FACULTY OF MATHEMATICS AND PHYSICS
Charles University

PRELIMINARY MASTER THESIS

Juraj Bodik

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.; RNDr. Milan Paluš, DrSc.
Study programme:	Probability, Mathematical Statistics and Econometrics
Study branch:	Spatial Modelling

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

Dedication. TODO

Title: Detection of causality in time series using extreme values

Author: Juraj Bodik

department: Department of Probability and Mathematical Statistics

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

Consultant: RNDr. Milan Paluš, DrSc.

Abstract: This thesis deals 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, which, under some assumptions, correctly detect 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 of this approach is that this method works even if non-linear relations and common ancestors are present. Moreover, we will mention how this method can help detect a time delay between the two time series. Finally, we will show some simulations and an actual application of how this method performs in practice.

V tejto praci riesime nasledovny problem: Mame dve stacionarne casove rady ktorych marginalne distribucie maju tazke chvosty. My chceme zistit, ci maju kauzalny vztah, teda ci zmena v jednej z nich sposobi zmenu v druhej. Otazka, ci nahodne premenne maju kauzalny suvis alebo su iba korelovane, je dolezita v mnohych oblastiach vedy. Bezne metody na detekciu kauzalit nefunguju dobre ak sa vzajomne vztahy prejavuju vyhradne pri extremnych hodnotach. V tejto praci navrhneme novy sposob, ako v takomto netradicnom pripade rozlisovat medzi korelaciou a kauzalitou. Definujeme si tzv kauzalny chvostovy koeficient, ktory za istych predpokladov detekuje asymetricke kauzalne vztahy medzi dvoma casovymi radami. Toto tvrdenie rigorozne dokazeme a navrhneme sposob akym kauzalny chvostovy koeficient statisticky odhadneme iba z konecneho mnozstva dat. Vyhodou je ze tato metoda funguje aj pri nelinearnych vztahoch medzi casovymi radami a aj pri pritomnosti spolocnej priciny. Navyse spomenieme sposob akym tato metoda moze pomoct pri zistovani casoveho posunu medzi dvoma casovymi radami. Konecne, na simulaciach a aplikacii ukazeme ako tato metoda funguje v praxi.

Keywords: Causality Time series Extremal value theory Heavy tail VAR processes

Contents

Introduction

0.0.1 Background

The ultimate goal of causal inference is to predict future values and to understand relations between random variables. It can be used in almost every scientific field. In medicine, to test whether a vaccine or some drug causes an improvement in health. 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 defines the mathematical definition of causality for time series was Granger [1980]. 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. 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 is known to follow a heavy-tailed distribution, there is almost none 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 requires linear relations (i.e. assuming a simple VAR model). Of course, in most real-life cases, restricting to linear dependencies is enough to obtain accurate results. On the other hand, it can be interesting in theory to deal with more general cases.

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 bivariate strong stationary time series, defined by the following recurrent relations

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

where $\varepsilon_t^X, \varepsilon_t^Y$ are iid, following a Pareto distribution. A sample realization of such a model is at the 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).

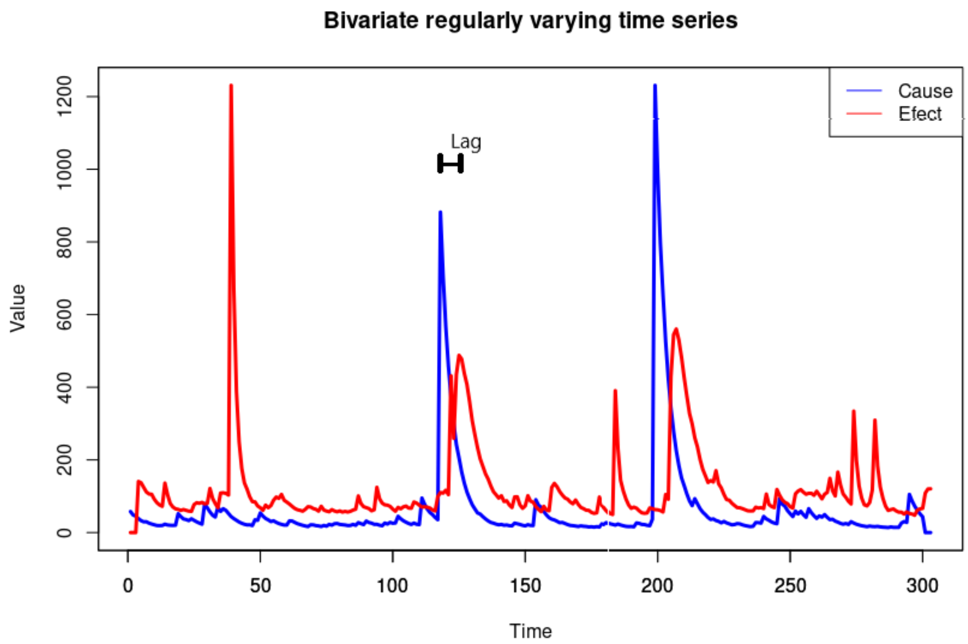

Figure 1: We can see that large values of blue time series causes large values of red time series, but not the other way. Here, it is easy to see that the blue is causing the red. The lag represents the time delay between the time series, in this case is equal to 5.

0.0.4 Main idea

Consider we have data such as in Figure 1; we want to detect a causal relation 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,

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

where, in our case, is $q=5$. 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, 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 holds. Similar idea was first used in Gnecco et al. [2020], which was not dealing with time series but a structural causal models.

0.0.5 Thesis organization

The thesis is organised as follows. The first chapter gives some preliminaries about time series, their specific models, extremal value theory, some causal definitions, and finally, we propose a specific VAR time series generalisation. The second chapter consists of the main results, together with a model example of the method. The third chapter 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 mentioned above. Chapter four TODO. Finally, chapter 5 and 6 consist of proofs, where chapter 5 deals with auxiliary propositions, and chapter 6 with direct proofs of theorems from the previous chapters. The thesis finishes with a Table of notations and Bibliography.

1. Preliminaries

This chapter will give a short review of some results from extremal value theory, causal inference, and multivariate time series analysis. In Section 1.4. we propose a generalization of the VAR process, which we call $GAM(q)$ processes.

1.1 Time series

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

If the reader is not familiar with this topic, we recommend the references Prášková [2017] for the univariate time series, and Lutkepohl [2007] for multivariate time series and VAR processes.

1.1.1 Some basics

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a sample space, $\mathcal{F}$ is a sigma 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 X_t are defined on the same probability space. 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.

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, and by a stationary process, we will always mean strict stationarity.

1.1.2 Summability of random variables

Let $(X_i, i \in \mathbb{Z})$ be an univariate stochastic process. We say that the sum $\sum_{i=1}^{\infty} X_i$ is summable (in probability, resp. almost surely), if $S_n = \sum_{i=1}^{n} X_i$ converges to some random variable (in probability , resp. a.s.).

Theorem 1.1. *If $(X_i, i \in \mathbb{Z})$ are independent random variables, then the sum $\sum_{i=1}^{\infty} X_i$ is summable in probability if and only if it is summable almost surely.*

Theorem 1.2. *Let α_i be a sequence of real constants such that $\sum_{i=1}^{\infty} |\alpha_i| < \infty$. If $(\varepsilon_i, i \in \mathbb{Z})$ are iid random variables with finite variance, then $\sum_{i=1}^{\infty} \alpha_i \varepsilon_i$ is a.s. summable.*

1.1.3 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 such that* $\sum_{i=0}^{\infty} |\alpha_i| < \infty$ *. 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.

Definition 1.2. *Let* $\alpha_1, \dots, \alpha_q$ *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* $AR(q)$*.*

Theorem 1.3. *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* $\frac{1}{f(x)} = \sum_{i=0}^{\infty} \beta_i x^i$*.*

1.1.4 VAR(q) models

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

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

$$Z_t = A_1 Z_{t-1} + \cdots + A_q Z_{t-q} + \varepsilon_t, \tag{1.1}$$

is called vector autoregressive model of order q*, notation VAR(q).*

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 an $VAR(q+h)$ process. We will refer to the maximal q such that $A_q \neq \mathbf{0}$ as the *minimal order*.

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

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

where I_d *denotes unit d-dimensional matrix.*

Theorem 1.4 (Causal representation)**.** *If 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*

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

Theorem 1.5. *If a $VAR(q)$ process is stable, then it is also stationary. The other direction is not true in general.*

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

1.2 Extremal value theory

If the reader is not familiar with the concept of heavy-tailed distributions, we recommend Foss et al. [2009], 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 1.5. *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*

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

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

Lemma 1.1. *For every regularly varying function, there exist slowly varying function L such that $f(x) = x^{\theta} L(x)$.*

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

Definition 1.6. *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 1.7. *Random variable X satisfies 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. Typical examples of heavy-tailed distributions with positive tail index are Cauchy distribution, Pareto distributions or Burr distributions.

Lemma 1.2. *Suppose $X \sim RV(\theta)$. Then $E|X|^{\beta} < \infty$ for $\beta < \theta$, and $E|X|^{\beta} = \infty$ for $\beta > \theta$.*

Theorem 1.6 (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. The previous theorem 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 following result is also a particular case of so-called Breiman's lemma.

Consequence 1. *For $X \sim RV(\theta)$ and $\alpha > 0$ holds $P(\alpha X > u) \sim \alpha^\theta P(X > u)$.*

1.2.1 Causal linear process with heavy-tailed noise

Proof of the theorems can be found in Resnick [1987], Lemma 4.24.

Theorem 1.7 (Sufficient condition)**.** *Let $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$. Let α_i be a sequence of non-negative constants, such that one of the following conditions hold:*

1. *For $\theta > 1$ is $E(\varepsilon_i) = 0$ and $\sum_{i=0}^{\infty} \alpha_i < \infty$*
2. *For $\theta \in (0,1]$ $\exists \delta > 0$ 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. *$P(\sum_{i=0}^{\infty} \alpha_i \varepsilon_i > u) \sim [\sum_{i=0}^{\infty} \alpha_i^\theta] P(\varepsilon_1 > u)$.*

Theorem 1.8 (Necessary condition)**.** *Let $(\varepsilon_i, i \in \mathbb{Z}) \overset{iid}{\sim} RV(\theta)$. 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.

Remark. If we assume that we have AR(q) process with heavy-tailed noise, which is stable ($\implies$ stationary), we will automatically assume that the necessary condition from the Theorem 1.8 is fulfilled.

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].

1.3.1 Structural causal model

Let X, Y be real random variables.

Definition 1.8. *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. In this case, we say that X causes Y, or that X is the cause and Y is the effect.

Lemma 1.3. *For every joint distribution $P_{X,Y}$, there exists an SCM, where $Y = f_Y(X, N_Y)$ where f is some measurable function and X is independent of N_Y.*

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 1.9. *Linear models with non-gaussian additive noise (LiNGAMs) are a specific case 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 1.4. *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.*

Definition 1.10. *Nonlinear additive noise models (ANMs) are a 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

There is a large number of different notions of causality. Generally, process X causes Y if the knowledge of X can improve the prediction of Y.

Definition 1.11. *Let $(X, Y)^\top = ((X_t, Y_t)^\top, t \in \mathbb{Z})$ be bivariate stationary time series. Let $\sigma_t^Y = \sigma\{(Y_s, s \leq t)\}$ and $\sigma_t^{X,Y} = \sigma\{(X_s, Y_s, s \leq t)\}$. X (Granger) causes Y, if and only if there exists a measurable set A such that*

$$P(Y_{t+1} \in A \mid \sigma_t^Y) \neq P(Y_{t+1} \in A \mid \sigma_t^{X,Y}).$$

In case of VAR processes, this definition is equivalent to the following definition (sometimes called Sims causality), which will be used in the thesis.

Definition 1.12. *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 exist* $\delta_i \neq 0$.

In the notion of Theorem 1.4, 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 some $d_i \neq 0$.

Definition 1.13. *Let* $(X,Y)^\top = ((X_t,Y_t)^\top, t \in \mathbb{Z})$ *follow stationary* $GAM(1)$ *model (defined in 1.15), 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 so called Granger test. For $VAR(q)$ model as in 1.12, it uses linear regression to test the hypothesis if the sub-model $Y_t = \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q}$ 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 not significantly larger).

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

1.4 Generalization of VAR model

We will define generalizations of the VAR model. They can be seen as a counterpart of additive noise models and structural causal models for time series. We did not find any similar context in the literature.

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

Definition 1.14. *Let* $F : \mathbb{R}^{(q+1)\times d} \to \mathbb{R}^d$ *be measurable function. Stochastic process* $(Z_t, t \in \mathbb{Z})$*, defined by*

$$Z_t = F(Z_{t-1}, \ldots, Z_{t-q}, \varepsilon_t),$$

is called functional vector autoregressive model of order q*, notation* $FVAR(q)$.

Definition 1.15. *Let $F^k = (f^k_{i,j})^d_{i,j=1}$ be a $d \times d$ matrices whose elements are real measurable functions, for $k \le q$. For vector $v \in \mathbb{R}^d$, we define a multiplication $F^k(v) \in \mathbb{R}^d$, where the elements are $F^k(v)_i = \sum_{j=1}^d f^k_{i,j}(v_j)$. 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 generalized additive model of order q, notation $GAM(q)$.

Remark. For $d = 2$ and $Z = (X, Y)^\top$, the $GAM(1)$ has the form

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

for some measurable functions f_1, f_2, g_1, g_2.

Lemma 1.5. *Let $(X_t, t \in \mathbb{Z})$ be stationary univariate process. Let it satisfy*

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

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

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

Example. For

$$\begin{aligned} X_t &= \frac{1}{2} X_{t-1} + \varepsilon^X_t \\ Y_t &= \frac{1}{2} Y_{t-1} + \sqrt{X_{t-5}} + \varepsilon^Y_t, \end{aligned}$$

where $\varepsilon_t^\cdot$ are iid non-negative, $(X, Y)^\top$ is $GAM(5)$ process.

2. Causal tail coefficient for time series

Let $(X,Y)^\top = ((X_t,Y_t)^\top, t \in \mathbb{Z})$ be 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.

Definition 2.1. *Let $(X,Y)^\top = ((X_t,Y_t)^\top, t \in \mathbb{Z})$ be bivariate (strong) 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_1, Y_1, respectively. $\Gamma^{time}_{X,Y}(q;-0)$ will denote $\Gamma^{time}_{X,Y}(q)$ without the zero term $F_Y(Y_0)$.

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

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_1 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 relation.

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

2.1 Models

Definition 2.2 (Model 1)**.** *Let $(X,Y)^\top$ follow stable $VAR(q)$ model, specified by*

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

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: we assume that $\alpha_i, \beta_i, \gamma_i, \delta_i \geq 0$ (this assumption is not necessary and will be discussed in Chapter 3). Moreover, we assume that $\varepsilon^X_t, \varepsilon^Y_t \overset{iid}{\sim} RV(\theta)$, and that $\sum_{i=0}^\infty a_i^\theta < \infty, \sum_{i=0}^\infty b_i^\theta < \infty, \sum_{i=0}^\infty c_i^\theta < \infty, \sum_{i=0}^\infty d_i^\theta < \infty$. Then, we will say that $(X,Y)^\top$ follow Model 1.

Definition 2.3 (Model 2). *Let $(X,Y)^\top$ follow stationary GAM(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 $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} \frac{f(x)}{x} < 1$.

Moreover, let $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$ be non-negative. Then, we will say that $(X,Y)^\top$ follow Model 2.

2.2 Causal direction

Theorem 2.1. *Let $(X,Y)^\top$ be times series which follow either Model 1 or Model 2. If X causes Y, then $\Gamma^{time}_{X,Y}(q) = 1$.*

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

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

Theorem 2.2. *Let $(X,Y)^\top$ be times series which follow either Model 1 or Model 2. If Y is not causing X then $\Gamma^{time}_{Y,X}(p) < 1$ 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. 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* which has 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

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

(see also Section 3.4 for discussion). Let us give some vague computation of this coefficient. From Theorem 2.1 is $\Gamma^{time}_{X,Y} = 1$. 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 independence, second from Lemma 5.2):

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

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 datasets, 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 Non-direct dependences

Up to now, we assumed that large X causes large Y. In other words, we assumed that all coefficients in our Model 1 are non-negative. In this section, we will discuss the extension to possibly negative coefficients and non-direct proportional dependencies.

3.1.1 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. 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 $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} \textit{Cauchy}$. 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. 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

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

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

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

We call that it satisfy an extremal causal condition, if there exist $p \le q$ such that the following implication hold:

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

Lemma 3.1. *Extremal causal condition holds in Model 1, where X causes Y.*

Proof. In the notion of the definition of Model 1 and the previous theorem, if $\delta_p > 0$, then

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

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

Remark. 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 times series which follow Model 1, 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 $\Gamma^{time}_{|Y|,|X|}(q) < 1$.*

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

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

Such a coefficient does not have good properties and is still not applicable for a general class of VAR coefficients.

3.2 Derivatives

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

$$X_t^* := X_t - X_{t-1}$$
$$Y_t^* := Y_t - Y_{t-1}.$$

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

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

$$X_t = 0.999X_{t-1} + \varepsilon_t^X$$
$$Y_t = 0.5X_{t-1} + \varepsilon_t^Y.$$

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 the example, 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.

3.3 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 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 previous result holds even if we do not observe the common cause. However, the common cause still needs to fulfil 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. Let $(X,Y,Z)^\top$ follow 3 dimensional stable $VAR(q)$ model, specified by

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

where $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} Pareto(2,2)$ (i.e. with tail index 2) and $\varepsilon_t^Z \overset{\text{iid}}{\sim} Pareto(1,1)$ (i.e. with tail index 1). Then, $\Gamma_{X,Y}^{time}(1) = \Gamma_{Y,X}^{time}(1) = 1$ even though X does not cause Y. This claim follows from Proposition 5.4, where we used that the linear causal representation is in the form $X_0 = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{-i}^X + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{-i}^Z$ and $Y_1 = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{1-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{1-i}^Z$.

3.4 Estimating lag q

3.4.1 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 $\Gamma^{time}_{Y,X}(q)$ 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.

3.4.2 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

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

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.

3.4.3 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?

Definition 3.2 (Lag). *Let $(X, Y)^\top$ follow 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 Model 1, 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 $r \geq p$.*

Notation. *For simplicity, we denote $\Gamma^{time}_{X,Y}(-q) := \Gamma^{time}_{Y,X}(q)$ and $\hat{\Gamma}^{time}_{X,Y}(-q) := \hat{\Gamma}^{time}_{Y,X}(q)$ for any positive q.*

Definition 3.3. *Let $\Psi := \{q \in \mathbb{Z} : \Gamma^{time}_{X,Y}(q) = 1\}$. Let us have consistent tests of the hypothesis $\Gamma^{time}_{X,Y}(q) = 1$, and $\hat{\Psi} := \{q \in \mathbb{Z}$: We do not reject the hypothesis $\Gamma^{time}_{X,Y}(q) = 1\}$. We define*

$$\hat{q} := \min_{q \in \hat{\Psi}} |q|.$$

In case that $\hat{\Psi} = \emptyset$, we refer to a minimum of an empty set as ∞.

Theorem 3.3. *Let $(X,Y)^\top$ follow Model 1. Then, $\hat{q}$ is a consistent estimator of the minimal lag.*

Proof. Assume that X, Y are not independent, and let p be the true minimal lag. WLOG let X cause Y. It follows from Lemma 3.2, that for large n, with large probability we will reject the hypothesis $\Gamma^{time}_{X,Y}(p-h) = 1, \forall h \in \{1, \dots, 2p-1\}$ and we will not reject the hypothesis $\Gamma^{time}_{X,Y}(p) = 1$. Therefore, with large probability is $\hat{q} = p$, what we wanted to show.

In case that X, Y are independent, $\Gamma^{time}_{X,Y}(q) < 1$ for every q, and therefore $\hat{q} \to \infty$ as $n \to \infty$. $\square$

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. In economy is often used so-called *conditional tail expectation*, which is only a slight modification of the causal tail coefficient 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 future research to show some connections between an extremogram and causal tail coefficient.

Another interesting approach for causal detection can be using an extremal index 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. E.g. 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.

Neviem nakolko sa mi paci takato diskusia.

4. Estimations and simulations

Chapter not finished

All methods proposed in this chapter are programmed in R language, and can be found on github webpage https://github.com/jurobodik/Master_thesis

4.1 Non-parametric estimator

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

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

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 of only those values of Y_i where X_i is larger than some threshold. For simplicity, assume that we observe $n+q$ data.

Definition 4.1.

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

where $\tau^X_k = X_{(n-k+1)}$ is the $k-$ th largest value of X_i, and $\hat{F}_X(X_i) = \frac{1}{n}\sum_{j=1}^n 1\{X_j \leq X_i\}$.

Number k represents a 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

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

Theorem 4.1. *Let $(X,Y)^\top$ be a strong stationary bivariate time series, for which $\Gamma^{time}_{X,Y}(q)$ exists. Let the empirical distribution functions $\hat{F}_X(x),\hat{F}_Y(x)$ be uniformly consistent. If k_n fulfill (4.1), then $\hat{\Gamma}^{time}_{X,Y}(q)$ is asymptotically unbiased estimator of $\Gamma^{time}_{X,Y}(q)$.*

Consequence 2. *Let $(X,Y)^\top$ follow model 1, where X causes Y. If k_n fulfill (4.1), then $\hat{\Gamma}^{time}_{X,Y}(q) \xrightarrow{P} \Gamma^{time}_{X,Y}(q)$ as $n\to\infty$, i.e. is consistent estimator.*

Proof. Consequence of the fact that for VAR sequences is the empirical distribution function uniformly consistent, previous theorem and Theorem 2.1. Because X causes Y, according to Theorem 2.1 is $\Gamma^{time}_{X,Y} = 1$. From the previous theorem therefore holds $\mathbb{E}\,\hat{\Gamma}^{time}_{X,Y}(q) \to 1$. Necessary $var(\hat{\Gamma}^{time}_{X,Y}(q)) \to 0$ because $\hat{\Gamma}^{time}_{X,Y}(q) \leq 1$ a.s. Therefore, the consistency holds. □

Remark. Unfortunately, the estimator is not consistent in general. A counterexamples include (non-trivial) self-similar time series with $\Gamma^{time}_{X,Y} < 1$. A vague idea why is this true is the following: if our time series are self-similar, then for large n the estimate will be computed only from one largest "jump". But if $\Gamma^{time}_{X,Y} < 1$, then the estimate can not be consistent if it is computed only from one "jump".

Treba to viac popisat? alebo menej?

In the following methods we take $k_n = \sqrt{n}$. Note that this choice is not optimal- it should (ideally) depend on the structure of the time series.

4.1.1 Testing and bootstrap

We want to test the hypothesis $\Gamma^{time}_{X,Y}(q) = 1$ against the alternative $\Gamma^{time}_{X,Y}(q) < 1$. From the simulations, it seems that the statistic $\hat{\Gamma}^{time}_{X,Y}(q)$ asymptotically follows a normal distribution, but proving it is beyond the scope of this thesis. One possibility is to put a threshold, e.g. we say that we do not reject the hypothesis 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.

Another method is to use block bootstrap technique (sometimes called stationary bootstrap). The idea is to resample the blocks of data with replacement, to obtain new datasets from which we obtain new estimate $\hat{\Gamma}^{time}_{X^\star,Y^\star}(q)$. We can obtain a p-value with the following algorithm:

1. Let $m \in \mathbb{N}$ divide n (m will be number of blocks) and let $K >> 0$ be large (K will be number of repetitions).

2. For $i \in \{0, \ldots, m-1\}$, choose uniformly randomly a number ω_i between 1 and $n - m$. Put $X^\star_{im+j} = X_{\omega_i+j}$ and $Y^\star_{im+j} = Y_{\omega_i+j}$ for all $j \in \{1, \ldots, m\}$.

3. Compute $2\hat{\Gamma}^{time}_{X,Y}(q) - \hat{\Gamma}^{time}_{X^\star,Y^\star}(q)$.

4. Repeat step 2 and step 3 exactly K times . Compute 95% quantile of those values. We will call it the bootstrap confidence interval. Compute number of those values that exceed 1. We will call this number divided by K the bootstrap p-value.

It is well known that confidence intervals obtained by bootstrap are in general smaller than the correct ones, therefore such tests are anti-conservative.

4.2 Simulation results

We will simulate how the estimations of causal tail coefficient for time series works for a series of models.

4.2.1 Performance under non-heavy tailed noise

We will deal with the model, where $(X, Y)^\top$ follow

$$X_t = 0.5X_{t-1} + \varepsilon_t^X \tag{4.2}$$
$$Y_t = 0.5Y_{t-1} + \delta X_{t-2} + \varepsilon_t^Y. \tag{4.3}$$

We will show results for $\delta = 0.1, 0.5$ and 0.9. Distributions of $\varepsilon_t^X, \varepsilon_t^Y \overset{\text{iid}}{\sim} \mathfrak{L}$ where $\mathfrak{L}$ will be standard Pareto and standard normal distributions. Table 4.1 shows results for number of data $n = 100, 1000, 10000$. In each cell is estimated $\hat{\Gamma}_{X,Y}^{time} := \hat{\Gamma}_{X,Y}^{time}(2)$, computed as a mean of 200 simulations from the model with corresponding δ, distribution and number of data-points.

Example. $\hat{\Gamma}_{X,Y}^{time} = 0.5 \pm 0.1$ means that out of all 200 simulated series 4.2, $\hat{\Gamma}_{X,Y}^{time}(2)$ was on average equal to 0.5 and exactly 190 of those simulations were $\hat{\Gamma}_{X,Y}^{time} \leq 0.6$. We write $\pm$, because the 5% quantiles were in all cases symmetrical, i.e. cca 190 of those simulations fulfilled also $\hat{\Gamma}_{X,Y}^{time} \geq 0.4$.

Results of the simulations 1. *The method works surprisingly well under violating the heavy-tails assumption. If we consider Gaussian noise, the correct theoretical value is $\Gamma_{X,Y}^{time} = \Gamma_{Y,X}^{time} = 1$. But if we estimate these values, for finite $u < 1$ is $\mathbb{E}[\max\{F_Y(Y_0), \ldots, F_Y(Y_q)\} \mid F_X(X_0) > u]$ still larger than in the other direction. This results into seemingly correct causal directions. On the other hand, if the cause has heavier tails than the effect, our method suggests $\Gamma_{X,Y}^{time} = \Gamma_{Y,X}^{time} = 1$. In this case, for large n, both estimates are very close to 1 and this results into the wrong causal directions.*

4.2.2 Bootstrap performance and comparison with Granger tests

We perform a simple simulation study to show how well the bootstrap works. We will consider two different time series models, one the simplest VAR model and the second with (unobserved) common cause and non-linear dependencies.

Definition 4.2. *Time series 1 will follow*

$$X_t = 0.5X_{t-1} + \varepsilon_t^X$$
$$Y_t = 0.5Y_{t-1} + 0.5X_{t-3} + \varepsilon_t^Y,$$

where distributions of $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} \mathfrak{L}$ where $\mathfrak{L}$ will have Student t-distribution with 2 *degrees of freedom.*

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: We have 200 simulated time series following VAR model 4.2. Each cell represents different coefficient δ and a different number of data-points n and a different noise-distribution. Each value $\hat{\Gamma}^{time}_{X,Y} = \cdot \pm \cdot$ represents 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, the true fact is that X causes Y.

Definition 4.3. *Time series 2 will follow*

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

where distributions of $\varepsilon_t^X, \varepsilon_t^Y, , \varepsilon_t^Z \overset{iid}{\sim} \mathfrak{L}$ *where* $\mathfrak{L}$ *will have standard Pareto distribution. In this case, process* Z *represents (not observed) common cause. Sample realization can be found in Figure 4.1.*

We will test two different hypothesis: first is that X causes Y, the second is that Y does not cause X. We compute two accuracy indexes, measuring the performance of different tests:

- Sensitivity, i.e. $\frac{TP}{TP+FN}$ where TP is true positive and FN is false negative. In our case where X causes Y, sensitivity is the estimated probability of correctly inferring that X causes Y.
- Specificity, i.e. $\frac{TN}{TN+FP}$ where TN is true negative and FP is false positive. In our case, where X causes Y, specificity is the estimated probability of correctly inferring that Y is not causing X.

We will discuss three different tests for detecting the causal directions. First is the bootstrap method discussed in section 4.1.1. Second is estimating $\hat{\Gamma}^{time}_{X,Y}(3)$ and concluding that X causes Y if and only if $\hat{\Gamma}^{time}_{X,Y}(3) \geq \tau$ with the choice $\tau = 0.9$. The third method is using classical Granger test. [1] All results use the significance level $\alpha = 0.05$.

We perform 100 simulation of the aforementioned time series, with number of data $n = 500, 5000$. For each of them, we estimate the causal direction using these three methods. Finally, we compute the percentage of correctly inferring the causal directions $X \to Y$ and $Y \not\to X$. Table 4.2 shows the results in percentage.

Results of the simulations 2. *Even for moderate number of data such as* 5000*, bootstrap method does not give satisfactory results. In more than* 25% *of cases we incorrectly rejected the hypothesis* $\Gamma^{time}_{X,Y} = 1$*. This result corresponds to the property of bootstrap which states that the estimated confidence intervals are shorter. Therefore, tests more often reject the hypothesis than they should.*

Results of the simulations 3. *Granger tests behave well under simple heavy tailed VAR model. This is not surprising, and similar observations were discussed in Mittnik et al. [2001]. Granger test does not behave properly when a common cause is present. In our artificial time series, the common cause creates a spurious causal direction from* $Y \to X$*. Granger test can not deal with such complex systems.*

Surely the best results are obtained by the method which uses $\hat{\Gamma}^{time} > \tau = 0.9$. At in cases when we have enough data. Even for such complex non-linear series with common cause, we obtained always a correct result. On the other hand, this method does not provide any rigorous testing or p-value.

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

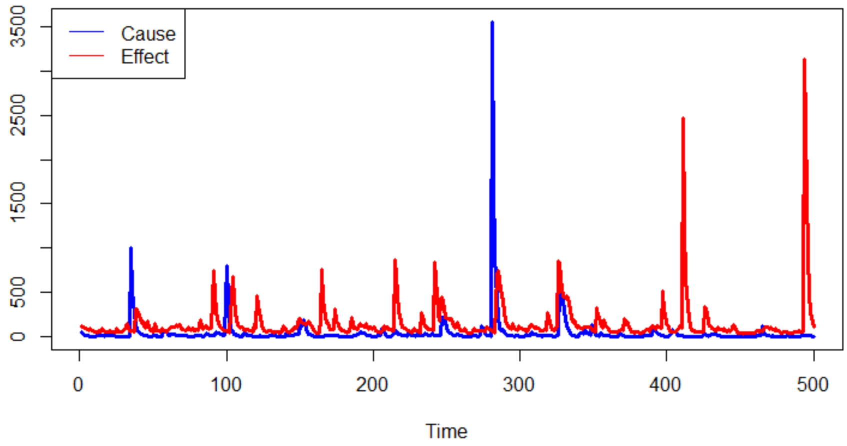

Figure 4.1: A sample realization of the time series from the Definition 4.3.

4.3 Application

We will show in the application how the method works for real data. Our data contains 6598 data, representing the amount of rainfall in india and the NAO index TODO

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

Table 4.2: We consider two time series models, one is simple VAR model, the second is more complex non-linear model with common cause. We performed three tests concerning the causal direction. 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 second and third method, the results represents the percentage of cases when the corresponding p-value was less than $\alpha = 0.05$. For the first method, sensitivity 99% represents that in 99% of cases was $\hat{\Gamma}_{X,Y} > \tau$. Specificity 79% represents that in 79% of cases was $\hat{\Gamma}_{X,Y} \leq \tau$.

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 it holds that*

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

or more generally,

$$P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1 \mid \sum_{i=1}^{\infty} b_i \varepsilon_i > M_2) \geq P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1), \tag{5.1}$$

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

Proof. First, we will prove that $\forall n \in \mathbb{N}$ holds

$$P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1 \mid \sum_{i=1}^{n} b_i \varepsilon_i > M_2) \geq P(\sum_{i=1}^{n} a_i \varepsilon_i > M_1). \tag{5.2}$$

We will use the following fact: Let $\varepsilon = (\varepsilon_1, \dots \varepsilon_n)$, 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 $f(x_1, \dots, x_n) = 1\{\sum_{i=1}^{n} a_i x_i > M_1\}, g(x_1, \dots, x_n) = 1\{\sum_{i=1}^{n} b_i x_i > M_2\}$. They are non-decreasing because $a_i, b_i \geq 0$. Therefore, we obtain

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

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. Let $\delta > 0$, and choose $n_0 \in \mathbb{N}$ such that $P(\sum_{i=n_0}^{\infty} |a_i \varepsilon_i| + |b_i \varepsilon_i| > \delta) < \delta$. Then we can use that

$$\begin{aligned}
&P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1) \\
&= P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1, \sum_{i=n_0+1}^{\infty} a_i \varepsilon_i < \delta) + P(\sum_{i=1}^{\infty} a_i \varepsilon_i > M_1, \sum_{i=n_0+1}^{\infty} a_i \varepsilon_i \geq \delta) \\
&\geq P(\sum_{i=1}^{n_0} a_i \varepsilon_i > M_1 + \delta)
\end{aligned}$$

and also $P(\sum_{i=1}^{\infty} a_i\varepsilon_i > M_1) \leq P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 - \delta) + \delta$. Similarly for $\sum b_i\varepsilon_i$. Finally, we can write

$$\frac{P(\sum_{i=1}^{\infty} a_i\varepsilon_i > M_1 \mid \sum_{i=1}^{\infty} b_i\varepsilon_i > M_2)}{P(\sum_{i=1}^{\infty} a_i\varepsilon_i > M_1)} = \frac{P(\sum_{i=1}^{\infty} a_i\varepsilon_i > M_1; \sum_{i=1}^{\infty} b_i\varepsilon_i > M_2)}{P(\sum_{i=1}^{\infty} b_i\varepsilon_i > M_2)P(\sum_{i=1}^{\infty} a_i\varepsilon_i > M_1)}$$
$$\geq \frac{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 + \delta; \sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 + \delta)}{(P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 - \delta) + \delta)(P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 - \delta) + \delta)}$$
$$= \frac{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 + \delta; \sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 + \delta)}{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 + \delta)P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 + \delta)}$$
$$\cdot \frac{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 + \delta)P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 + \delta)}{(P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 - \delta) + \delta)(P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 - \delta) + \delta)}$$
$$\geq 1 \cdot \frac{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 + \delta)}{P(\sum_{i=1}^{n_0} a_i\varepsilon_i > M_1 - \delta) + \delta} \cdot \frac{P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 + \delta)}{(P(\sum_{i=1}^{n_0} b_i\varepsilon_i > M_2 - \delta) + \delta} \xrightarrow{\delta\to 0} 1,$$

where we used the finite case result and that our variables are continuous.

□

5.2 Proposition 5.2

Proposition 5.2.
- *Let $(\varepsilon_i^X, \varepsilon_i^Y, i \in \mathbb{Z})$ be independent $RV(\theta)$ continuous random variables with compatible tails.*
- *Let $a_i, b_i, c_i \geq 0$ be constants, such that all $\sum_{i=0}^{\infty} a_i\varepsilon_i^Y, \sum_{i=0}^{\infty} b_i\varepsilon_i^X, \sum_{i=0}^{\infty} c_i\varepsilon_i^X$ are a.s. summable and non-trivial.*
- *Denote $A = \sum_{i=0}^{\infty} a_i^\theta, B = \sum_{i=0}^{\infty} b_i^\theta, C = \sum_{i=0}^{\infty} c_i^\theta$ (it holds that $A, B, C \in (0,\infty)$, see Theorem 1.8).*
- *Let $\Phi = \{i \in \mathbb{N}^0 : b_i > 0 = a_i\}$.*

Then, $\forall M > 0$ such that $P(\sum_{i=0}^{\infty} a_i\varepsilon_i^X < M) > 0$ is

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

We will prove this theorem using the following series of lemmas.

Lemma 5.1. *Let $X, Y \sim RV(\theta)$ be independent. Then, $\forall M > 0$ is*

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

Lemma 5.2. *Under the conditions from Proposition 5.2, holds $\forall n \in \mathbb{N}$ that*

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

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 holds $\forall n \in \mathbb{N}$ that*

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

Proof of Proposition 5.2. Let $\delta > 0$ and choose n_0 such that $P(|\sum_{i=n_0}^{\infty} b_i\varepsilon_i^X| + |\sum_{i=n_0}^{\infty} a_i\varepsilon_i^X| > \delta) < \delta$.

Denote

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

Then, all E, F, G, H, Z are independent. With our notation, we *want* to prove that

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

It is enough, because due to Theorem 1.8, we have

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

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

First, due to Lemma 5.1 is $\lim_{u\to\infty} P(H > \delta \mid H + (G + Z) > u) = P(H > \delta) \lim_{u\to\infty} \frac{P(H>u)}{P(H+G+Z>u)} < \delta$.

Second, we will show that $\lim_{u\to\infty} P(F > \delta \mid G + H + Z > u) < \delta$. Using previous results and independence of F and G, Z we obtain

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

Finally, using Lemma 5.3, we obtain

$$
\begin{aligned}
&\lim_{u\to\infty} P(E+F<M \mid G+H+Z>u)\\
&\geq \lim_{u\to\infty} P(E+2\delta<M \mid G+(H+Z)>u)\\
&= P(E+2\delta<M)\lim_{u\to\infty}\frac{P(H+Z>u)}{P(H+Z>u)+P(G>u)}\\
&\geq P(E+2\delta<M)\lim_{u\to\infty}\frac{(1+\delta)P(Z>u)}{P(Z>u)+P(G>u)}\\
&\geq P(E+F<M-4\delta)\lim_{u\to\infty}\frac{(1+\delta)P(Z>u)}{P(Z>u)+P(G+H>u)},
\end{aligned}
$$

and also

$$
\begin{aligned}
&P(E+F<M \mid G+H+Z>u)\\
&\leq P(E+F<M+4\delta)\lim_{u\to\infty}\frac{1}{(1-\delta)}\frac{P(Z>u)}{P(Z>u)+P(G+H>u)}.
\end{aligned}
$$

We send $\delta\to 0$ and both right sides of the equations converge to

$$
P(E+F<M)\lim_{u\to\infty}\frac{P(Z>u)}{P(G+H>u)+P(Z>u)},
$$

what we wanted to prove. □

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

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

If we denote truncated random variable $W := X1_{\{X<M\}}$, 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 N. H. Bingham [2006] to obtain $lim_{u\to\infty}\frac{P(Y+W>u)}{P(Y>u)} = 1$.

Finally, using max-sum equivalence, we get

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

what we wanted to prove. □

Proof of Lemma 5.2. WLOG $\Phi=\emptyset$, otherwise we have only lower n. 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:

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

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

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

□

Proof of Lemma 5.3. WLOG $\Phi = \emptyset$, otherwise we have only lower n. In this proof only, we will denote $B = \sum_{i=0}^n b_i$, and $A = \sum_{i=0}^n a_i$. Denote $W := \sum_{i=0}^n b_i\varepsilon_i^X 1_{\{\sum_{i=0}^n a_i\varepsilon_i^X < M\}}$. 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. 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. □

Consequence 3. *Under the conditions from Proposition 5.2, $\forall M > 0$ such that $P(\sum_{i=0}^\infty a_i|\varepsilon_i^X| < M) > 0$ also holds*

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

Proof. Proof can be done analogously as proof of the previous proposition. 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

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

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

$\square$

5.3 Proposition 5.3

Proposition 5.3. *Let $(X,Y)^\top$ follow $GAM(q)$ 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 satisfy $\lim_{x\to\infty} f(x) = \infty$ and $\lim_{x\to\infty} \frac{f(x)}{x} < 1$.Moreover, let $\varepsilon, \varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} RV(\theta)$. 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, it holds that*

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

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

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

Therefore, we have $\lim_{u\to\infty} \frac{P(X_t > u)}{P(\varepsilon > u)} = \frac{1}{1-c} < \infty$. $\square$

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 $f(x+y) \le (K+cx) + (K+cy)$. Then, a.s. holds

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

Finally,

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

where we used independence of X_i and ε_i^Y, the previous Lemma and Theorem 1.8.

□

Remark. We proved a stronger claim. We showed that for every Model 2, there exist 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.

5.4 Proposition 5.4

Proposition 5.4. *Let $\varepsilon_t^X, \varepsilon_t^Y \overset{iid}{\sim} Pareto(2,2)$ (i.e. with tail index 2) and $\varepsilon_t^Z \overset{iid}{\sim} Pareto(1,1)$ (i.e. with tail index 1) be jointly independent, $t \in \mathbb{Z}$. Then,*

$$\lim_{u\to\infty} P(\sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{1-i}^Y + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{1-i}^Z > \frac{u}{4} \mid \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{-i}^X + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{-i}^Z > u) = 1.$$

Proof. Let $X = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{-i}^X$, and $Y = \sum_{i=0}^{\infty} \frac{1}{2^i}\varepsilon_{1-i}^Y + \frac{\varepsilon_0^Z}{2}$. Then, we want to show that

$$\lim_{u\to\infty} \frac{P(Y + \sum_{i=0}^{\infty} \frac{i+1}{2^{i+1}}\varepsilon_{-i}^Z > \frac{u}{4}; X + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{-i}^Z > u)}{P(X + \sum_{i=0}^{\infty} \frac{i}{2^i}\varepsilon_{-i}^Z > u)} = 1.$$

Because X has lower tail index then ε_i^Z, from extremal value theory follows that

$$P(X + \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > u) \sim P(\sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > u).$$

Let us focus on the numerator of the fraction. Choose $\delta > 0$, and we have

$$\begin{aligned}
&\lim_{u\to\infty} P(Y + \sum_{i=0}^{\infty} \frac{i+1}{2^{i+1}} \varepsilon_{-i}^Z > \frac{u}{4}; X + \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > u) \\
&\geq \lim_{u\to\infty} P(\sum_{i=0}^{\infty} \frac{i+1}{2^{i+1}} \varepsilon_{-i}^Z > (1+\delta)\frac{u}{4}; \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u; X > -\delta u; Y > -\delta u) \\
&= \lim_{u\to\infty} P(X > -\delta u) P(Y > -\delta u) P(\sum_{i=0}^{\infty} \frac{i+1}{2^i} \varepsilon_{-i}^Z > (1+\delta)\frac{u}{2}; \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u) \\
&\geq \lim_{u\to\infty} 1 \cdot 1 \cdot P(\sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Z > -(1+\delta)\frac{u}{2}; \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u),
\end{aligned}$$

where we used independence of X, Y, ε^Z, and in the last step we simply subtracted second term from the first term. Finally, we have

$$\begin{aligned}
&\lim_{u\to\infty} \frac{P(\sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Z > -(1+\delta)\frac{u}{2}; \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u)}{P(\sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > u)} \\
&= \lim_{u\to\infty} \frac{P(\sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Z > -(1+\delta)\frac{u}{2}; \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u)}{(1+\delta)^2 P(\sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u)} \\
&= \frac{1}{(1+\delta)^2} \lim_{u\to\infty} P(\sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Z > -(1+\delta)\frac{u}{2} \mid \sum_{i=0}^{\infty} \frac{i}{2^i} \varepsilon_{-i}^Z > (1+\delta)u) \\
&\geq \frac{1}{(1+\delta)^2} \lim_{u\to\infty} P(\sum_{i=0}^{\infty} \frac{1}{2^i} \varepsilon_{-i}^Z > -(1+\delta)\frac{u}{2}) = \frac{1}{(1+\delta)^2},
\end{aligned}$$

where we used Proposition 5.1 in the last step. Finally, by sending $\delta \to 0$, we obtain the desired equality.

□

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

Proof. Trivial. □

6.1 Theorem 2.1.

Model 1.

Theorem 2.1. (Model 1). Let $(X, Y)^\top$ be times series which follow Model 1. 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

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

Now, using Causal representation 1.4, we can rewrite all

$$X_0 = \sum_{i=0}^{\infty} a_i \varepsilon^X_{-i} + \sum_{i=0}^{\infty} c_i \varepsilon^Y_{-i}$$

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

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

We obtain

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

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.

$\square$

Model 2.

Theorem 2.1. (Model 2). Let $(X, Y)^\top$ be times series which follow Model 2. If X causes Y then $\Gamma^{time}_{X,Y}(q) = 1$.

Proof. We proceed very similarly as in the proof of model 1. 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

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

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, surely for all $u > x_0$ is

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

Finally, we only use the fact that ε^Y_t and g_1 are non-negative, therefore also

$$\lim_{u\to\infty} P(g_1(Y_{q-1}) + g_2(X_0) + \varepsilon^Y_t > M \mid X_0 > u) \geq \lim_{u\to\infty} P(g_2(X_0) > M \mid X_0 > u) = 1,$$

what we wanted to prove.

$\square$

6.2 Theorem 2.2.

Model 1.

Theorem 2.2. (Model 1). Let $(X, Y)^\top$ be times series which follow either Model 1 or Model 2. 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

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

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

We rewrite

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

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

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

for $\phi_i = a_i + \cdots + a_{i-p+1}$ (we define $a_j = 0$ for $j < 0$). Finally, it follows from the consequence of Theorem 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 requires non-trivial sums, but if $\forall i : d_i = 0$ then series are independent and this inequality holds trivially).

□

Model 2.

Theorem 2.2. (Model 2). Let $(X, Y)^\top$ be times series which follow Model 2. If Y is not causing X then $\Gamma_{Y,X}^{time}(q) < 1$.

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 $P(\varepsilon_0^X < M - \sup_{x<M} f(x)) > 0$. Denote $M^\star = \sup_{x<M} f(x)$. Similarly as in the proof of model 1, we will only work with

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

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

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

We only need to show that $\lim_{u\to\infty} P(X_0 > M \mid Y_0 > u) < 1$. Let $Z = g_1(Y_{-1}) + g_2(X_{-q})$, $Z \perp\!\!\!\perp \varepsilon_0^Y$. After rewriting, we obtain

$$\begin{aligned}
\lim_{u\to\infty} P(X_0 > M \mid Y_0 > u) &= \lim_{u\to\infty} P(X_0 > M \mid \varepsilon_0^Y + Z > u) \\
&= \lim_{u\to\infty} \frac{P(X_0 > M; \varepsilon_0^Y + Z > u)}{P(\varepsilon_0^Y + Z > u)}.
\end{aligned}$$

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

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

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

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

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 $Z \sim RV(\theta)$ (trivially from the definition, tails behave the same up to a constant). Therefore, we obtain

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

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.

□

6.3 Theorem 3.1.

Theorem 3.1. Let $(X, Y)^\top$ be times series which follow Model 1, 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 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, it holds that

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

see e.g. page 6 in Jessen and Mikosch [2006]. Second step follows simply from Theorem 1.8. Finally, we use this fact and the triangle inequality to obtain

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

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 $\delta_p \neq 0$. Using causal representation with the same notation as in the proof of Theorem 2.1., we obtain

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

where we used the same trick as in the first part of the proof. We are obtaining the classical non-negative case. The result follows from the previous theory- using Lemma 5.2 we obtain finite result

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

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

□

6.4 Theorem 3.2

Theorem 3.2. Let $(X, Y, Z)^\top$ follow 3 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:

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

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

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

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

□

6.5 Lemma 3.2

Lemma 3.2. Let $(X,Y)^\top$ follow Model 1, 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. 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 for some $M > 0$ is $\lim_{u\to\infty} P(Y_{p-1} < M|X_0 > u) > 0$.By rewriting to its causal representation, we obtain

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

We only need to realize that $d_i = 0$ for $i \in \{1,\dots,p-1\}$ because p is the minimal lag. Therefore, ε^X_0 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^X_0 is the second "sum"). □

6.6 Theorem 4.1

Theorem 4.1. Let $(X,Y)^\top$ be a strong stationary bivariate time series, for which $\Gamma^{time}_{X,Y}(q)$ exists. Let the empirical distribution functions $\hat{F}_X(x), \hat{F}_Y(x)$ be uniformly consistent. If k_n fulfill (4.1), then $\hat{\Gamma}^{time}_{X,Y}(q)$ is asymptotically unbiased estimator of $\Gamma^{time}_{X,Y}(q)$.

Proof. First, notice that

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

Second, notice that

$$\begin{aligned}
\mathbb{E}\,\hat{\Gamma}^{time}_{X,Y}(q) &= \mathbb{E}\,\frac{1}{k_n}\sum_{i:X_i\geq\tau^X_{k_n}} \max\{\hat{F}_Y(Y_i),\dots,\hat{F}_Y(Y_{i+q})\} \\
&= \mathbb{E}\,\frac{1}{n}\sum_{i=1}^{n}\frac{n}{k_n}\max\{\hat{F}_Y(Y_i),\dots,\hat{F}_Y(Y_{i+q})\}1[\hat{F}_X(X_i) > 1-\frac{k_n}{n}] \\
&= \frac{n}{k_n}\,\mathbb{E}\,[\hat{F}_Y(max\{Y_0,\dots,Y_q\})1[\hat{F}_X(X_0) > 1-\frac{k_n}{n}]] \\
&= \frac{n}{k_n}P(\hat{F}_X(X_0) > 1-\frac{k_n}{n})\,\mathbb{E}\,[\hat{F}_Y(max\{Y_0,\dots,Y_q\}) \mid \hat{F}_X(X_0) > 1-\frac{k_n}{n}] \\
&= \mathbb{E}\,[\hat{F}_Y(max\{Y_0,\dots,Y_q\}) \mid \hat{F}_X(X_0) > 1-\frac{k_n}{n}].
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\,[\hat{F}_Y(max\{Y_0,\ldots,Y_q\}) \mid \hat{F}_X(X_0) > 1 - \frac{k_n}{n}] \\
&= \mathbb{E}\,[F_Y(max\{Y_0,\ldots,Y_q\}) \mid \hat{F}_X(X_0) > 1 - \frac{k_n}{n}] \\
&+ \mathbb{E}\,[(\hat{F}_Y - F_Y)(max\{Y_0,\ldots,Y_q\}) \mid \hat{F}_X(X_0) > 1 - \frac{k_n}{n}].
\end{aligned}$$

The second term is less than $E[\sup_{x\in\mathbb{R}} |\hat{F}_Y(x) - F_Y(x)|] \to 0$ 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

$$\begin{aligned}
&\mathbb{E}\,[F_Y(max\{Y_0,\ldots,Y_q\}) \mid \hat{F}_X(X_0) > 1 - \frac{k_n}{n}] \\
&= \mathbb{E}\,[F_Y(max\{Y_0,\ldots,Y_q\}) \mid F_X(X_0) > F_X(X_{(n-k_n+1)})].
\end{aligned}$$

But because $F_X(X_{(n-k_n+1)}) \overset{n\to\infty}{\to} 1^-$ almost surely, we obtain the desired result. $\square$

Conclusion

TODO

7. Table of notations

$f(x) \sim g(x)$	$\triangleq$	$lim_{x\to\infty}\frac{f(x)}{g(x)} = 1$
$\varepsilon_i \overset{\text{iid}}{\sim} Cauchy$	$\triangleq$	random variables ε_i are independent and are identically distributed according to Cauchy distribution.
a.s.	$\triangleq$	almost surely
$\mathbb{N}^0$	$\triangleq$	set of all natural numbers with zero
F_X	$\triangleq$	distribution function of random variable X
EX , $\mathbb{E}\,X$	$\triangleq$	expected value of X
$Var(X)$	$\triangleq$	variance of X
$X \sim N(0,1)$	$\triangleq$	X has standard normal distribution
$X \perp\!\!\!\perp Y$	$\triangleq$	X is independent of Y
iid	$\triangleq$	independent and identically distributed
$X_{(k)}$	$\triangleq$	k-th order statistic
.	$\triangleq$	.

Table 7.1: Table of notations and basic definitions.

Bibliography

Richard A. Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 15(4):977–1009, Nov 2009. ISSN 1350-7265. doi: 10.3150/09-bej213. URL `http://dx.doi.org/10.3150/09-BEJ213`.

J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474, 1967. ISSN 00034851. URL `http://www.jstor.org/stable/2238962`.

Sergey Foss, Dmitry Korschunov, and Stan Zachary. An introduction to heavy-tailed and subexponential distributions, 2009.

Nicola Gnecco, Nicolai Meinshausen, Jonas Peters, and Sebastian Engelke. Causal discovery in heavy-tailed models, 2020.

C.W.J. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352, 1980. ISSN 0165-1889. doi: https://doi.org/10.1016/0165-1889(80)90069-X. URL `https://www.sciencedirect.com/science/article/pii/016518898090069X`.

A. Jessen and T. Mikosch. Regularly varying functions. *Publications De L'institut Mathematique*, 80:171–192, 2006.

Rafal Kulik and Philippe Soulier. *Heavy-Tailed Time Series*. 01 2020. ISBN 978-1-0716-0735-0. doi: 10.1007/978-1-0716-0737-4.

Helmut Lutkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Publishing Company, Incorporated, 2007. ISBN 3540262393.

Thomas Mikosch. Regular variation, subexponentiality and their applications in probability theory. *International Journal of Production Economics - INT J PROD ECON*, 01 1999.

S. Mittnik, V. Paulauskas, and S.T. Rachev. Statistical inference in regression with heavy-tailed integrated variables. *Mathematical and Computer Modelling*, 34(9):1145–1158, 2001. ISSN 0895-7177. doi: https://doi.org/10.1016/S0895-7177(01)00123-6. URL `https://www.sciencedirect.com/science/article/pii/S0895717701001236`.

Nicholas Moloney, Davide Faranda, and Yuzuru Sato. An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29:022101, 02 2019. doi: 10.1063/1.5079656.

Edward Omey N. H. Bingham, Charles M. Goldie. Regularly varying probability densities. *Publications de L'Institut mathematique*, 80(94), 2006.

Abdelhakim Necir, Rassoul Abdelaziz, and Ricardas Zitikis. Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, 2010, 04 2010. doi: 10.1155/2010/596839.

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

Milan Paluš, Katerina Hlaváčková-Schindler, Martin Vejmelka, and Joydeep Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1 – 46, 2007. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2006.12.004. URL http://www.sciencedirect.com/science/article/pii/S0370157307000403.

Jonas Peters, Dominik Janzing, and Bernhard Schlkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017. ISBN 0262037319.

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

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

Achim Zeileis and Torsten Hothorn. Diagnostic checking in regression relationships. *R News*, 2(3):7–10, 2002. URL https://CRAN.R-project.org/doc/Rnews/.