P(F > \delta, G + Z > u - \delta)}{P(G + H + Z > u)} \\
&= \lim_{u\to\infty} P(H > \delta \mid G + H + Z > u) + \frac{P(F > \delta) P(G + Z > u - \delta)}{P(G + H + Z > u)} \\
&< \delta + P(F > \delta) \lim_{u\to\infty} \frac{P(G + Z > u)}{P(G + Z > u) + P(H > u)} \le \delta + P(F > \delta) < 2\delta.
\end{aligned}$$

Finally, we obtain (first inequality is trivial, second uses the previous inequality and independence, the equality follows from Lemma 5.3, the next two inequalities follow from the sum-equivalence and trivial $P(H > u) \ge 0$, and the last inequality follows from $P(F + \delta > 0) > 1 - \delta$):

$$\begin{aligned}
&\lim_{u\to\infty} P(E + F < M \mid G + H + Z > u) \\
&\ge \lim_{u\to\infty} P(E + \delta < M; F < \delta \mid G + (H + Z) > u) \\
&\ge (1 - 2\delta) \lim_{u\to\infty} P(E + \delta < M \mid G + (H + Z) > u) \\
&= (1 - 2\delta) P(E < M - \delta) \lim_{u\to\infty} \frac{P(H + Z > u)}{P(H + Z > u) + P(G > u)} \\
&\ge (1 - 2\delta) P(E < M - \delta) \lim_{u\to\infty} \frac{P(H > u) + P(Z > u)}{P(Z > u) + P(H > u) + P(G > u)} \\
&\ge (1 - 2\delta) P(E < M - \delta) \lim_{u\to\infty} \frac{P(Z > u)}{P(Z > u) + P(G + H > u)} \\
&\ge (1 - 2\delta)(1 - \delta) P(E + (F + \delta) < M - \delta) \lim_{u\to\infty} \frac{P(Z > u)}{P(Z > u) + P(G + H > u)}.
\end{aligned}$$

When we send $\delta \to 0$ we finally obtain

$$\begin{aligned}
&P(E + F < M \mid G + H + Z > u) \\
&\ge P(E + F < M) \lim_{u\to\infty} \frac{P(Z > u)}{P(Z > u) + P(G + H > u)},
\end{aligned}$$

what we wanted to show. The inequality in the other direction can be done analogously. $\square$

Consequence. *Under the conditions from Proposition 5.2*

$$\lim_{u\to\infty} P(\sum_{i=0}^{\infty} a_i|\varepsilon_i^X| < M \mid \sum_{i=0}^{\infty} b_i\varepsilon_i^X + \sum_{i=0}^{\infty} c_i\varepsilon_i^Y > u) = P(\sum_{i=0}^{\infty} a_i|\varepsilon_i^X| < M)\frac{C + \sum_{i\in\Phi} b_i^\theta}{C+B}.$$

Proof. The proof can be done analogously as the proof of the Proposition 5.2. Modified Lemma 5.1 and Lemma 5.3 are still valid, just with $|\varepsilon_i^X|$ instead of ε_i^X in the equations. Modification for Lemma 5.2 is trivial, because

$$\begin{aligned}
&\lim_{u\to\infty} P(\sum_{i=0}^{n} a_i|\varepsilon_i^X| < M \mid \sum_{i=0;i\notin\Phi}^{n} b_i\varepsilon_i^X > u) \\
&\leq \lim_{u\to\infty} P(\sum_{i=0}^{n} a_i\varepsilon_i^X < M \mid \sum_{i=0;i\notin\Phi}^{n} b_i\varepsilon_i^X > u) = 0.
\end{aligned}$$

The limiting argument for $n\to\infty$ remains the same. $\square$

5.3 Proposition 5.3

Proposition 5.3. *Let $(X,Y)^\top$ follow NAR model, specified by*

$$\begin{aligned}
X_t &= f(X_{t-1}) + \varepsilon_t^X, \\
Y_t &= g_1(Y_{t-1}) + g_2(X_{t-q}) + \varepsilon_t^Y,
\end{aligned}$$

where f, g_1, g_2 are continuous and satisfy $\lim_{x\to\infty} h(x) = \infty$ and $\lim_{x\to\infty}\frac{h(x)}{x} < 1$, $h = f, g_1, g_2$. Moreover, let $\varepsilon, \varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$ be non-negative. If $(X,Y)^\top$ is stationary, then

$$\lim_{u\to\infty}\frac{P(Y_t > u)}{P(\varepsilon > u)} < \infty.$$

Lemma 5.4. *Under assumptions of Proposition 5.3*

$$\lim_{u\to\infty}\frac{P(X_t > u)}{P(\varepsilon > u)} < \infty.$$

Proof of Lemma 5.4. Let $c = \lim_{x\to\infty}\frac{f(x)}{x} \in [0,1)$. First, notice that

$$\lim_{u\to\infty}\frac{P(f(X_t) > u)}{P(X_t > u)} = c^\theta.$$

Compute

$$\begin{aligned}
&\lim_{u\to\infty}\frac{P(X_t > u)}{P(\varepsilon > u)} = \lim_{u\to\infty}\frac{P(f(X_{t-1}) + \varepsilon_t^X > u)}{P(\varepsilon > u)} \\
&= 1 + \lim_{u\to\infty}\frac{P(f(X_{t-1}) > u)}{P(\varepsilon > u)} \leq 1 + c^\theta \lim_{u\to\infty}\frac{P(X_{t-1} > u)}{P(\varepsilon > u)} \\
&= 1 + c^\theta \lim_{u\to\infty}\frac{P(X_t > u)}{P(\varepsilon > u)}.
\end{aligned}$$

We have used max-sum equivalence, independence and the previous equation. Therefore, we have $\lim_{u\to\infty}\frac{P(X_t>u)}{P(\varepsilon>u)} = \frac{1}{1-c^\theta} < \infty$. □

Proof of Proposition 5.3. Find $c < 1, K \in \mathbb{R}$ such that for all $x > 0$ is

$$f(x) < K + cx, g_1(x) < K + cx, g_2(x) < K + cx.$$

Specially note that $f(x + y) \leq (K + cx) + (K + cy)$. Then, a.s. holds

$$\begin{aligned}
Y_0 &= \varepsilon_0^Y + g_2(X_{-q}) + g_1(Y_{-1}) \leq \varepsilon_0^Y + g_2(X_{-q}) + K + cY_{-1} \\
&\leq \varepsilon_0^Y + g_2(X_{-q}) + K + c(\varepsilon_{-1}^Y + g_2(X_{-q-1}) + K + cY_{-2}) \\
&\leq (\varepsilon_0^Y + c\varepsilon_{-1}^Y + c^2\varepsilon_{-2}^Y + \dots)+ \\
&+ (g_2(X_{-q}) + cg_2(X_{-q-1}) + c^2 g_2(X_{-q-2}) + \dots) + (K + cK + c^2K + \dots) \\
&= \sum_{i=0}^{\infty} c^i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} c^iK + \sum_{i=0}^{\infty} c^i g_2(X_{q-i}) \leq \sum_{i=0}^{\infty} c^i\varepsilon_{-i}^Y + \frac{2K}{1-c} + \sum_{i=0}^{\infty} c^{i+1}X_{q-i}.
\end{aligned}$$

Finally, because X_i and ε_j^Y are all independent, holds

$$\begin{aligned}
&\lim_{u\to\infty}\frac{P(Y_t > u)}{P(\varepsilon > u)} \leq \lim_{u\to\infty}\frac{P(\sum_{i=0}^{\infty} c^i\varepsilon_{-i}^Y + \frac{2K}{1-c} + \sum_{i=0}^{\infty} c^{i+1}X_{q-i} > u)}{P(\varepsilon > u)} \\
&= \lim_{u\to\infty}\frac{P(\sum_{i=0}^{\infty} c^i\varepsilon_{-i}^Y > u) + P(\sum_{i=0}^{\infty} c^{i+1}X_{q-i} > u)}{P(\varepsilon > u)} \\
&= \sum_{i=0}^{\infty} c^{i\theta} + \lim_{u\to\infty}\frac{P(\sum_{i=0}^{\infty} c^{i+1}X_{q-i} > u)}{P(\varepsilon > u)} < \infty,
\end{aligned}$$

where we used regular variation property, sum-equivalence, and the Lemma 5.4. □

Remark. We proved a stronger claim. We showed that for every Heavy-tailed NAR model, there exists stable $VAR(q)$ sequence which is a.s. larger. Note that $VAR(q)$ process defined by

$$\begin{aligned}
X_t &= aX_{t-1} + \varepsilon_t^X, \\
Y_t &= bY_{t-1} + dX_{t-q} + \varepsilon_t^Y,
\end{aligned}$$

with $0 < a, b, d < 1$, is stable.

6. Proofs of theorems

Observation: Let X, Y be continuous random variables with support on some neighbourhood of infinity, and F_X, F_Y their distribution functions. Then,

$$\lim_{u\to 1^-} \mathbb{E}\left[F_Y(Y) \mid F_X(X) > u\right] = 1$$

if and only if $\lim_{u\to\infty} P(Y > M \mid X > u) = 1$ for every $M \in \mathbb{R}$.

Proof. Trivial. □

6.1 Theorem 2.1.

Theorem 2.1. (Heavy-tailed VAR model). Let $(X, Y)^\top$ be a time series which follows Heavy-tailed VAR model. If X causes Y then $\Gamma^{time}_{X,Y}(q) = 1$.

Proof. Because X causes Y, for some $p \leq q$ is $\delta_p > 0$.

Then,

$$\begin{aligned}\Gamma^{time}_{X,Y}(q) &= \lim_{u\to 1^-} \mathbb{E}\left[\max\{F_Y(Y_0), \dots, F_Y(Y_q)\} \mid F_X(X_0) > u\right] \\ &\geq \lim_{u\to 1^-} \mathbb{E}\left[F_Y(Y_p) \mid F_X(X_0) > u\right] = \lim_{u\to\infty} \mathbb{E}\left[F_Y(Y_p) \mid X_0 > u\right].\end{aligned}$$

Now, if we prove that $\forall M \in \mathbb{R}$ is $\lim_{u\to\infty} P(Y_p > M \mid X_0 > u) = 1$, it will imply that $\lim_{u\to\infty} \mathbb{E}\left[F_Y(Y_p) \mid X_0 > u\right] = 1$. Rewrite

$$\begin{aligned}&\lim_{u\to\infty} P(Y_p > M \mid X_0 > u) \\ &= \lim_{u\to\infty} P(\delta_p X_0 + \sum_{i=1}^{q} \beta_i Y_{p-i} + \sum_{i=1; i\neq p}^{q} \delta_i X_{p-i} + \varepsilon_p^Y > M \mid X_0 > u) \\ &\geq \lim_{u\to\infty} P(\delta_p u + \sum_{i=1}^{q} \beta_i Y_{p-i} + \sum_{i=1; i\neq p}^{q} \delta_i X_{p-i} + \varepsilon_p^Y > M \mid X_0 > u).\end{aligned}$$

Now, using causal representation (Theorem 1.2), we can rewrite all

$$X_0 = \sum_{i=0}^{\infty} a_i \varepsilon_{-i}^X + \sum_{i=0}^{\infty} c_i \varepsilon_{-i}^Y$$

$$\sum_{i=1}^{q} \beta_i Y_{p-i} + \sum_{i=1; i\neq p}^{q} \delta_i X_{p-i} + \varepsilon_p^Y = \sum_{i=0}^{\infty} \phi_i \varepsilon_{p-i}^X + \sum_{i=0}^{\infty} \psi_i \varepsilon_{p-i}^Y$$

for some $\phi_i, \psi_i \geq 0$.

We obtain

$$\begin{aligned}&\lim_{u\to\infty} P(\delta_p u + \sum_{i=1}^{q} \beta_i Y_{p-i} + \sum_{i=1; i\neq p}^{q} \delta_i X_{p-i} > M \mid X_0 > u) \\ &= \lim_{u\to\infty} P(\sum_{i=0}^{\infty} \phi_i \varepsilon_{q-i}^X + \sum_{i=0}^{\infty} \psi_i \varepsilon_{q-i}^Y > M - \delta_p u \mid \sum_{i=0}^{\infty} a_i \varepsilon_{-i}^X + \sum_{i=0}^{\infty} c_i \varepsilon_{-i}^Y > u) \\ &\geq \lim_{u\to\infty} P(\sum_{i=0}^{\infty} \phi_i \varepsilon_{q-i}^X + \sum_{i=0}^{\infty} \psi_i \varepsilon_{q-i}^Y > M - \delta_p u) = 1,\end{aligned}$$

where we used Proposition 5.1 in the last step. Therefore, $\lim_{u\to\infty} P(Y_p > M \mid X_0 > u) \geq 1$, which proves the theorem. □

Theorem 2.1. (Heavy-tailed NAR model). Let $(X, Y)^\top$ be a time series which follows Heavy-tailed NAR model. If X causes Y then $\Gamma^{time}_{X,Y}(q) = 1$.

Proof. We proceed very similarly as in the proof of Heavy-tailed VAR model. We rewrite $\Gamma^{time}_{X,Y}(q) \geq \lim_{u\to\infty} \mathbb{E}\left[F_Y(Y_q) \mid X_0 > u\right]$, which is equal to 1 if $\forall M \in \mathbb{R}$ is $\lim_{u\to\infty} P(Y_q > M \mid X_0 > u) = 1$. We rewrite

$$\lim_{u\to\infty} P(Y_q > M \mid X_0 > u) = \lim_{u\to\infty} P(g_1(Y_{q-1}) + g_2(X_0) + \varepsilon_q^Y > M \mid X_0 > u).$$

Because X causes Y, it holds that g_2 is not constant and $\lim_{x\to\infty} g_2(x) = \infty$. This implies that there exists $x_0 \in \mathbb{R} : \forall x \geq x_0 : g_2(x) > M$. Therefore, for all $u > x_0$ holds

$$P(g_2(X_0) > M \mid X_0 > u) = 1.$$

Finally, we only use the fact that ε_t^Y and g_1 are non-negative.

$$\begin{aligned}
&\lim_{u\to\infty} P(g_1(Y_{q-1}) + g_2(X_0) + \varepsilon_t^Y > M \mid X_0 > u) \\
&\geq \lim_{u\to\infty} P(g_2(X_0) > M \mid X_0 > u) = 1,
\end{aligned}$$

what we wanted to prove. □

6.2 Theorem 2.2.

Theorem 2.2. (Heavy-tailed VAR model). Let $(X, Y)^\top$ be a time series which follows Heavy-tailed VAR model. If Y is not causing X then $\Gamma^{time}_{Y,X}(p) < 1$ for all $p \in \mathbb{N}$.

Proof. Let $M \in \mathbb{R}$ such that $P(X_0 < M) > 0$. We will show that

$$\lim_{u\to\infty} P(\max(X_0, \ldots, X_p) < M \mid Y_0 > u) > 0,$$

from which it follows that $\lim_{u\to\infty} \mathbb{E}\left[\max(F_X(X_0), \ldots, F_X(X_p)) \mid Y_0 > u\right] < 1$.

Rewrite

$$\begin{aligned}
P(\max(X_0, \ldots, X_p) < M \mid Y_0 > u) &= P(X_0 < M, \ldots, X_p < M \mid Y_0 > u) \\
&\geq P(|X_0| + |X_1| + \cdots + |X_p| < M \mid Y_0 > u).
\end{aligned}$$

Now, we will use causal representation of the time series, which, because we know that Y is not causing X, can be written in the form

$$\begin{aligned}
X_t &= \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}^X, \\
Y_t &= \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}^Y + \sum_{i=0}^{\infty} d_i \varepsilon_{t-i}^X.
\end{aligned}$$

We obtain

$$
\begin{aligned}
&P(\sum_{t=0}^{p}|X_t| < M \mid Y_0 > u) = P(\sum_{t=0}^{p}|\sum_{i=0}^{\infty} a_i\varepsilon_{t-i}^X| < M \mid Y_0 > u) \\
&\geq P(\sum_{t=0}^{p}\sum_{i=0}^{\infty} a_i|\varepsilon_{t-i}^X| < M \mid Y_0 > u) \\
&= P(\sum_{i=0}^{\infty} \phi_i|\varepsilon_{p-i}^X| < M \mid \sum_{i=0}^{\infty} b_i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{-i}^X > u),
\end{aligned}
$$

for $\phi_i = a_i + \cdots + a_{i-p}$ (we define $a_j = 0$ for $j < 0$). Finally, it follows from the consequence of Proposition 5.2 that

$$
\lim_{u\to\infty} P(\sum_{i=0}^{\infty} \phi_i|\varepsilon_{p-i}^X| < M \mid \sum_{i=0}^{\infty} b_i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{-i}^X > u) > 0,
$$

what we wanted to prove (Theorem 5.2 requires non-trivial sums, but if $\forall i : d_i = 0$ then the series are independent and this inequality holds trivially). $\square$

Theorem 2.2. (Heavy-tailed NAR model). Let $(X, Y)^\top$ be a time series which follows Heavy-tailed NAR model. If Y is not causing X then $\Gamma^{time}_{Y,X}(p) < 1$ for all $p \in \mathbb{N}$.

Proof. We have

$$
\begin{aligned}
X_t &= f(X_{t-1}) + \varepsilon_t^X, \\
Y_t &= g_1(Y_{t-1}) + g_2(X_{t-q}) + \varepsilon_t^Y.
\end{aligned}
$$

Choose large $M \in \mathbb{R}$, such that $\sup_{x<M} f(x) < M$ *and* such that [1]

$$
P(\varepsilon_0^X < M - \sup_{x<M} f(x)) > 0.
$$

Denote $M^\star = \sup_{x<M} f(x)$. Rewrite

$$
\begin{aligned}
&P(\max(X_0, \dots, X_q) < M \mid Y_0 > u) = P(X_0 < M, \dots, X_q < M \mid Y_0 > u) \\
&= \prod_{i=0}^{q} P(X_i < M \mid X_0 < M, \dots, X_{i-1} < M, Y_0 > u).
\end{aligned}
$$

Then, as in the proof of Heavy-tailed VAR model case, if we show that this is strictly larger than 0, it will imply that $\Gamma^{time}_{Y,X}(q) < 1$. We know that for every $i \geq 1$ holds the following

$$
\begin{aligned}
&\lim_{u\to\infty} P(X_i < M \mid X_0 < M, \dots, X_{i-1} < M, Y_0 > u) \\
&= \lim_{u\to\infty} P(f(X_{i-1}) + \varepsilon_i^X < M \mid X_0 < M, \dots, X_{i-1} < M, Y_0 > u) \\
&\geq \lim_{u\to\infty} P(M^\star + \varepsilon_i^X < M \mid X_0 < M, \dots, X_{i-1} < M, Y_0 > u) \\
&= P(M^\star + \varepsilon_i^X < M) > 0.
\end{aligned}
$$

[1] This is possible from the assumptions on continuity and the limit behaviour of f.

We only need to show for the case when $i = 0$ that $\lim_{u\to\infty} P(X_0 > M \mid Y_0 > u) < 1$. Let $Z = g_1(Y_{-1}) + g_2(X_{-q})$, Z is independent with ε_0^Y. After rewriting, we obtain

$$P(X_0 > M \mid Y_0 > u) = P(X_0 > M \mid \varepsilon_0^Y + Z > u) = \frac{P(X_0 > M; \varepsilon_0^Y + Z > u)}{P(\varepsilon_0^Y + Z > u)}.$$

Let $\frac{1}{2} < \delta < 1$ (we will send $\delta \to 1$). Now, note the following events relation

$$\begin{aligned}
&\{X_0 > M; \varepsilon_0^Y + Z > u\} \\
&\subseteq \{X_0 > M; \varepsilon_0^Y > \delta u\} \cup \{Z > \delta u\} \cup \{Z > (1-\delta)u; \varepsilon_0^Y > (1-\delta)u\}.
\end{aligned}$$

Applying it to the previous equation, we obtain

$$\begin{aligned}
&\lim_{u\to\infty} \frac{P(X_0 > M; \varepsilon_0^Y + Z > u)}{P(\varepsilon_0^Y + Z > u)} \\
&\leq \lim_{u\to\infty} \frac{P(X_0 > M; \varepsilon_0^Y > \delta u) + P(Z > \delta u) + P(Z > (1-\delta)u; \varepsilon_0^Y > (1-\delta)u)}{P(\varepsilon_0^Y + Z > u)} \\
&= \lim_{u\to\infty} \frac{P(X_0 > M)P(\varepsilon_0^Y > \delta u)}{P(\varepsilon_0^Y + Z > u)} + \frac{P(Z > \delta u)}{P(\varepsilon_0^Y + Z > u)} \\
&+ \lim_{u\to\infty} P(Z > (1-\delta)u) \frac{(\frac{1}{1-\delta})^\theta P(\varepsilon_0^Y > u)}{P(\varepsilon_0^Y + Z > u)} \\
&= \frac{1}{\delta^\theta} \lim_{u\to\infty} \frac{P(X_0 > M)P(\varepsilon_0^Y > u)}{P(\varepsilon_0^Y + Z > u)} + \frac{P(Z > \delta u)}{P(\varepsilon_0^Y + Z > u)} + 0.
\end{aligned}$$

The last element is 0 because $\lim_{u\to\infty} P(Z > (1-\delta)u) = 0$ and $\frac{P(\varepsilon_0^Y > u)}{P(\varepsilon_0^Y + Z > u)} \leq 1$ (simply because Z is non-negative random variable).

Now, we will use the result from Proposition 5.3. In the case when $\lim_{u\to\infty} \frac{P(Z>u)}{P(\varepsilon_0^Y > u)} = 0$, we obtain (see e.g. Lemma 1.3.2 in Kulik and Soulier [2020]) $\lim_{u\to\infty} \frac{P(\varepsilon_0^Y > u)}{P(\varepsilon_0^Y + Z > u)} = 1$ and $\lim_{u\to\infty} \frac{P(Z>u)}{P(\varepsilon_0^Y + Z > u)} = 0$. Therefore,

$$\lim_{u\to\infty} \frac{\frac{1}{\delta^\theta} P(X_0 > M)P(\varepsilon_0^Y > u) + P(Z > \delta u)}{P(\varepsilon_0^Y + Z > u)} = \frac{1}{\delta^\theta} P(X_0 > M) < 1,$$

for δ close enough to 1.

On the other hand, if $\lim_{u\to\infty} \frac{P(Z>u)}{P(\varepsilon_0^Y > u)} = c \in \mathbb{R}^+$, we also have that $Z \sim RV(\theta)$ (this follows trivially from the definition of regular variation, tails behaviour is the same up to a constant). Therefore, we can apply sum-equivalence and we obtain

$$\begin{aligned}
&\lim_{u\to\infty} \frac{\frac{P(X_0 > M)}{\delta^\theta} P(\varepsilon_0^Y > u) + P(Z > \delta u)}{P(\varepsilon_0^Y + Z > u)} \\
&= \frac{1}{\delta^\theta} \lim_{u\to\infty} \frac{P(X_0 > M)P(\varepsilon_0^Y > u) + P(Z > u)}{P(\varepsilon_0^Y > u) + P(Z > u)} \\
&= \frac{1}{\delta^\theta} \lim_{u\to\infty} \frac{P(X_0 > M)P(\varepsilon_0^Y > u) + cP(\varepsilon_0^Y > u)}{P(\varepsilon_0^Y > u) + cP(\varepsilon_0^Y > u)} \\
&= \frac{1}{\delta^\theta} \frac{P(X_0 > M) + c}{1 + c},
\end{aligned}$$

which is less than 1 for δ close enough to 1. Therefore, we obtained $\lim_{u\to\infty} P(X_0 > M \mid Y_0 > u) < 1$, what we wanted to prove. $\square$

6.3 Theorem 3.1.

Theorem 3.1. Let $(X,Y)^\top$ be a time series which follows Heavy-tailed VAR model, with possibly negative coefficients, satisfying the extremal causal condition. Moreover, let $\varepsilon_t^X, \varepsilon_t^Y$ have full support on $\mathbb{R}$, are iid satisfying tail balance condition. If X causes Y, but Y does not cause X, then $\Gamma^{time}_{|X|,|Y|}(q) = 1$, and $\Gamma^{time}_{|Y|,|X|}(q) < 1$.

Proof. First, we will show that if Y does not cause X, then $\Gamma^{time}_{|Y|,|X|}(q) < 1$. This holds even without the extremal causal condition. Similarly as in the proof of Theorem 2.1, it is enough to show that for some $M > 0$ is $\lim_{u\to\infty} P(|\sum_{i=0}^{\infty} a_i\varepsilon_{t-i}^X| > M \mid |\sum_{i=0}^{\infty} b_i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{-i}^X| > u) < 1$ for $t \le q$.

We will use the following fact. Because we assumed that ε_i^X are $RV(\theta)$ and satisfy tail balance condition, the following holds:

$$P(|\sum_{i=0}^{\infty} a_i\varepsilon_{t-i}^X| > u) \sim [\sum_{i=0}^{\infty} |a_i|^\theta]P(|\varepsilon_0^X| > u) \sim P(\sum_{i=0}^{\infty} |a_i||\varepsilon_{t-i}^X| > u),$$

see e.g. page 6 in Jessen and Mikosch [2006]. Second step follows simply from the sum-equivalence. Finally, we use this fact and the triangle inequality to obtain the following relations

$$\begin{aligned}
&P(|\sum_{i=0}^{\infty} a_i\varepsilon_{t-i}^X| > M \mid |\sum_{i=0}^{\infty} b_i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{-i}^X| > u)\\
&\le \frac{P(\sum_{i=0}^{\infty} |a_i||\varepsilon_{t-i}^X| > M; \sum_{i=0}^{\infty} |b_i||\varepsilon_{-i}^Y| + \sum_{i=0}^{\infty} |d_i||\varepsilon_{-i}^X| > u)}{P(|\sum_{i=0}^{\infty} b_i\varepsilon_{-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{-i}^X| > u)}\\
&\sim \frac{P(\sum_{i=0}^{\infty} |a_i||\varepsilon_{t-i}^X| > M; \sum_{i=0}^{\infty} |b_i||\varepsilon_{-i}^Y| + \sum_{i=0}^{\infty} |d_i||\varepsilon_{-i}^X| > u)}{P(\sum_{i=0}^{\infty} |b_i||\varepsilon_{-i}^Y| + \sum_{i=0}^{\infty} |d_i||\varepsilon_{-i}^X| > u)}\\
&= P(\sum_{i=0}^{\infty} |a_i||\varepsilon_{t-i}^X| > M \mid \sum_{i=0}^{\infty} |b_i||\varepsilon_{-i}^Y| + \sum_{i=0}^{\infty} |d_i||\varepsilon_{-i}^X| > u).
\end{aligned}$$

This is for $u \to \infty$ less than 1 due to the classical non-negative case from Proposition 5.2 (for any $M \in \mathbb{R}$ such that $P(|\sum_{i=0}^{\infty} a_i\varepsilon_{t-i}^X| > M) < 1$).

Second, we will show that if X causes Y, then $\Gamma^{time}_{|X|,|Y|}(q) = 1$. Similarly, as in the proof of Theorem 2.1, it is enough to show that for every $M \in \mathbb{R}$ is

$$\lim_{u\to\infty} P(|Y_p| < M \mid |X_0| > u) = 0.$$

Here, $p \le q$ is some index with $\delta_p \ne 0$. Using causal representation with the same notation as in the proof of Theorem 2.1,

$$\begin{aligned}
&\lim_{u\to\infty} P(|\sum_{i=0}^{\infty} b_i\varepsilon_{p-i}^Y + \sum_{i=0}^{\infty} d_i\varepsilon_{p-i}^X| < M \mid |\sum_{i=0}^{\infty} a_i\varepsilon_{-i}^X| > u)\\
&\le \lim_{u\to\infty} P(\sum_{i=0}^{\infty} |b_i||\varepsilon_{p-i}^Y| + \sum_{i=0}^{\infty} |d_i||\varepsilon_{p-i}^X| < M \mid |\sum_{i=0}^{\infty} |a_i||\varepsilon_{-i}^X| > u),
\end{aligned}$$

where we used the same trick as in the first part of the proof. Therefore, we simplified our model and obtained the classical non-negative case. The result follows from the previous theory. Using Lemma 5.2 we obtain the result for finite n,

$$\lim_{u\to\infty} P(\sum_{i=0}^{n} |b_i||\varepsilon^Y_{p-i}| + \sum_{i=0}^{n} |d_i||\varepsilon^X_{p-i}| < M \mid |\sum_{i=0}^{n} |a_i||\varepsilon^X_{-i}| > u) = 0,$$

because due to the extremal causal condition is $\Phi = \emptyset$. The argument for limiting case $n \to \infty$ follows the same steps as those in the proof of Proposition 5.2. $\square$

6.4 Theorem 3.2

Theorem 3.2. Let $(X, Y, Z)^\top$ follow three-dimensional stable $VAR(q)$ model, with non-negative coefficients, where independent noise variables have $RV(\theta)$ distribution. Let Z be a common cause of both X and Y, and neither X nor Y are causing Z. If Y does not cause X, then $\Gamma^{time}_{Y,X}(q) < 1$.

Proof. Let our series have the following representation:

$$\begin{aligned} Z_t &= \sum_{i=0}^{\infty} a_i \varepsilon^Z_{t-i}, \\ X_t &= \sum_{i=0}^{\infty} b_i \varepsilon^X_{t-i} + \sum_{i=0}^{\infty} c_i \varepsilon^Z_{t-i}, \\ Y_t &= \sum_{i=0}^{\infty} d_i \varepsilon^X_{t-i} + \sum_{i=0}^{\infty} e_i \varepsilon^Y_{t-i} + \sum_{i=0}^{\infty} f_i \varepsilon^Z_{t-i}. \end{aligned}$$

Just as in the proof of Theorem 2.2, it is enough to show that $\lim_{u\to\infty} P(X_t > M|Y_0 > u) < 1$ for some $M > 0$. After rewriting,

$$\lim_{u\to\infty} P(\sum_{i=0}^{\infty} b_i \varepsilon^X_{t-i} + \sum_{i=0}^{\infty} c_i \varepsilon^Z_{t-i} > M \mid \sum_{i=0}^{\infty} d_i \varepsilon^X_{-i} + \sum_{i=0}^{\infty} e_i \varepsilon^Y_{-i} + \sum_{i=0}^{\infty} f_i \varepsilon^Z_{-i} > u) < 1,$$

which follows from Proposition 5.2 (two countable sums can be written as one countable sum). $\square$

6.5 Lemma 3.2

Lemma 3.2. Let $(X, Y)^\top$ follow the Heavy-tailed VAR model, where X causes Y. Let p be the minimal lag. Then, $\Gamma^{time}_{X,Y}(r) < 1$ for all $r < p$, and $\Gamma^{time}_{X,Y}(r) = 1$ for all $r \geq p$.

Proof. The second part, i.e. proving that $\Gamma^{time}_{X,Y}(r) = 1$ for all $r \geq p$, is an obvious consequence of the proof of Theorem 2.1 (in the first row of the proof, instead of choosing *some* $p \leq q : \delta_p > 0$, we choose p to be the minimal lag).

Concerning the first part, we only need to prove that $\Gamma^{time}_{X,Y}(p-1) < 1$, because then also $\Gamma^{time}_{X,Y}(p-i) \leq \Gamma^{time}_{X,Y}(p-1) < 1$. As in the proof of Theorem 2.2, we only need to show that $\lim_{u\to\infty} P(Y_{p-1} < M|X_0 > u) > 0$ for some $M > 0$. By rewriting to its causal representation, we obtain the following relation

$$\lim_{u\to\infty} P(\sum_{i=0}^{\infty} b_i \varepsilon_{p-1-i}^Y + \sum_{i=0}^{\infty} d_i \varepsilon_{p-1-i}^X < M \mid \sum_{i=0}^{\infty} a_i \varepsilon_{-i}^X + \sum_{i=0}^{\infty} c_i \varepsilon_{-i}^Y > u) > 0.$$

We only need to realize that $d_i = 0$ for $i \in \{1, \dots, p-1\}$ because p is the minimal lag. Therefore, ε_0^X is independent of Y_{p-1} and the rest follows from Proposition 5.2 (where we deal with the two sums as one, and single ε_0^X is the second "sum"). □

6.6 Theorem 4.1

Theorem 4.1. Let $(X, Y)^\top = ((X_t, Y_t)^\top, t \in \mathbb{Z})$ be a stationary bivariate time series, whose marginal distributions are absolutely continuous with support on some neighbourhood of infinity. Let $\Gamma_{X,Y}^{time}(q)$ exists. Let k_n satisfy (4.1) and

$$\frac{n}{k_n} P(\frac{n}{k_n} \sup_{x\in\mathbb{R}} |\hat{F}_X(x) - F(x)| > \delta) \overset{n\to\infty}{\to} 0, \ \ \forall \delta > 0. \tag{4.2}$$

Then, $\mathbb{E}\, \hat{\Gamma}_{X,Y}^{time}(q) \overset{n\to\infty}{\to} \Gamma_{X,Y}^{time}(q)$ [2].

Proof. Throughout the proof, we will use the copula fact that $P(F_X(X_1) \le t) = t$ for $t \in [0, 1]$ and the fact following from the stationarity $P(\hat{F}_X(X_1) \le \frac{k}{n}) = P(X_1 \le X_{(k)}) = \frac{k}{n}$, for $k \le n, k \in \mathbb{N}$. Please note that $X_{(k)}$ is always meant with respect to (not written) index n.

First, notice the following (third equation follows from the linearity of expectation and stationarity of our series. Fourth equation follows from the definition of conditional expectation. Fifth is quite trivial):

$$\begin{aligned}
\mathbb{E}\, \hat{\Gamma}_{X,Y}^{time}(q) &= \mathbb{E}\, \frac{1}{k_n} \sum_{i: X_i \ge \tau_{k_n}^X} \max\{\hat{F}_Y(Y_i), \dots, \hat{F}_Y(Y_{i+q})\} \\
&= \mathbb{E}\, \frac{1}{n} \sum_{i=1}^{n} \frac{n}{k_n} \max\{\hat{F}_Y(Y_i), \dots, \hat{F}_Y(Y_{i+q})\} 1[\hat{F}_X(X_i) > 1 - \frac{k_n}{n}] \\
&= \frac{n}{k_n}\, \mathbb{E}\, [\hat{F}_Y(\max\{Y_1, \dots, Y_{q+1}\}) 1[\hat{F}_X(X_1) > 1 - \frac{k_n}{n}]] \\
&= \frac{n}{k_n} P(\hat{F}_X(X_1) > 1 - \frac{k_n}{n}) \cdot \\
&\qquad \cdot \mathbb{E}\, [\hat{F}_Y(\max\{Y_1, \dots, Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}] \\
&= \mathbb{E}\, [\hat{F}_Y(\max\{Y_1, \dots, Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}].
\end{aligned}$$

Now, use $\hat{F} = F + \hat{F} - F$ to obtain

$$\begin{aligned}
&\mathbb{E}\, [\hat{F}_Y(\max\{Y_1, \dots, Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}] \\
&= \mathbb{E}\, [F_Y(\max\{Y_1, \dots, Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}] \\
&+ \mathbb{E}\, [(\hat{F}_Y - F_Y)(\max\{Y_1, \dots, Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}].
\end{aligned}$$

[2] Do not forget that $\hat{\Gamma}_{X,Y}^{time}(q)$ depends on n.

The second term is less than $\mathbb{E}\left[\sup_{x\in\mathbb{R}} |\hat{F}_Y(x) - F_Y(x)|\right] \to 0$ as $n \to \infty$ from the assumptions. All we need to show is that the first term converges to $\Gamma^{time}_{X,Y}(q)$. Rewrite

$$\begin{aligned}
&\mathbb{E}\left[F_Y(\max\{Y_1,\dots,Y_{q+1}\}) \mid \hat{F}_X(X_1) > 1 - \frac{k_n}{n}\right] \\
&= \mathbb{E}\left[F_Y(\max\{Y_1,\dots,Y_{q+1}\}) \mid X_1 > X_{(n-k_n)}\right].
\end{aligned}$$

Therefore, all we *need* to show is the following

$$\begin{aligned}
\Gamma^{time}_{X,Y}(q) &= \lim_{u\to\infty} \mathbb{E}\left[F_Y(\max\{Y_1,\dots,Y_{q+1}\}) \mid X_1 > u\right] \\
&\stackrel{?}{=} \lim_{n\to\infty} \mathbb{E}\left[F_Y(\max\{Y_1,\dots,Y_{q+1}\}) \mid X_1 > X_{(n-k_n)}\right].
\end{aligned}$$

Denote $Z = F_Y(\max\{Y_1,\dots,Y_{q+1}\})$. Choose $u_n \in \mathbb{R}$ such as $1 - \frac{k_n}{n}$ quantiles of X_1, i.e. numbers fulfilling $P(X_1 > u_n) = \frac{k_n}{n}$. Because $u_n \to \infty$ it is sufficient to show that

$$\lim_{n\to\infty} \mathbb{E}\left[Z \mid X_1 > u_n\right] \stackrel{?}{=} \lim_{n\to\infty} \mathbb{E}\left[Z \mid X_1 > X_{(n-k_n)}\right].$$

Rewrite (using identity $1[a > b] = 1[c > a > b] + 1[a > c > b] + 1[a > b > c]$ when no ties are present):

$$\begin{aligned}
&\mathbb{E}\left[Z \mid X_1 > u_n\right] = \frac{1}{P(X_1 > u_n)} \int_\Omega Z \cdot 1[X_1 > u_n] dP = \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > u_n] dP \\
&= \frac{n}{k_n} \int_\Omega Z \cdot 1[X_{(n-k_n)} > X_1 > u_n] dP + \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > X_{(n-k_n)} > u_n] dP \\
&+ \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > u_n > X_{(n-k_n)}] dP.
\end{aligned}$$

On the other hand, rewrite also

$$\begin{aligned}
&\mathbb{E}\left[Z \mid X_1 > X_{(n-k_n)}\right] = \frac{1}{P(X_1 > X_{(n-k_n)})} \int_\Omega Z \cdot 1[X_1 > X_{(n-k_n)}] dP \\
&= \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > X_{(n-k_n)}] dP = \frac{n}{k_n} \int_\Omega Z \cdot 1[u_n > X_1 > X_{(n-k_n)}] dP \\
&+ \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > X_{(n-k_n)} > u_n] dP + \frac{n}{k_n} \int_\Omega Z \cdot 1[X_1 > u_n > X_{(n-k_n)}] dP.
\end{aligned}$$

Note that these two equations differ only in the first term. Therefore, to show the equality, we only need to show that

$$\lim_{n\to\infty} \frac{n}{k_n} \int_\Omega Z \cdot 1[X_{(n-k_n)} > X_1 > u_n] dP - \frac{n}{k_n} \int_\Omega Z \cdot 1[u_n > X_1 > X_{(n-k_n)}] dP = 0.$$

We will show that the first term goes to 0. The second term can be shown analogously that it also converges to 0, and from that we will have proven that this limit goes to 0.

We know that $0 \leq Z \leq 1$, so we know that for the first term holds the following:

$$
\begin{aligned}
&\frac{n}{k_n}\int_\Omega Z\cdot 1[X_{(n-k_n)}>X_1>u_n]dP \le \frac{n}{k_n}P(X_{(n-k_n)}>X_1>u_n)\\
&= P(X_{(n-k_n)}>X_1 \mid X_1>u_n) = P(X_{(n-k_n)}>X_1 \mid F_X(X_1)>1-\frac{k_n}{n})\\
&= 1-P(X_1\ge X_{(n-k_n)} \mid F_X(X_1)>1-\frac{k_n}{n})\\
&= 1-P(\hat F_X(X_1)\ge 1-\frac{k_n}{n} \mid F_X(X_1)>1-\frac{k_n}{n})\\
&= 1-P(F_X(X_1)+(\hat F_X(X_1)-F_X(X_1))\ge 1-\frac{k_n}{n} \mid F_X(X_1)>1-\frac{k_n}{n})\\
&\le 1-P(F_X(X_1)-\sup_{x\in\mathbb{R}}|\hat F_X(x)-F_X(x)|\ge 1-\frac{k_n}{n} \mid F_X(X_1)>1-\frac{k_n}{n}).
\end{aligned}
$$

Denote $S_n := \sup_{x\in\mathbb{R}}|\hat F_X(x)-F_X(x)|$. It is sufficient for our proof to show that

$$
P(F_X(X_1)-S_n\ge 1-\frac{k_n}{n} \mid F_X(X_1)>1-\frac{k_n}{n}) \overset{n\to\infty}{\to} 1.
$$

Choose $\varepsilon>1$, define $\delta = 1-\frac{1}{\varepsilon}$. Rewrite

$$
\begin{aligned}
&P(F_X(X_1)-S_n\ge 1-\frac{k_n}{n} \mid F_X(X_1)>1-\frac{k_n}{n})\\
&= \frac{n}{k_n}P(F_X(X_1)-S_n\ge 1-\frac{k_n}{n}; F_X(X_1)>1-\frac{k_n}{n})\\
&\ge \frac{n}{k_n}P(F_X(X_1)-S_n\ge 1-\frac{k_n}{n}; F_X(X_1)>1-\frac{k_n/\varepsilon}{n})\\
&\ge \frac{n}{k_n}P(S_n\le \frac{k_n-k_n/\varepsilon}{n}; F_X(X_1)>1-\frac{k_n/\varepsilon}{n})\\
&= \frac{n}{k_n}P(\frac{n}{k_n}S_n\le\delta; F_X(X_1)>1-\frac{k_n/\varepsilon}{n}).
\end{aligned}
$$

Use the identity $P(A\cap B) = 1-P(A^c)-P(B^c)+P(A^c\cap B^c)$ and continue

$$
\begin{aligned}
&\frac{n}{k_n}P(\frac{n}{k_n}S_n\le\delta; F_X(X_1)>1-\frac{k_n/\varepsilon}{n})\\
&= \frac{n}{k_n}[1-P(\frac{n}{k_n}S_n>\delta)-P(F_X(X_1)\le 1-\frac{k_n/\varepsilon}{n})\\
&+P(\frac{n}{k_n}S_n>\delta; F_X(X_1)\le 1-\frac{k_n/\varepsilon}{n}))]\\
&\ge \frac{n}{k_n}[1-P(\frac{n}{k_n}S_n>\delta)-(1-\frac{k_n/\varepsilon}{n})+0]\\
&= \frac{n}{k_n}[\frac{k_n/\varepsilon}{n}-P(\frac{n}{k_n}S_n>\delta)] = \frac{1}{\varepsilon}-\frac{n}{k_n}P(\frac{n}{k_n}S_n>\delta) \overset{n\to\infty}{\to} \frac{1}{\varepsilon} \overset{\varepsilon\to 1}{\to} 1.
\end{aligned}
$$

All together, we proved that $\lim_{n\to\infty}\mathbb{E}[Z \mid X_1>u_n] = \lim_{n\to\infty}\mathbb{E}[Z \mid X_1 > X_{(n-k_n)}]$, from which the theorem follows. $\square$

Conclusion

In this thesis, we dealt with an open problem of detecting causal relations between two possibly nonlinear heavy-tailed time series. We proposed an original method how to estimate the causal influence from the extremes. The causal tail coefficient for time series,

$$\Gamma^{time}_{X,Y}(q) := \lim_{u \to 1^-} \mathbb{E}\left[\max\{F_Y(Y_0), \dots, F_Y(Y_q)\} \mid F_X(X_0) > u\right],$$

was defined in Chapter 2. Under certain assumptions is $\Gamma^{time}_{X,Y}(q) = 1$ if and only if X causes Y (for an appropriate q). It is the core of our method for detecting the causal directions. This holds for regularly varying time series, even in a certain class of nonlinear relations and without any assumptions on the bulk of the distributions. To our knowledge, no literature deals with such a case.

We rigorously proved some properties of this coefficient $\Gamma^{time}_{X,Y}$ and discussed its extensions. Many examples of how this coefficient can be computed were provided, and it was shown that even an unobserved common cause does not change the output. We also discussed some possibilities and methods on how to estimate the lag between the time series.

The problem of estimating this coefficient was solved by choosing a statistic

$$\hat{\Gamma}^{time}_{X,Y}(q) := \frac{1}{k} \sum_{i: X_i \geq \tau^X_k} \max\{\hat{F}_Y(Y_i), \dots, \hat{F}_Y(Y_{i+q})\}.$$

We showed on a simulation study how such a method works in practice. We also programmed everything in an R language. Finally, we applied our method to a real dataset concerning a geomagnetic storms. We confirmed results demonstrated by another article, implying that the interplanetary magnetic field from the Sun is a common cause of both geomagnetic storms and substorms, using NASA dataset.

This work can potentially have a broad impact on the causal inference theory. It sheds light on some connections between causality and extremes. Many scientific disciplines use causal inference as a baseline of their work. A method that can detect a causal direction in complex heavy-tailed datasets can be very useful in some domains.

This topic provides a wide range of possibilities for future research. For example, can $\Gamma^{time}_{X,Y}(q)$ be written as some function of an extremogram? What is the distribution of $\hat{\Gamma}^{time}_{X,Y}(q)$? Can we create some better (consistent) statistic where we cancel the negative bias of $\hat{\Gamma}^{time}_{X,Y}(q)$ and provide better testing than by a bootstrap? Does this method work even for some light-tailed time series? Can a similar method be used in neural networks or machine learning? These questions can lead to potential future research and important results.

Bibliography

J. Andel. On nonlinear models for time series. *Statistics*, 1989. doi: 10.1080/02331888908802217.

R. N. Bhattacharya and C. Lee. Ergodicity of nonlinear first order autoregressive models. *J Theor Probab*, 1995. doi: 10.1007/BF02213462.

N. Bingham, Ch. Goldie, and E. Omey. Regularly varying probability densities. *Publications de L'Institut mathematique*, 2006. doi: 10.2298/PIM0694047B.

R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 2009. doi: 10.3150/09-bej213.

J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 1967. doi: 10.1214/aoms/11776987012.

S. Foss, D. Korschunov, and S. Zachary. An introduction to heavy-tailed and subexponential distributions. *Springer*, 2013. doi: 10.1007/978-1-4614-7101-1.

N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke. Causal discovery in heavy-tailed models. 2020. URL `https://arxiv.org/abs/1908.05097`.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969. URL `http://www.jstor.org/stable/1912791`.

A. Gut. *Probability: A Graduate Course*. Springer, 2013. ISBN 978-1-4614-4708-5. doi: 10.1007/b138932.

T. Hesterberg. What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 2014. doi: 10.1080/00031305.2015.1089789.

A. Jessen and T. Mikosch. Regularly varying functions. *Publications De L'institut Mathematique*, 2006. doi: 10.2298/PIM0694171J.

R. Kulik and P. Soulier. *Heavy-Tailed Time Series*. Springer, 2020. ISBN 978-1-0716-0735-0. doi: 10.1007/978-1-0716-0737-4.

H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Publishing Company, Incorporated, 2007. ISBN 3540262393. doi: 10.1007/978-3-540-27752-1.

P. Manshour, G. Balasis, G. Consolini, C. Papadimitriou, and M. Paluš. Causality and information transfer between the solar wind and the magnetosphere–ionosphere system. *Entropy*, 2021. doi: 10.3390/e23040390.

T. Mikosch. Regular variation, subexponentiality and their applications in probability theory. *International Journal of Production Economics - INT J PROD ECON*, 1999. URL `https://www.eurandom.tue.nl/reports/1999/013-report.pdf`.

T. Mikosch and G. Samorodnitsky. The supremum of a negative drift random walk with dependent heavy-tailed steps. *The Annals of Applied Probability*, 2000. doi: 10.1214/aoap/1019487517.

N. Moloney, D. Faranda, and Y. Sato. An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2019. doi: 10.1063/1.5079656.

A. Necir, R. Abdelaziz, and R. Zitikis. Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, 2010. doi: 10.1155/2010/596839.

S. Padoan and G. Stupfler. *ExtremeRisks: Extreme Risk Measures*, 2020. URL `https://CRAN.R-project.org/package=ExtremeRisks`. R package version 0.0.4.

S. Palachy. Inferring causality in time series data, 2019. URL `https://towardsdatascience.com/inferring-causality-in-time-series-data-b8b75fe52c46`.

M. Paluš, K. Hlaváčková-Schindler, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 2007. doi: 10.1016/j.physrep.2006.12.004.

J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017. ISBN 0262037319.

Z. Prášková. *Stochastic Processes 2 - Lecture Notes*. 2017. URL `https://www2.karlin.mff.cuni.cz/~praskova/anglicky2016.pdf`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer Publishing Company, Incorporated, 1987. ISBN 978-0-387-75953-1. doi: 10.1007/978-0-387-75953-1.

R. Tsay and R. Chen. *Nonlinear Time Series Analysis*. John Wiley and Sons, 2018. ISBN 9781119514312. doi: 10.1002/9781119514312.

N. Wiener. The theory of prediction. *Modern Mathematics for Engineers*, 1956.

J. Yang and A. Hongzhi. Nonlinear autoregressive models with heavy-tailed innovation. *Science China-mathematics*, 2005. doi: 10.1360/03za00321.

A. Zeileis and T. Hothorn. Diagnostic checking in regression relationships. *R News*, 2002. URL `https://CRAN.R-project.org/doc/Rnews/`.