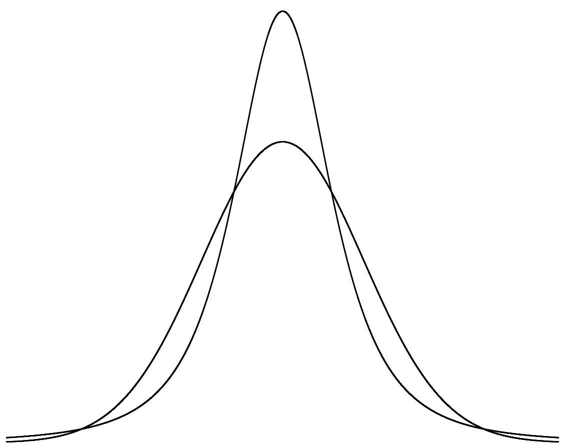
Tail dependence: OLS estimators under heavy-tails

Master Thesis Quantitative Finance

Jochem Antal Oorschot (374143) Supervisor: Chen Zhou

August 20, 2017



Abstract

This thesis shows that if the errors in a multiple regression model are heavy-tailed, the OLS estimators are tail dependent. The presence of heavy-tailed regression errors does not impair the F-test for joint tests on the regression coefficients. However, we show that it does change the dependence structure between the two main components of the F-statistic: the fitted sum of squares (FSS) and the residual sum of squares (RSS) become strongly tail dependent — which is in contrast to the fact that they are independent when the regression errors follow a normal distribution. The tail dependence between the OLS estimators, and that between the components of the F-statistic, follow from the fact that stochastic linear (and quadratic) combinations of heavy-tailed random variables are tail dependent.

Contents

1 Introduction

Financial markets are notorious for their crashing behaviour. During the most recent financial crisis of 2007-2008, many stock indices have experienced a daily loss of around 10% (Phani Kumar, 2008). Losses so large relative to the returns observed on most other days are hard to reconcile with a normal data generating process. For this reason, financial returns are often better modeled by heavy-tailed distributions. Heavy-tailed distributions possess tails that decay in power speed. In turn, extreme values are more likely to occur compared to thin-tailed distributions with exponentially decaying tails, such as the normal distribution. For example, Gopikrishnan et al. (1998) finds that the tails of the distributions of stock returns over short horizons can be described by a power law with a cubic tail exponent — which implies that these tails decay substantially slower than the tails of a normal distribution.

Since at least Mandelbrot's work on the stable Paretian hypothesis in the early 1960s, the finance literature has been extensively concerned with the heavy-tailed behaviour of financial data since it can affect the behaviour of statistical estimators. In the context of regression analysis it is necessary to address heavy-tails for at least two reasons. Firstly, it may be possible to find more efficient, or more robust, estimators than the ordinary least squares (OLS) estimator if the regression errors follow heavy-tailed distributions. Secondly, since the underlying data generating process is always unknown, one may run an OLS regression without knowing that the regression errors are heavy-tailed. It is therefore of interest to study how the OLS estimator behaves under heavy-tailed regression errors. This thesis serves the latter purpose by considering the joint tail distribution of multiple OLS estimators in a multiple regression with heavy-tailed errors, instead of focusing on deficiencies of the OLS estimator or attempting to overcome them.

Recently, Mikosch & de Vries (2013) show that when running a simple OLS regression, the estimator of the regression coefficient inherits the heavy-tailed behaviour of the error terms. Whereas Mikosch & de Vries (2013) analyses the marginal behaviour of a single OLS estimator, as a follow-up study, this thesis studies the joint behaviour of multiple OLS estimators — in particular, the tail dependence between the OLS estimators. In addition, we will examine the impact of having heavy-tailed regression errors on the F-statistic for jointly testing the regression coefficients.

The results in this thesis arise from the following property of heavy-tailed random variables: stochastic linear (and quadratic) combinations of heavy-tailed random variables exhibit tail dependence. Consequently, since the OLS estimators are stochastic linear combinations of the error terms we have that they are tail dependent under heavy-tailed errors. This is true for any arbitrary finite number of observations but is especially relevant in small samples. Furthermore, although it is known that the F-test is still valid under certain heavy-tailed distributions, we show that the numerator, the fitted sum of squares (FSS), and the denominator, the residual sum of squares (RSS), of the F-statistic are tail dependent if the regression errors are heavy-tailed. Tail dependence between the FSS and RSS follows from the fact that both are stochastic quadratic combinations of the heavy-tailed regression errors. The tail dependence between the FSS and RSS is in stark contrast with a regression model with normally distributed errors, under which the FSS and RSS are independent (Heij et al., 2004, p. 165).

This study contributes to the literature of regression analysis with heavy-tailed data. There are two streams of literature in this direction: finding more robust, or more efficient, estimators than the OLS estimator, and describing the behaviour of the OLS estimator when the regression errors are heavy-tailed. We will start with reviewing the stream of literature aiming at the former objective.

OLS regression analysis requires the existence of the variance of the regression errors. Otherwise, the OLS estimator minimizes a sum whose expectation is infinite, which in turn causes the OLS estimator to lose some of its favorable properties (see Fama (1963) for a discussion). This can be of practical relevance: for example, Mandelbrot (1963) argues that changes in the price of cotton should be described by a stable Paretian distribution with a tail index lower than two. A reason for concern here is the fact that the tail exponent bounds the number of finite moments, i.e. a tail exponent equal or lower than two implies that all moments higher than and including the second moment do not exist. Mandelbrot and Fama have suggested the use of the minimum sum of absolute errors (MSAE) estimator in the case of regression errors with infinite variance (Mandelbrot, 1963; Fama, 1963).

In an effort to address the efficiency loss of the OLS estimator under errors with infinite variance, Blattberg & Sargent (1971) compare the performance of other estimators to that of the OLS estimator in a simulation experiment. In particular, they examine the performance of the MSAE estimator and a class of best linear unbiased estimators compared to the OLS estimator in a regression model with stable Paretian distributed errors. For values of the tail exponent ranging in between one and two, they evaluate the performance by a measure of dispersion of the distributions of the estimators. They find that the turning point in efficiency lies at a tail exponent of approximately 1.5: for lower values of the tail exponent the MSAE estimator performs considerably better than the OLS estimator, while the OLS estimator performs only slightly better for higher values.

Rather than evaluating efficiency, He et al. (1990) investigates the robustness of regression estimators with errors belonging to a broad class of distributions with either exponential or algebraic tails, or light and heavy tails respectively. More specifically, they evaluate the robustness of the estimators by means of their 'tail performance', a concept they show to be inherently related to the concept of the breakdown point of an estimator (See Donoho & Huber (1983) for breakdown points). In essence, the tail performance criterion measures how fast the probability that an estimator deviates from its true parameter value by some value a tends to zero, compared to how fast the probability that an individual observation exceeds the value a tends to zero, as a goes to infinity. They show that the OLS estimator has very poor tail performance when the errors follow heavy-tailed distributions. Its tail performance does not depend on the sample size and is effectively as worse as with a single observation. Other estimators such as the MSAE estimator and especially the least median of squares (LMS) estimator have considerably better tail performance. However, estimators such as the LMS estimator may suffer considerably in asymptotic convergence rates.

Whereas most studies have focused on the possible deficiencies of the OLS estimator in terms of efficiency or robustness, to the best of my knowledge, the only study in the second stream is Mikosch & de Vries (2013), which provides analytical results on the distribution of the OLS estimator when the error terms follow heavy-tailed distributions. They show that the approximate distribution for the OLS estimator prescribed by the central limit theorem is not a good representation of the true distribution in small samples; the OLS estimator is more likely to experience extreme values when the error terms follow a heavy-tailed distribution compared to the situation in which the error terms are normally distributed. These discrepancies in the distribution of the OLS estimator under heavy-tailed and normally distributed errors become increasingly apparent as one goes further into the tails. Specifically, they find that for a simple linear regression model with additive or multiplicative error terms, the distribution function of the difference between the OLS estimator and the true regression coefficient is regularly varying. To establish this, they use generalizations of the Feller theorem (see Feller (1971), p. 278) and Breiman's result (see

Proposition 3 in Breiman (1965)) to handle the finite sample tail behaviour of the OLS estimator. In addition, they illustrate that their result may be used for hypothesis testing in an empirical application. Since the tail behaviour of the OLS estimator also depends on the generally unknown properties of the stochastic regressors, they extrapolate high quantiles or p-values from intermediate levels. With this approach, the unknown properties of the regressors do not play a role in analyzing the tail behaviour of the OLS estimator. Instead, all that remains to be estimated from the data is the tail index. A prerequisite for empirical applicability of their results is either that the properties of the regressors are known (or can be easily estimated), or that quantiles of the OLS estimator may be cross-sectionally observed. In essence, this thesis extends the results of Mikosch & de Vries (2013) multivariately.

Besides contributing to the literature of regression analysis with heavy-tailed data, this study adds to the methodology for dealing with stochastic linear combinations of heavy-tailed random variables. Previously, van Oordt (2013) has shown that positive linear combinations of positive heavy-tailed random variables exhibit tail dependence. Tail dependence between the OLS estimators follows from an extension of this result to stochastic linear combinations of heavy-tailed random variables. Although the results in this thesis are specific in their application, they can be applied to stochastic linear combinations of heavy-tailed random variables outside the context of regression analysis; for example, to obtain the tail dependence between portfolios with common risk factors.

We will establish the results in this thesis as follows. First, we will define a class of heavy-tailed random variables and review properties of probabilities involving sums and products of random variables belonging to this class in Section 2. Next, by assuming that the distribution of the regression errors belongs to this class we derive the tail dependence between the OLS estimators in Section 3 — first in a regression model with only two regressors and later in a regression model with k regressors. After the theory in Section 3, we examine the tail dependence between the OLS estimators in a regression model with two explanatory variables by means of a simulation study in Section 4. Finally, we illustrate the robustness of the F-test and relate it to the dependence behaviour of the FSS and RSS by a simulation experiment in Section 5. Section 5 is concluded by a derivation of the tail dependence between the FSS and RSS. Section 6 concludes the thesis.

2 Preliminaries

To analyze the finite sample tail behaviour of the OLS estimator when the regression error terms are heavy-tailed, we start with defining heavy-tailed random variables. Heuristically, the difference between thin-tailed and heavy-tailed variables may be seen from the conditional expectation $\mathbb{E}(X|X > x)$ as $x \to \infty$: for thin-tailed variables this limiting conditional expectation will be at the same level of x, while it will be at a higher level than x for heavy-tailed variables. This difference occurs because the speed of probability decay is substantially slower for heavy-tailed variables. This section reviews essential parts of the Extreme Value Theory (EVT) literature to make the previous intuition more rigorous and to provide a framework in which we can obtain the tail dependence between the OLS estimators.

2.1 Univariate Regular Variation

In what follows, the "small-o" notation $h(x) = o(f(x))$ is shorthand for $h(x)/f(x) \to 0$ as $x \to \infty$. Furthermore, for positive functions a and b, $a(x) \sim b(x)$ is shorthand for $a(x)/b(x) \to 1$ as $x \to \infty$.

By Bingham et al. (1989) we have the following definitions.

Definition 1. *Let L be a positive measurable function, defined on $[0, \infty)$, and satisfying*

$$\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1, \qquad \forall\lambda > 0,$$

then L is called a slowly varying function at infinity.

Constant functions trivially satisfy Definition 1. Other functions satisfying Definition 1 include logarithmic functions and powers thereof.

Definition 2. *Let $\alpha \in R$, let L be a slowly varying function at infinity and let g be a positive measurable function defined as*

$$g(x) = x^{\alpha} L(x),$$

then the function g is said to be regularly varying at infinity with index α.

For slowly varying functions such as constants or logarithms it is easy to verify that for $\alpha > 0$ we have

$x^{-\alpha}L(x) \to 0$ as $x \to \infty$. In fact, this property holds for all slowly varying functions, see Proposition 1.3.6 in Bingham et al. (1989).

Definition 3. *A random variable X or its distribution function F is called regularly varying with tail index $\alpha > 0$ if there exists $p, q \geq 0$ with $p + q = 1$ and a slowly varying function L such that as $x \to \infty$*

$$F(-x) = qx^{-\alpha}L(x)(1 + o(1)), \;\; 1 - F(x) = px^{-\alpha}L(x)(1 + o(1)). \tag{1}$$

Condition (1) is referred to as a tail balance condition. The constants q and p represent the limit of the fractions $\frac{F(-x)}{F(-x)+1-F(x)}$ and $\frac{1-F(x)}{F(-x)+1-F(x)}$ respectively as $x \to \infty$. Here, the $o(1)$ term in (1) indicates that the left and right tail of the distribution function are not necessarily equal to $qx^{-\alpha}L(x)$ and $px^{-\alpha}L(x)$ respectively for all x. The following example may help to clarify the tail balance condition

Example 1. *Let X have a Pareto distribution with parameters $\alpha, A > 0$. Then, the survival function of X satisfies*

$$1 - F(x) = Ax^{-\alpha}, \qquad \textit{for } x > A^{\frac{1}{\alpha}}.$$

The Pareto distribution is a special case that strictly satisfies the tail balance condition for all x. Here, we may recognize $L(x) = A$ and $p = 1$.

From (1) we observe that regularly varying random variables satisfy that

$$\lim_{x\to\infty} \mathrm{P}(X > \lambda x | X > x) = \lambda^{-\alpha}, \qquad \text{for } \lambda > 1. \tag{2}$$

This conditional probability represents a scaling difference and illustrates that the parameter α determines the speed of probability decay. The lower the value of α, the heavier the tail of the distribution function in the sense that it decays more slowly. By (5.20) in McNeil et al. (2015) we have as $x \to \infty$,

$$\frac{\mathbb{E}(X|X > x)}{x} \to \frac{\alpha}{\alpha - 1}.$$

For the thin-tailed distributions such as the normal distribution, this limit relation can be read as the limit case $\alpha \to +\infty$. That is, $\mathbb{E}(X|X > x) \sim x$ as $x \to \infty$. In this sense the asymptotic behaviour of thin-tailed variables is fundamentally different from heavy-tailed variables.

2.2 Multivariate Regular Variation

There are several equivalent ways to formulate multivariate regular variation for random vectors (see Theorem 6.1 in Resnick (2007)). First, we will address some notation. Let $|\cdot|$ denote any norm on $\mathbb{R}^d$ and let ∂B denote the boundary of any Borel set $B \subset \bar{\mathbb{R}}^d_0$, where $\bar{\mathbb{R}}^d_0 = \bar{\mathbb{R}}^d \setminus \{0\}$. We are now able to formulate the following definition of a regularly varying random vector and its distribution function F.

Definition 4. *Let $X = (X_1, ..., X_d)'$ be a random vector taking values in $\mathbb{R}^d$. X or its distribution function F is called multivariate regularly varying with index $\alpha > 0$ if there exists a Radon measure μ on $\bar{\mathbb{R}}^d_0$ such that for every Borel set $B \subset \bar{\mathbb{R}}^d_0$ bounded away from the origin satisfying that $\mu(\partial B) = 0$,*

$$\lim_{x\to\infty} \frac{\mathrm{P}(X \in xB)}{\mathrm{P}(|X| > x)} = \mu(B),$$

with the homogeneity condition $\mu(xB) = x^{-\alpha}\mu(B)$.

In general, univariate regular variation of the random variables $X_1, ..., X_d$ does not imply multivariate regular variation of the random vector X. A sufficient condition for multivariate regular variation of X is that the regularly varying random variables $X_1, ..., X_d$ are independent. In that case, the measure μ does not assign mass off the axis (see Chapter 6.5 in Resnick (2007)). Multivariate regular variation is, however, especially useful to deal with more complex dependence structures among the individual components.

2.3 Sums and Products involving Regularly Varying Random Variables

The following theorem states a convenient property of sums of independent regularly varying variables, see (Feller, 1971, p. 278).

Theorem 1. *Let η_1 and η_2 be two positive, independent regularly varying random variables with tail index α. Then, the sum $\eta_1 + \eta_2$ is regularly varying with tail index α. Furthermore, as $x \to \infty$,*

$$\mathrm{P}(\eta_1 + \eta_2 > x) \sim \mathrm{P}(\eta_1 > x) + \mathrm{P}(\eta_2 > x). \tag{3}$$

From (3) it follows that as x becomes very large, the event $\eta_1+\eta_2 > x$ is solely caused by an extremely large value of either η_1 or η_2. Also, it illustrates that a sum of heavy-tailed variables asymptotically behaves as the maximum of the variables involved in the sum (See, for example, Remark 1.3.7 in Mikosch (1999)).

The following theorem handles the product of a regularly varying random variable and a random variable with a 'lighter tail', see Proposition 3 in Breiman (1965).

Theorem 2. *Let η_1 be a positive, regularly varying random variable with tail index α. Let C_1 be a positive random variable independent of η_1 that satisfies $\mathbb{E}(C_1^{\alpha+\delta}) < \infty$ for some $\delta > 0$. Then, it holds that the product $C_1\eta_1$ is regularly varying with tail index α and that as $x \to \infty$,*

$$\mathrm{P}(C_1\eta_1 > x) \sim \mathbb{E}(C_1^{\alpha})\mathrm{P}(\eta_1 > x). \tag{4}$$

Notice that from (2) we have for non-stochastic $c_1 > 0$ as $x \to \infty$,

$$\mathrm{P}(c_1\eta_1 > x) \sim c_1^{\alpha}\mathrm{P}(\eta_1 > x). \tag{5}$$

We thus make the following analogy between (5) and (4): as long as the random variable C_1 is of 'lighter tail' than η_1, it will behave as if it were a constant as x becomes really large.

The following theorem extends Breiman's result to a multivariate regularly varying random vector X and another random vector $Y = (Y_1, ..., Y_d)'$, see Lemma 3.4 in Mikosch & de Vries (2013).

Theorem 3. *Assume that X is multivariate regularly varying in $\mathbb{R}^d$ with index $\alpha > 0$ and is independent of the random vector Y, which satisfies $\mathbb{E}(|Y|^{\alpha+\epsilon}) < \infty$ for some $\epsilon > 0$. Then the scalar product $Z = X'Y$ is regularly varying with index α. Moreover, if X has independent components, then as $x \to \infty$*

$$P(Z > x) \sim P(|X| > x) \,(\sum_{i=1}^{d} c_i^{+}\mathbb{E}(Y_i^{\alpha}\mathbb{I}(Y_i > 0) + \sum_{i=1}^{d} c_i^{-}\mathbb{E}(|Y_i|^{\alpha}\mathbb{I}(Y_i < 0)),$$
$$\textit{with } c_i^{+} = \lim_{x\to\infty}\frac{P(X_i > x)}{P(|X| > x)},\ c_i^{-} = \lim_{x\to\infty}\frac{P(X_i < -x)}{P(|X| > x)}.$$

Crudely, we can understand the expansion of $P(Z > x)$ in Theorem 3 by considering it to be the result of repeatedly applying the Feller theorem and Breiman's result to the 'positive part' of the terms $X_i Y_i$.

3 Theory

3.1 Regression Model with Two Explanatory Variables

We start with deriving the tail dependence between the OLS estimators in a regression model with two explanatory variables . For this purpose, consider the regression model

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \eta_t, \qquad \text{for } t = 1, 2, ..., n, \tag{6}$$

where (η_t) is an independent and identically distributed (i.i.d.) error sequence of random variables, $(X_{1,t}, X_{2,t})$ is an i.i.d. sequence of bivariate random vectors containing explanatory variables, independent of the error sequence (η_t). β_1 and β_2 are parameters to be estimated. Furthermore, we denote the vectors $Y = (Y_1, ..., Y_n)'$, $X_1 = (X_{1,1}, X_{1,2}, ..., X_{1,n})'$, $X_2 = (X_{2,1}, X_{2,2}, ..., X_{2,n})'$, $\eta = (\eta_1, \eta_2, ..., \eta_n)'$ and the matrix $X = (X_1 \;\; X_2)$.

The OLS estimators for β_1 and β_2 are given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1} X'Y. \tag{7}$$

We substitute the random vector Y in (7) by the model in (6) and obtain the expansion of the OLS estimators for β_1 and β_2 as

$$\hat{\beta}_1 = \beta_1 + \frac{X_2' X_2 X_1' - X_1' X_2 X_2'}{X_1' X_1 X_2' X_2 - (X_1' X_2)^2}\eta, \;\; \hat{\beta}_2 = \beta_2 + \frac{X_1' X_1 X_2' - X_2' X_1 X_1'}{X_1' X_1 X_2' X_2 - (X_1' X_2)^2}\eta. \tag{8}$$

From (8) we observe that the OLS estimators consist of randomly weighted sums of the error terms in (η_t). Write $\hat{\beta}_1 = \beta_1 + \sum_{t=1}^{n} C_t \eta_t$ and $\hat{\beta}_2 = \beta_2 + \sum_{t=1}^{n} D_t \eta_t$, where

$$C_t = \frac{X_{1,t} \sum_{t=1}^{n} {X_{2,t}}^2 - X_{2,t} \sum_{t=1}^{n} X_{1,t} X_{2,t}}{\sum_{t=1}^{n} {X_{1,t}}^2 \sum_{t=1}^{n} {X_{2,t}}^2 - \left(\sum_{t=1}^{n} X_{1,t} X_{2,t}\right)^2}, \quad D_t = \frac{X_{2,t} \sum_{t=1}^{n} {X_{1,t}}^2 - X_{1,t} \sum_{t=1}^{n} X_{1,t} X_{2,t}}{\sum_{t=1}^{n} {X_{1,t}}^2 \sum_{t=1}^{n} {X_{2,t}}^2 - \left(\sum_{t=1}^{n} X_{1,t} X_{2,t}\right)^2},$$

such that (C_t, D_t) is an identically distributed, but not independent, sequence of random vectors. Consequently, the marginal and joint behaviour of $\hat{\beta}_1$ and $\hat{\beta}_2$ depends on the distribution of η_t. To continue, we need to make a distributional assumption on η_t.

In what follows we assume that η_t is regularly varying with tail index α. That is, the distribution function of η_t, F_η, satisfies Definition 3 such that $1-F_\eta(x) = px^{-\alpha}L(x)(1+o(1))$ and $F_\eta(-x) = qx^{-\alpha}L(x)(1+o(1))$ as $x \to \infty$. Note that this assumption on η_t is semi-parametric in the sense that it does not impose any restrictions on moderate levels of the distribution. In the next sections we will evaluate the tail dependence among the OLS estimators that arises from this assumption.

3.2 Conditional Joint Tail Behaviour of the OLS Estimators

Tail dependence between any two random variables Z and W is defined as

$$\lambda_{Z,W} =: \lim_{u \to 1} \mathrm{P}(Z > Q_u(Z) | W > Q_u(W)),$$

where $Q_u(X)$ denotes the quantile of the CDF of the random variable X, as $Q_u(X) = \inf\{l : F_X(l) \geq u\}$. Equivalently, we can write the tail dependence as

$$\lambda_{Z,W} = \lim_{u \to 1} \frac{\mathrm{P}(Z > Q_u(Z), W > Q_u(W))}{1-u} \tag{9}$$

Consequently, as a first step to obtain the tail dependence between the OLS estimators we should analyze the joint probability that both estimators exceed large thresholds. Since $\hat{\beta}_1$ and $\hat{\beta}_2$ are dependent through weighted sums of heavy-tailed error terms, we can obtain an expansion of this joint probability similar to the Feller theorem. Here the weights C_t and D_t are stochastic. To avoid dealing with stochastic weights directly, we use the relation

$$\mathrm{P}(\hat{\beta}_1 - \beta_1 > x,\ \hat{\beta}_2 - \beta_2 > y) = \mathbb{E}(\mathrm{P}(\hat{\beta}_1 - \beta_1 > x,\ \hat{\beta}_2 - \beta_2 > y | X_1, X_2)). \tag{10}$$

Notice that conditional on the regressors X_1 and X_2, the weights C_t and D_t can be regarded as non-stochastic. We thus need to handle non-stochastic linear combinations first. Therefore, we devote this section to obtaining an expansion for $\mathrm{P}(\hat{\beta}_1 - \beta_1 > x,\ \hat{\beta}_2 - \beta_2 > y | X_1, X_2)$ in terms of the weights C_t and D_t for large x and y.

To deal with the joint probability of two linear combinations of regularly varying random variables, we use a method of proof similar to the proof of the Feller theorem in Theorem 3, see (Feller, 1971, p. 278). Whereas the random variables are assumed to be postive in the Feller theorem, both the weights (C_t, D_t) and η_t can be real-valued in our case. In order to handle that, we split the calculation of the joint tail probability into several lemmas gearing towards the final result in Theorem 4 below. The following lemma first deals with positive coefficients and positive varying random variables only.

Lemma 1. *Let η_1 and η_2 be positive regularly varying random variables. Assume that they are independent. Also, assume $c_i, d_i > 0$ are fixed coefficients for $i = 1, 2$. In addition, assume that $\frac{x}{y} \to c$ as $x \to \infty$, with $c > 0$. Then, as $x \to \infty$,*

$$\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \sim \mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}). \tag{11}$$

Proof of Lemma 1. The joint probability on the left hand side of (11) is handled by providing its upper and lower bounds. This is done by finding an appropriate subset and superset of the joint region. For any $0 < \delta < 1/2$ we have that

$$\{c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y\} \supset \{\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}\} \cup \{\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}\}, \tag{12}$$

$$
\begin{aligned}
\{c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y\} \subset & \{\eta_1 > (1-\delta)(\frac{x}{c_1} \vee \frac{y}{d_1})\} \cup \{\eta_2 > (1-\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})\} \\
& \cup \{\eta_1 > \frac{\delta x}{c_1} \wedge \frac{\delta y}{d_1},\ \eta_2 > \frac{\delta x}{c_2} \wedge \frac{\delta y}{d_2}\}.
\end{aligned}
\tag{13}
$$

By the independence of η_1 and η_2, the set relation (12) implies a lower bound on the joint probability as

$$
\begin{aligned}
\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \geq & \mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}) \\
& - \mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1})\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}).
\end{aligned}
\tag{14}
$$

Now we show that the last term on the right hand side of (14) is negligible compared to the previous two terms for large x and y. Since we have that as $x \to \infty$

$$
\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) \to 0,\ \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}) \to 0.
$$

it follows that

$$
\frac{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1})\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \to 0,\ \text{as } x \to \infty,
$$

because the product of these probabilities is of higher order than their sum. Consequently, we get that

$$
\liminf_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \geq 1.
\tag{15}
$$

Next, from (13) we get an upper bound on the joint probability as

$$
\begin{aligned}
\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \leq & \mathrm{P}(\eta_1 > (1-\delta)(\frac{x}{c_1} \vee \frac{y}{d_1})) + \mathrm{P}(\eta_2 > (1-\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})) \\
& + \mathrm{P}(\eta_1 > \frac{\delta x}{c_1} \wedge \frac{\delta y}{d_1})\mathrm{P}(\eta_2 > \frac{\delta x}{c_2} \wedge \frac{\delta y}{d_2}) =: \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3.
\end{aligned}
$$

Similar to the proof above for the lower bound, I_3 is negligible with respect to $\mathrm{I}_1 + \mathrm{I}_2$ as $x \to \infty$.

Furthermore, since η_1 and η_2 are random variables with tail index α we get that

$$\frac{\mathrm{I}_1}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1})} \to (1-\delta)^{-\alpha}, \quad \frac{\mathrm{I}_2}{\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \to (1-\delta)^{-\alpha}, \text{ as } x \to \infty.$$

Therefore,

$$\frac{\mathrm{I}_1 + \mathrm{I}_2}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \to (1-\delta)^{-\alpha}, \text{ as } x \to \infty.$$

Consequently, we obtain an upper bound on the joint probability as

$$\limsup_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \leq (1-\delta)^{-\alpha}.$$

Now because δ is arbitrarily chosen within $(0,\ 1/2)$, we take $\delta \to 0$ to obtain

$$\limsup_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \leq 1 \tag{16}$$

By combining the upper and lower bound in (15) and (16) respectively, (11) is proved. $\square$

The intuition for (11) is analogous to (3): as x becomes very large, the events $c_1\eta_1 + c_2\eta_2 > x$ and $d_1\eta_1 + d_2\eta_2 > y$ happen simultaneously solely due to a sufficiently large value of either η_1 or η_2.

Next, we extend Lemma 1 to real-valued coefficients and positive random variables. We relax the assumption of independence to the weaker assumption that either the random variables are independent or $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0) = 0$.

Lemma 2. *Let η_1 and η_2 be positive regularly varying random variables. Assume they are either independent or $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0) = 0$. Also, assume that c_i, d_i are non-zero coefficients for $i = 1, 2$. In addition, assume that $\frac{x}{y} \to c$ as $x \to \infty$, with $c > 0$. It holds that as $x \to \infty$,*

$$\begin{aligned}
\mathrm{P}(c_1\eta_1+c_2\eta_2>x,\ d_1\eta_1+d_2\eta_2>y) \sim & \mathbb{I}\{c_1>0,d_1>0\}\mathrm{P}(\eta_1>\frac{x}{c_1}\vee\frac{y}{d_1}) \\
& +\mathbb{I}\{c_2>0,d_2>0\}\mathrm{P}(\eta_2>\frac{x}{c_2}\vee\frac{y}{d_2}),
\end{aligned}$$

if $\mathbb{I}\{c_1>0,d_1>0\}+\mathbb{I}\{c_2>0,d_2>0\}=0$, *the above relation should be read as*

$$\mathrm{P}(c_1\eta_1+c_2\eta_2>x,\ d_1\eta_1+d_2\eta_2>y)=o(\mathrm{P}(\eta_1>\frac{x}{c_1}\vee\frac{y}{d_1})+\mathrm{P}(\eta_2>\frac{x}{c_2}\vee\frac{y}{d_2})).$$

Proof of Lemma 2. To prove Lemma 2 we handle set manipulations equations analogous to (12) and (13) for real-valued coefficients. Due to the inherent symmetry in the problem we only consider 7 of the 16 cases.

Case 1: $c_1<0$ and $c_2,d_1,d_2>0$

In this case upper and lower bounds on the joint probability can be found by noting that for any $0<\delta<1/2$ we have

$$\{c_1\eta_1+c_2\eta_2>x,\ d_1\eta_1+d_2\eta_2>y\}\supset\{\eta_1<\frac{\delta x}{-c_1},\ \eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2})\}, \tag{17}$$

$$\{c_1\eta_1+c_2\eta_2>x,\ d_1\eta_1+d_2\eta_2>y\}\subset\{\eta_2>(1-\delta)(\frac{x}{c_2}\vee\frac{y}{d_2})\}\cup\{\eta_1>\frac{\delta y}{d_1},\ \eta_2>\frac{x}{c_2}\}. \tag{18}$$

From (17) we get that the lower bound on the joint probability is

$$\mathrm{P}(c_1\eta_1+c_2\eta_2>x,\ d_1\eta_1+d_2\eta_2>y)\geq\mathrm{P}(\eta_1<\frac{\delta x}{-c_1},\ \eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2})).$$

If $\mathrm{P}(\eta_1>0,\ \eta_2>0=0)$ we have that $\mathrm{P}(\eta_1<\frac{\delta x}{-c_1},\ \eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2}))=\mathrm{P}(\eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2}))$. If η_1 and η_2 are independent we have $\mathrm{P}(\eta_1<\frac{\delta x}{-c_1},\ \eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2}))=\mathrm{P}(\eta_1<\frac{\delta x}{-c_1})\mathrm{P}(\eta_2>(1+\delta)(\frac{x}{c_2}\vee\frac{y}{d_2}))$. Note that $\mathrm{P}(\eta_1<\frac{\delta x}{-c_1})\to1$ as $x\to\infty$. Similar to the proof of Lemma 1, by first letting $x\to\infty$ and then letting $\delta \to 0$ we have

$$\liminf_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \geq 1.$$

From (18) we get that the upper bound on the joint probability is

$$\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \leq \mathrm{P}(\eta_2 > (1-\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})) + \mathrm{P}(\eta_1 > \frac{\delta y}{d_1},\ \eta_2 > \frac{y}{d_2}). \qquad (19)$$

Analogous to the proof of Lemma 1, the last term on the right hand side of (19) is negligible compared to $\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$ as $x \to \infty$. This holds if η_1 and η_2 are independent and holds trivially if $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0) = 0$. By letting $\delta \to 0$ we have

$$\limsup_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \leq 1.$$

Combining the upper and lower bound gives the result.

Case 2: $c_1, c_2 < 0$ and $d_1, d_2 > 0$

The only feasible regions can appear in the second or fourth quadrant of the η_1, η_2 plane. Because η_1 and η_2 are positive-valued we obtain that $\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) = 0$ for $x, y > 0$.

Case 3: $c_1, d_1 < 0$ and $c_2, d_2 > 0$

In this case upper and lower bounds on the joint probability can be found by noting that for any $0 < \delta < 1/2$ we have

$$\{c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y\} \supset \{\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1},\ \eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})\}, \qquad (20)$$

$$\{c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y\} \subset \{\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}\}. \qquad (21)$$

From (20) we get that the lower bound on the joint probability is

$$\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \geq \mathrm{P}(\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1},\ \eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})).$$

If $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0 = 0)$ we have that $\mathrm{P}(\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1},\ \eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})) = \mathrm{P}(\eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2}))$. If η_1 and η_2 are independent we have $\mathrm{P}(\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1},\ \eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2})) = \mathrm{P}(\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1})$ $\mathrm{P}(\eta_2 > (1+\delta)(\frac{x}{c_2} \vee \frac{y}{d_2}))$. Note that $\mathrm{P}(\eta_1 < \frac{\delta x}{-c_1} \wedge \frac{\delta y}{-d_1}) \to 1$ as $x \to \infty$. By first letting $x \to \infty$ and then letting $\delta \to 0$ we have

$$\liminf_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \geq 1.$$

From (21) we immediately have the upper bound

$$\limsup_{x\to\infty} \frac{\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)}{\mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})} \leq 1.$$

The result then follows by combining the upper and lower bound.

Case 4: $c_1, d_2 < 0$ and $d_1, c_2 > 0$

In this case the only feasible regions may occur in the third or first quadrant of the η_1, η_2 plane. Any region in the third quadrant has probability zero since η_1 and η_2 are positive-valued. Any region in the first quadrant can be contained in a region of the form $\mathrm{P}(\eta_1 > a, \eta_2 > b)$, where a and b are the η_1, η_2 coordinates of the point of intersection of the boundaries of the regions $\{c_1\eta_1 + c_2\eta_2 > x\}$ and $\{d_1\eta_1 + d_2\eta_2 > y\}$. As in Lemma 1, we have that if η_1 and η_2 are independent $\mathrm{P}(\eta_1 > a, \eta_2 > b)$ is negligible compared to $\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$ as $x \to \infty$. Therefore, we obtain that $\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y)$ is negligible compared to $\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$. If it holds that $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0) = 0$ then clearly the joint probability is negligible compared to $\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$ as well. Therefore, we have that $\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) = o(\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2}))$.

Case 5: $c_1, c_2, d_1 < 0$ and $d_2 > 0$

In this case feasible regions only occur in the third or fourth quadrant of the η_1, η_2 plane. Since η_1 and

η_2 are positive-valued we obtain that $\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) = 0$ for $x, y > 0$.

Case 6: $c_1, c_2, d_1, d_2 < 0$

In this case feasible regions only occur outside the first quadrant of the η_1, η_2 plane. Since η_1 and η_2 are positive-valued we obtain that $\mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) = 0$ for $x, y > 0$.

Case 7: $c_1, c_2, d_1, d_2 > 0$

The result follows immediately from Lemma 1 if η_1 and η_2 are independent. The proof of Lemma 1 only uses the independence between η_1 and η_2 to establish that the joint probability $\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{c_1},\ \eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$ is negligible with respect to $\mathrm{P}(\eta_1 > \frac{x}{c_1} \vee \frac{y}{c_1}) + \mathrm{P}(\eta_2 > \frac{x}{c_2} \vee \frac{y}{d_2})$ as $x \to \infty$. This is trivially established if $\mathrm{P}(\eta_1 > 0,\ \eta_2 > 0) = 0$. Therefore, the result follows by Lemma 1. $\square$

As a corollary of Lemma 2, we extend Lemma 2 to an arbitrary finite number of positive regularly varying random variables. That is, η_i satisfies $1 - F_i(x) = x^{-\alpha}L_i(x)(1 + o(1))$ as $x \to \infty$ for $i = 1, 2, ..., n$. This extension will be useful to handle real-valued random variables.

Corollary 1. *Let $\eta_1, \eta_2, ..., \eta_n$ be positive regularly varying random variables. Assume that either η_i is independent of η_j or $\mathrm{P}(\eta_i > 0, \eta_j > 0) = 0$ for any $i \neq j$. Also, assume that c_i, d_i are non-zero coefficients for $i = 1, 2, ..., n$. In addition, assume that $\frac{x}{y} \to c$ as $x \to \infty$, with $c > 0$. Then, it holds that as $x \to \infty$*

$$\mathrm{P}(\sum_{i=1}^{n} c_i\eta_i > x,\ \sum_{i=1}^{n} d_i\eta_i > y) \sim \sum_{i=1}^{n} \mathbb{I}\{c_i > 0, d_i > 0\}\mathrm{P}(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}).$$

In the final step, we extend Lemma 2 to real-valued regularly varying random variables. That is, η_i satisfies $1 - F_i(x) = px^{-\alpha}L_i(x)(1 + o(1))$ and $F_i(-x) = qx^{-\alpha}L_i(x)(1 + o(1))$ as $x \to \infty$, for $i = 1, 2$ whereas q was implicitly assumed to be zero before.

Theorem 4. *Let η_1 and η_2 be real-valued regularly varying random variables. Assume that they are independent. Also, assume that c_i, d_i are non-zero coefficients for $i = 1, 2$. In addition, assume that*

$\frac{x}{y} \to c$ *as* $x \to \infty$, *with* $c > 0$. *Then, it holds that as* $x \to \infty$

$$\begin{aligned} \mathrm{P}(c_1\eta_1 + c_2\eta_2 > x,\ d_1\eta_1 + d_2\eta_2 > y) \sim & \sum_{i=1}^{2} \mathbb{I}\{c_i > 0, d_i > 0\}\mathrm{P}(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}) \\ & + \sum_{i=1}^{2} \mathbb{I}\{c_i < 0, d_i < 0\}\mathrm{P}(\eta_i < \frac{x}{c_i} \wedge \frac{y}{d_i}). \end{aligned} \tag{22}$$

Proof of Theorem 4. Let $\eta_i^+ = \max\{\eta_i, 0\}$ and let $\eta_i^- = -\min\{\eta_i, 0\}$. Also, let $e_i = -c_i$ and $z_i = -d_i$. By definition it holds that $\mathrm{P}(\eta_i^+ > 0,\ \eta_i^- > 0) = 0$. Also by definition, it holds that

$$c_1\eta_1 + c_2\eta_2 = c_1\eta_1^+ + e_1\eta_1^- + c_2\eta_2^+ + e_2\eta_2^-, \quad d_1\eta_1 + d_2\eta_2 = d_1\eta_1^+ + z_1\eta_1^- + d_2\eta_2^+ + z_2\eta_2^-.$$

The random variables now satisfy the conditions of Corollary 1. Therefore, as $x \to \infty$

$$\begin{aligned} & \mathrm{P}(c_1\eta_1^+ + e_1\eta_1^- + c_2\eta_2^+ + e_2\eta_2^- > x,\ d_1\eta_1^+ + z_1\eta_1^- + d_2\eta_2^+ + z_2\eta_2^- > y) \sim \\ & \mathbb{I}\{c_1 > 0, d_1 > 0\}\mathrm{P}(\eta_1^+ > \frac{x}{c_1} \vee \frac{y}{d_1}) + \mathbb{I}\{c_2 > 0, d_2 > 0\}\mathrm{P}(\eta_2^+ > \frac{x}{c_2} \vee \frac{y}{d_2}) \\ & + \mathbb{I}\{e_1 > 0, z_1 > 0\}\mathrm{P}(\eta_1^- > \frac{x}{e_1} \vee \frac{y}{z_1}) + \mathbb{I}\{e_2 > 0, z_2 > 0\}\mathrm{P}(\eta_2^- > \frac{x}{e_2} \vee \frac{y}{z_2}) \\ & = \sum_{i=1}^{2} \mathbb{I}\{c_i > 0, d_i > 0\}\mathrm{P}(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}) + \sum_{i=1}^{2} \mathbb{I}\{c_i < 0, d_i < 0\}\mathrm{P}(\eta_i < \frac{x}{c_i} \wedge \frac{y}{d_i}). \end{aligned}$$

$\square$

The form of (22) is not surprising given the earlier results; the events $c_1\eta_1 + c_2\eta_2 > x$ and $d_1\eta_1 + d_2\eta_2 > y$ happen simultaneously solely due to a sufficiently large absolute value of η_1 or η_2 if their corresponding coefficients have the same sign. Theorem 4 can be extended to an arbitrary finite number of real-valued regularly varying random variables.

Corollary 2. *Let* $\eta_1, \eta_2, ..., \eta_n$ *be real-valued, regularly varying random variables. Assume that* η_i *is independent of* η_j *for* $i \neq j$. *Also, assume that* c_i, d_i *are non-zero coefficients for* $i = 1, 2, ..., n$. *In addition, assume that* $\frac{x}{y} \to c$ *as* $x \to \infty$, *with* $c > 0$. *Then, it holds that as* $x \to \infty$

$$\begin{aligned} \mathrm{P}(\sum_{i=1}^{n} c_i\eta_i > x,\ \sum_{i=1}^{n} d_i\eta_i > y) \sim & \sum_{i=1}^{n} \mathbb{I}\{c_i > 0, d_i > 0\}\mathrm{P}(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}) \\ & + \sum_{i=1}^{n} \mathbb{I}\{c_i < 0, d_i < 0\}\mathrm{P}(\eta_i < \frac{x}{c_i} \wedge \frac{y}{d_i}). \end{aligned}$$

Corollary 2 enables us to calculate the joint tail probability of the OLS estimators since we have as $x \to \infty$,

$$\begin{aligned} \mathrm{P}(\hat{\beta}_1 - \beta_1 > x,\ \hat{\beta}_2 - \beta_2 > y | X_1, X_2) \sim & \sum_{t=1}^{n} \mathbb{I}(C_t > 0, D_t > 0) \mathrm{P}(\eta_t > \frac{x}{C_t} \vee \frac{y}{D_t}) \\ & + \sum_{t=1}^{n} \mathbb{I}(C_t < 0, D_t < 0) \mathrm{P}(\eta_t < \frac{x}{C_t} \wedge \frac{y}{D_t}). \end{aligned}$$

As a final corollary of Theorem 4, we consider the joint tail probability of several linear combinations of regularly varying random variables. This will be used when we consider a regression model with k regressors.

Corollary 3. *Let $\eta_1, \eta_2, ..., \eta_n$ be real-valued regularly varying random variables. Assume that η_i is independent of η_j for $i \neq j$. Also, assume that $w_{1,i}, ..., w_{k,i}$ are non-zero coefficients for $i = 1, 2, ..., n$. In addition, assume that $\frac{x_1}{x_z} \to c_z$ as $x_1 \to \infty$, with $c_z > 0$ for $z = 2, ..., k$. Then, it holds that as $x_1 \to \infty$,*

$$\begin{aligned} \mathrm{P}(\sum_{i=1}^{n} w_{1,i}\eta_i > x_1\ , ..., \ \sum_{i=1}^{n} w_{k,i}\eta_i > x_k) \sim & \sum_{i=1}^{n} \mathbb{I}\{w_{1,i} > 0, ..., w_{k,i} > 0\} \mathrm{P}(\eta_i > \bigvee_{z=1}^{k} \frac{x_z}{w_{z,i}}) \\ & + \sum_{i=1}^{n} \mathbb{I}\{w_{1,i} < 0, ..., w_{k,i} < 0\} \mathrm{P}(\eta_i < \bigwedge_{z=1}^{k} \frac{x_z}{w_{z,i}}). \end{aligned}$$

3.3 Tail Dependence between the OLS Estimators

From (9) we infer that the tail dependence between the OLS estimators is given by

$$\lambda_{\hat{\beta}_1, \hat{\beta}_2} = \lim_{u \to 1} \frac{\mathrm{P}(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1),\ \hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2))}{1-u}.$$

Note that by relation (10) and interchanging the order of the expectation and limit operator (assuming this is allowed) we have

$$\lambda_{\hat{\beta}_1, \hat{\beta}_2} = \mathbb{E}\left(\lim_{u \to 1} \frac{\tilde{\mathrm{P}}(\hat{\beta}_1 - \beta_1 > \mathrm{Q}_u(\hat{\beta}_1 - \beta_1),\ \hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2))}{1-u} \right), \tag{23}$$

where $\tilde{\mathrm{P}}(...)$ is shorthand for $\mathrm{P}(...|X_1, X_2)$. Equation (23) forms the basis for calculating the tail dependence between the OLS estimators. First, we derive the 'conditional tail dependence', the term within the expectation operator on the right hand side of (23), using the results of the previous section. Finally, we will take the expectation over the conditional tail dependence to obtain the unconditional tail dependence.

The following lemma first deals with $Q_u(\hat{\beta}_1 - \beta_1)$. This lemma will be useful in a later step because it will effectively cancel the denominator, $1-u$.

Lemma 3. *Let $\eta_1, \eta_2, ..., \eta_n$ be real-valued regularly varying i.i.d. random variables satisfying $p, q > 0$. Assume that (C_t) is a sequence of identically distributed real-valued random variables independent of η_t for $t = 1, 2, ..., n$. Furthermore, assume that C_t satisfies $\mathbb{E}(|C_t|^{\alpha+\epsilon}) < \infty$, for some $\epsilon > 0$. Then, it holds as $u \to 1$,*

$$\begin{aligned} \mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t)) &\sim \frac{p(1-u)}{np\mathbb{E}(\mathbb{I}(C_t > 0)C_t^{\alpha}) + nq\mathbb{E}(\mathbb{I}(C_t < 0)(-C_t)^{\alpha})}, \\ \mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^{n} C_t\eta_t)) &\sim \frac{q(1-u)}{np\mathbb{E}(\mathbb{I}(C_t > 0)C_t^{\alpha}) + nq\mathbb{E}(\mathbb{I}(C_t < 0)(-C_t)^{\alpha})}. \end{aligned} \tag{24}$$

Proof of Lemma 3. To prove Lemma 3 we will use multivariate regular variation. Since (η_t) is an i.i.d sequence of regularly varying random variables with tail index α, we have that random vector η is multivariate regularly varying with tail index α. Furthermore, we denote the random vector C by $C = (C_1, C_2, ..., C_n)'$. Note that we have by definition

$$\mathrm{P}(\sum_{t=1}^{n} C_t\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t)) = 1 - u.$$

The equation above will be divided by the probability that η_t exceeds the quantile $Q_u(\sum_{t=1}^{n} C_t\eta_t)$ to obtain an expression that is independent of $Q_u(\sum_{t=1}^{n} C_t\eta_t)$. Now we will use the multivariate extension of Breiman's lemma to handle $\mathrm{P}(\sum_{t=1}^{n} C_t\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t))$. Since C and η satisfy the conditions of Theorem 3, we obtain for an arbitrary norm $|\cdot|$ on $\mathbb{R}^n$ as $u \to 1$,

$$\begin{aligned} \mathrm{P}(C'\eta > Q_u(\sum_{t=1}^{n} C_t\eta_t)) \sim \mathrm{P}(|\eta| > Q_u(\sum_{t=1}^{n} C_t\eta_t))\ (&\sum_{t=1}^{n} \lim_{u\to 1} \frac{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t))}{\mathrm{P}(|\eta| > Q_u(\sum_{t=1}^{n} C_t\eta_t))}\mathbb{E}(C_t^{\alpha}\mathbb{I}(C_t > 0)) \\ &+ \sum_{t=1}^{n} \lim_{u\to 1} \frac{\mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^{n} C_t\eta_t))}{\mathrm{P}(|\eta| > Q_u(\sum_{t=1}^{n} C_t\eta_t))}\mathbb{E}((-C_t)^{\alpha}\mathbb{I}(C_t < 0))) \end{aligned} \tag{25}$$

Since the norm was arbitrary, we can simplify (25) by taking the norm to be the max norm, defined by

$|\eta|_\infty =: \max_t(|\eta_t|)$, and noting that as $x \to \infty$,

$$\mathrm{P}(|\eta|_\infty > x) \sim \mathrm{P}(\max_t(\eta_t) > x) + \mathrm{P}(\min_t(\eta_t) < -x) \sim n(\mathrm{P}(\eta_t > x) + \mathrm{P}(\eta_t < -x)), \tag{26}$$

where the relations follow because (η_t) is a sequence of i.i.d. random variables. By the aid of (26), we have

$$\lim_{u\to 1} \frac{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t))}{\mathrm{P}(|\eta| > Q_u(\sum_{t=1}^n C_t\eta_t))} = \frac{p}{n}, \qquad \lim_{u\to 1} \frac{\mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^n C_t\eta_t))}{\mathrm{P}(|\eta| > Q_u(\sum_{t=1}^n C_t\eta_t))} = \frac{q}{n}.$$

By substituting these limits in (25) we obtain

$$\mathrm{P}(C'\eta > Q_u(\sum_{t=1}^n C_t\eta_t)) \sim \mathrm{P}(|\eta| > Q_u(\sum_{t=1}^n C_t\eta_t))\ (\frac{p}{n}\sum_{t=1}^n \mathbb{E}(C_t^\alpha \mathbb{I}(C_t > 0)) + \frac{q}{n}\sum_{t=1}^n \mathbb{E}((-C_t)^\alpha \mathbb{I}(C_t < 0))). \tag{27}$$

Now, by dividing both sides of (27) by $P(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t))$ and noting that (C_t) is an identically distributed sequence of random variables, we have as $u \to 1$,

$$\frac{1-u}{P(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t))} = \frac{\mathrm{P}(C'\eta > Q_u(\sum_{t=1}^n C_t\eta_t))}{P(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t))} \sim \frac{1}{p}(np\mathbb{E}(C_t^\alpha \mathbb{I}(C_t > 0)) + nq\mathbb{E}((-C_t)^\alpha \mathbb{I}(C_t < 0))). \tag{28}$$

It follows from (28) that as $u \to 1$,

$$\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t)) \sim \frac{p(1-u)}{np\mathbb{E}(\mathbb{I}(C_t > 0)C_t^\alpha) + nq\mathbb{E}(\mathbb{I}(C_t < 0)(-C_t)^\alpha)}. \tag{29}$$

To obtain the second equation of (24), we exploit the ratio between the right and left tail probability of a regularly varying random variable. From the definition of a regularly varying random variable, as $u \to 1$,

$$\frac{\mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^n C_t\eta_t))}{\mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^n C_t\eta_t)) + \mathrm{P}(\eta_t > Q_u(\sum_{t=1}^n C_t\eta_t))} \sim q. \tag{30}$$

Since $p+q=1$, it follows from (30) that as $u \to 1$,

$$\mathrm{P}(\eta_t < -Q_u(\sum_{t=1}^{n} C_t\eta_t)) \sim \frac{q}{p}\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t)). \tag{31}$$

Combining (29) and (31) completes the proof. □

As a corollary of Lemma 3, we show that the limit of $Q_u(\hat{\beta}_1 - \beta_1)/Q_u(\hat{\beta}_2 - \beta_2)$ is finite as $u \to 1$.

Corollary 4. *Let $\eta_1, \eta_2, ..., \eta_n$ be real-valued regularly varying i.i.d. random variables satisfying $p, q > 0$. Assume that (C_t, D_t) is a sequence of identically distributed real-valued random vectors independent of η_t for $t = 1, 2, ..., n$. Furthermore, assume that C_t and D_t satisfy $\mathbb{E}(|C_t|^{\alpha+\epsilon}) < \infty$, $\mathbb{E}(|D_t|^{\alpha+\delta}) < \infty$ for some $\epsilon > 0$, $\delta > 0$. Then, it holds that,*

$$\lim_{u\to 1} \frac{Q_u(\sum_{t=1}^{n} C_t\eta_t)}{Q_u(\sum_{t=1}^{n} D_t\eta_t)} = c, \quad \textit{for some positive constant } c. \tag{32}$$

Proof of Corollary 4. By applying Lemma 3 to $\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t))$ and $\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} D_t\eta_t))$ we obtain

$$\frac{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t))}{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} D_t\eta_t))} \sim \frac{p\mathbb{E}(\mathbb{I}(D_t > 0)D_t^{\alpha}) + q\mathbb{E}(\mathbb{I}(D_t < 0)(-D_t)^{\alpha})}{p\mathbb{E}(\mathbb{I}(C_t > 0)C_t^{\alpha}) + q\mathbb{E}(\mathbb{I}(C_t < 0)(-C_t)^{\alpha})}. \tag{33}$$

Note that the term on the right hand side of (33) is a positive constant. In addition, regular variation of η_t gives us,

$$\frac{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} C_t\eta_t))}{\mathrm{P}(\eta_t > Q_u(\sum_{t=1}^{n} D_t\eta_t))} \sim \left(\frac{Q_u(\sum_{t=1}^{n} C_t\eta_t)}{Q_u(\sum_{t=1}^{n} D_t\eta_t)}\right)^{-\alpha}. \tag{34}$$

By combining the limit relations in (33) and (34) we obtain the result in (32). □

By virtue of Lemma 3 and Corollary 4 we can prove the main result of this section.

Theorem 5. *Assume that there exists a regression model given by*

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \eta_t \quad \textit{for } t = 1, ..., n,$$

where (η_t) is an i.i.d. error sequence of regularly varying random variables satisfying $p, q > 0$, $(X_{1,t}, X_{2,t})$

is an i.i.d. sequence of bivariate random vectors containing explanatory variables, independent of the error sequence (η_t)*. Then, the tail dependence between the OLS estimators for* β_1 *and* β_2 *is given by*

$$\begin{aligned}\lambda_{\hat\beta_1,\hat\beta_2} = &\frac{p\mathbb{E}(\mathbb{I}(C_t>0,D_t>0)C_t^\alpha|C_t<\kappa D_t)P(C_t<\kappa D_t)}{p\mathbb{E}(\mathbb{I}(C_t>0)C_t^\alpha)+q\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^\alpha)} + \frac{q\mathbb{E}(\mathbb{I}(C_t<0,D_t<0)(-C_t)^\alpha|C_t>\kappa D_t)P(C_t>\kappa D_t)}{p\mathbb{E}(\mathbb{I}(C_t>0)C_t^\alpha)+q\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^\alpha)}\\ &+\frac{p\mathbb{E}(\mathbb{I}(C_t>0,D_t>0)D_t^\alpha|C_t>\kappa D_t)P(C_t>\kappa D_t)}{p\mathbb{E}(\mathbb{I}(D_t>0)D_t^\alpha)+q\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^\alpha)} + \frac{q\mathbb{E}(\mathbb{I}(C_t<0,D_t<0)(-D_t)^\alpha|C_t<\kappa D_t)P(C_t<\kappa D_t)}{p\mathbb{E}(\mathbb{I}(D_t>0)D_t^\alpha)+q\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^\alpha)},\end{aligned} \tag{35}$$

where κ *is the positive number given by*

$$\kappa = \left(\frac{p\mathbb{E}(\mathbb{I}(C_t>0)C_t^\alpha)+q\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^\alpha)}{p\mathbb{E}\mathbb{I}(D_t>0)D_t^\alpha)+q\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^\alpha)}\right)^{1/\alpha}. \tag{36}$$

Proof of Theorem 5. We start by deriving the conditional tail dependence between the OLS estimators. By the aid of Corollary 4, we can take $x = Q_u(\hat\beta_1-\beta_1)$ and $y = Q_u(\hat\beta_2-\beta_2)$ in Corollary 2 to obtain as $u \to 1$,

$$\frac{\tilde{\mathrm{P}}(\hat\beta_1-\beta_1 > Q_u(\hat\beta_1-\beta_1),\ \hat\beta_2-\beta_2 > Q_u(\hat\beta_2-\beta_2))}{1-u} \sim \frac{A_u}{1-u}+\frac{B_u}{1-u}, \tag{37}$$

where A_u and B_u are given by

$$\begin{aligned}A_u &= \sum_{t=1}^n \mathbb{I}(C_t>0,D_t>0)\tilde{\mathrm{P}}(\eta_t > \frac{Q_u(\hat\beta_1-\beta_1)}{C_t}) \wedge \tilde{\mathrm{P}}(\eta_t > \frac{Q_u(\hat\beta_2-\beta_2)}{D_t}),\\ B_u &= \sum_{t=1}^n \mathbb{I}(C_t<0,D_t<0)\tilde{\mathrm{P}}(\eta_t < \frac{Q_u(\hat\beta_1-\beta_1)}{C_t}) \wedge \tilde{\mathrm{P}}(\eta_t < \frac{Q_u(\hat\beta_2-\beta_2)}{D_t}).\end{aligned}$$

We will now handle the term involving A_u. By Lemma 3 and regular variation of η_t we have as $u \to 1$,

$$\begin{aligned}\tilde{\mathrm{P}}(\eta_t > \frac{Q_u(\hat\beta_1-\beta_1)}{C_t}) &\sim \frac{p(1-u)C_t^\alpha}{np\mathbb{E}(\mathbb{I}(C_t>0)C_t^\alpha)+nq\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^\alpha)} && \text{if } C_t>0,\\ \tilde{\mathrm{P}}(\eta_t > \frac{Q_u(\hat\beta_2-\beta_2)}{D_t}) &\sim \frac{p(1-u)D_t^\alpha}{np\mathbb{E}(\mathbb{I}(D_t>0)D_t^\alpha)+nq\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^\alpha)} && \text{if } D_t>0.\end{aligned}$$

The indicator functions in the expression of A_u allow us to only consider $C_t > 0$ and $D_t > 0$. Furthermore, note that the limit of the minimum of two convergent sequences is equal to the minimum of the two limits of the sequences. Consequently, we have for the term involving A_u

$$\lim_{u\to 1}\frac{A_u}{1-u}=\sum_{t=1}^{n}\frac{p\mathbb{I}(C_t>0,D_t>0)C_t^{\alpha}}{np\mathbb{E}(\mathbb{I}(C_t>0)C_t^{\alpha})+nq\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^{\alpha})}\wedge\frac{p\mathbb{I}(C_t>0,D_t>0)D_t^{\alpha}}{np\mathbb{E}(\mathbb{I}(D_t>0)D_t^{\alpha})+nq\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^{\alpha})}=:A.$$

Similarly, we have for the term involving B_u

$$\lim_{u\to 1}\frac{B_u}{1-u}=\sum_{t=1}^{n}\frac{q\mathbb{I}(C_t<0,D_t<0)(-C_t)^{\alpha}}{np\mathbb{E}(\mathbb{I}(C_t>0)C_t^{\alpha})+nq\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^{\alpha})}\wedge\frac{q\mathbb{I}(C_t<0,D_t<0)(-D_t)^{\alpha}}{np\mathbb{E}(\mathbb{I}(D_t>0)D_t^{\alpha})+nq\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^{\alpha})}=:B.$$

By relation (37) we have that

$$\lim_{u\to 1}\frac{\tilde{\mathrm{P}}(\hat{\beta}_1-\beta_1>Q_u(\hat{\beta}_1-\beta_1),\ \hat{\beta}_2-\beta_2>Q_u(\hat{\beta}_2-\beta_2))}{1-u}=A+B.$$

To obtain the unconditional tail dependence we need to take the expectation with respect to X_1, X_2 over the conditional tail dependence. It follows that the unconditional tail dependence between the OLS estimators is given by $\mathbb{E}(A+B)$. Since $(X_{1,t}, X_{2,t})$ is an i.i.d. random vector we obtain

$$\mathbb{E}(A)=\mathbb{E}\left(\frac{p\mathbb{I}(C_t>0,D_t>0)C_t^{\alpha}}{p\mathbb{E}(\mathbb{I}(C_t>0)C_t^{\alpha})+q\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^{\alpha})}\bigwedge\frac{p\mathbb{I}(C_t>0,D_t>0)D_t^{\alpha}}{p\mathbb{E}(\mathbb{I}(D_t>0)D_t^{\alpha})+q\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^{\alpha})}\right)=:\mathbb{E}(A_1\wedge A_2).$$

Similarly, we have

$$\mathbb{E}(B)=\mathbb{E}\left(\frac{q\mathbb{I}(C_t<0,D_t<0)(-C_t)^{\alpha}}{p\mathbb{E}(\mathbb{I}(C_t>0)C_t^{\alpha})+q\mathbb{E}(\mathbb{I}(C_t<0)(-C_t)^{\alpha})}\bigwedge\frac{q\mathbb{I}(C_t<0,D_t<0)(-D_t)^{\alpha}}{p\mathbb{E}(\mathbb{I}(D_t>0)D_t^{\alpha})+q\mathbb{E}(\mathbb{I}(D_t<0)(-D_t)^{\alpha})}\right)=:\mathbb{E}(B_1\wedge B_2).$$

By the law of total expectation we have

$$\mathbb{E}(A)=\mathbb{E}(A_1|A_1<A_2)\mathrm{P}(A_1<A_2)+\mathbb{E}(A_2|A_1>A_2)\mathrm{P}(A_1>A_2), \tag{38}$$

$$\mathbb{E}(B)=\mathbb{E}(B_1|B_1<B_2)\mathrm{P}(B_1<B_2)+\mathbb{E}(B_2|B_1>B_2)\mathrm{P}(B_1>B_2). \tag{39}$$

Note that the event $A_1>A_2$ is equivalent to the event $C_t>\kappa D_t$ where κ is the positive number given in

(36). Likewise, the event $B_1 > B_2$ is equivalent to the event $C_t < \kappa D_t$. Therefore, we can simplify (38) and (39) to get (35). $\square$

Following Theorem 5, we observe that the tail dependence between the two OLS estimators is related to the tail exponent α and the dependence across the regressors. In other words, the tail exponent α and the dependence across the regressors are two potential determinants of the tail dependence measure. The relation between α and $\lambda_{\hat{\beta}_1,\hat{\beta}_2}$ is difficult to be analyzed. However, we can conclude from (35) that the tail dependence between the two OLS estimators is stronger if $\mathrm{P}(C_t > 0, D_t > 0)$ is higher, ceteris paribus. Since C_t and D_t are the stochastic weights in the two OLS estimators, a higher $\mathrm{P}(C_t > 0, D_t > 0)$ implies that the two estimators are more likely to be exposed to the same heavy tailed error term in the same direction. Still, we do not have an analytical way to evaluate how the dependence among the regressors is related to $P(C_t > 0, D_t > 0)$. Therefore, instead of further analyzing the relations between the tail dependence measures and its potential determinants analytically, we will resort to a simulation study in Section 4 to illustrate the discussion above.

3.4 Generalizing on to the Multiple Regression Model

We can generalize the tail dependence analysis to a regression model with k regressors in two ways. Firstly, we can calculate the tail dependence between any two OLS estimators even if there are more than two regressors. Secondly, we can calculate a tail dependence measure that accounts for tail dependence among k OLS estimators.

We define the model with k regressors as

$$Y = X\beta + \eta, \tag{40}$$

where X is now an $n \times k$ matrix and β is an $k \times 1$ vector. Similar to Section 3.1, we have that (η_t) is an i.i.d error sequence of random variables making up the vector η. $(X_{1,t}, ..., X_{k,t})$ is an i.i.d sequence of k-dimensional random vectors containing explanatory variables, independent of the error sequence (η_t)

for $t = 1, ..., n$. The OLS estimator for β is

$$\hat{\beta} = (X'X)^{-1}X'Y = \beta + W'\eta$$

where W' is the $k \times n$ matrix given by $W' = (X'X)^{-1}X'$. Similar to the regression model with two explanatory variables, we have that the OLS estimator for the i-th regression coefficient is a linear combination of the errors since $\hat{\beta}_i = \beta_i + W_i'\eta$, where W_i' is the i-th row of W'. To obtain the tail dependence between the i-th and j-th OLS estimator, $\lambda_{\hat{\beta}_i,\hat{\beta}_j}$, we can simply denote $C_t = W_{i,t}$ and $D_t = W_{j,t}$ in (35) because the derivation of Theorem 5 only requires the weights to be identically distributed.

Next, we consider a multivariate tail dependence measure. That is, the probability that one of the OLS estimators exceeds its high quantile conditional on all other estimators exceeding their corresponding high quantiles. Mathematically, it is defined

$$\lambda_{\hat{\beta}_1|\hat{\beta}_2 ...,\hat{\beta}_k} =: \lim_{u\to 1} \mathrm{P}(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1)|\, \hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_k - \beta_k)).$$

In order to calculate this limiting conditional probability, we write it in terms of two joint probabilities. Define

$$\lambda_{\hat{\beta}_i,...,\hat{\beta}_k} =: \lim_{u\to 1} \frac{\mathrm{P}(\hat{\beta}_i - \beta_i > Q_u(\hat{\beta}_i - \beta_i), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_k - \beta_k))}{1-u}, \quad \text{for any } i < k.$$

Consequently, we have

$$\lambda_{\hat{\beta}_1|\hat{\beta}_2,...,\hat{\beta}_k} = \lim_{u\to 1} \frac{\mathrm{P}(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_k - \beta_k))}{\mathrm{P}(\hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_k - \beta_k))} = \frac{\lambda_{\hat{\beta}_1,...,\hat{\beta}_k}}{\lambda_{\hat{\beta}_2,...,\hat{\beta}_k}}.$$

To calculate $\lambda_{\hat{\beta}_1|\hat{\beta}_2,...,\hat{\beta}_k}$, we need to deal with the limits $\lambda_{\hat{\beta}_1,...,\hat{\beta}_k}$ and $\lambda_{\hat{\beta}_2,...,\hat{\beta}_k}$. Fortunately, these limits are similar to the limit appearing in the bivariate tail dependence measure in Section 3.3. By using Corollary 3 and mimicking the proof of Theorem 5 we have the following result.

Theorem 6. *Assume that there exists a regression model with k regressors as*

$$Y = X\beta + \eta,$$

where (η_t) is an i.i.d. error sequence of regularly varying random variables satisfying $p, q > 0$, $(X_{1,t}, ..., X_{k,t})$ is an i.i.d. sequence of k-dimensional random vectors containing the explanatory variables, independent of the error sequence (η_t) for $t = 1, ..., n$. In addition, assume that $\lambda_{\hat{\beta}_2,...,\hat{\beta}_k} \neq 0$. Then, the multivariate tail dependence among the k OLS estimators is given by

$$\lambda_{\hat{\beta}_1|\hat{\beta}_2,...,\hat{\beta}_k} = \frac{\lambda_{\hat{\beta}_1,...,\hat{\beta}_k}}{\lambda_{\hat{\beta}_2,...,\hat{\beta}_k}},$$

where $\lambda_{\hat{\beta}_z,...,\hat{\beta}_k}$ is given by

$$\begin{aligned}\lambda_{\hat{\beta}_z,...,\hat{\beta}_k} =& \sum_{j=z}^{k} \frac{p\mathbb{E}(\mathbb{I}(W_{z,t} > 0, ..., W_{k,t} > 0)W_{j,t}^{\alpha})}{p\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + q\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \mathrm{P}(\bigvee_{i=z}^{k} c_i W_{i,t} = c_j W_{j,t}) \\ &+ \sum_{j=z}^{k} \frac{q\mathbb{E}(\mathbb{I}(W_{z,t} < 0, ..., W_{k,t} < 0)(-W_{j,t})^{\alpha})}{p\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + q\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \mathrm{P}(\bigwedge_{i=z}^{k} c_i W_{i,t} = c_j W_{j,t}), \quad \textit{for } z = 1, 2,\end{aligned} \tag{41}$$

and c_i is the positive number given by

$$c_i = \frac{1}{p\mathbb{E}(\mathbb{I}(W_{i,t} > 0)W_{i,t}^{\alpha}) + q\mathbb{E}(\mathbb{I}(W_{i,t} < 0)(-W_{i,t})^{\alpha})}, \qquad \textit{for } i = 1, ..., k. \tag{42}$$

Proof of Theorem 6. Without loss of generality, we assume that $z = 1$ and only prove the result for $\lambda_{\hat{\beta}_1,...,\hat{\beta}_k}$. Denote $\bar{\mathrm{P}}(...)$ as $\mathrm{P}(...|X)$. By taking $x_i = Q_u(\hat{\beta}_i - \beta_i)$ in Corollary 3 we obtain as $u \to 1$,

$$\frac{\bar{\mathrm{P}}(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_2 - \beta_2))}{1 - u} \sim \frac{A_u}{1 - u} + \frac{B_u}{1 - u}, \tag{43}$$

where A_u and B_u are given by

$$A_u = \sum_{t=1}^{n} \mathbb{I}(W_{1,t} > 0, ..., W_{k,t} > 0) \bigwedge_{j=1}^{k} \bar{\mathrm{P}}(\eta_t > \frac{Q_u(\hat{\beta}_j - \beta_j)}{W_{j,t}}),$$

$$B_u = \sum_{t=1}^{n} \mathbb{I}(W_{1,t} < 0, ..., W_{k,t} < 0) \bigwedge_{j=1}^{k} \bar{\mathrm{P}}(\eta_t < \frac{Q_u(\hat{\beta}_j - \beta_j)}{W_{j,t}}),$$

We will now handle the term involving A_u. By Lemma 3 and regular variation of η_t we have as $u \to 1$,

$$\bar{\mathrm{P}}(\eta_t > \frac{Q_u(\hat{\beta}_j - \beta_j)}{W_{j,t}}) \sim \frac{p(1-u)W_{j,t}^{\alpha}}{np\mathbb{E}(\mathbb{I}(W_{j,t} > 0)C_t^{\alpha}) + nq\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \qquad \text{if } W_{j,t} > 0,$$
$$\bar{\mathrm{P}}(\eta_t < \frac{Q_u(\hat{\beta}_j - \beta_j)}{W_{j,t}}) \sim \frac{q(1-u)(-W_{j,t})^{\alpha}}{np\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + nq\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \qquad \text{if } W_{j,t} < 0.$$

The indicator functions in the expression A_u allow us to only consider the cases above. Furthermore, note that the limit of the minimum of two convergent sequences is equal to the minimum of the two limits of the sequences. Consequently, we have for the term involving A_u

$$\lim_{u \to 1} \frac{A_u}{1-u} = \sum_{t=1}^{n} \mathbb{I}(W_{1,t} > 0, ..., W_{k,t} > 0) \bigwedge_{j=1}^{k} \left(\frac{pW_{j,t}^{\alpha}}{np\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + nq\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \right) =: A.$$

Similarly, we have for the term involving B_u

$$\lim_{u \to 1} \frac{A_u}{1-u} = \sum_{t=1}^{n} \mathbb{I}(W_{1,t} < 0, ..., W_{k,t} < 0) \bigwedge_{j=1}^{k} \left(\frac{q(-W_{j,t})^{\alpha}}{np\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + nq\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \right) =: B.$$

By relation (43) we have that

$$\lim_{u \to 1} \frac{\bar{\mathrm{P}}(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1), ..., \hat{\beta}_k - \beta_k > Q_u(\hat{\beta}_k - \beta_k))}{1-u} = A + B.$$

To obtain the 'unconditional limit' we need to take the expectation with respect to $X_1, ..., X_k$. It follows that the unconditional limit is given by $\mathbb{E}(A + B)$. Since $(X_{1,t}, .., X_{k,t})$ is an i.i.d. random vector we obtain

$$\mathbb{E}(A) = \mathbb{E} \left(\bigwedge_{j=1}^{k} \frac{\mathbb{I}(W_{1,t} > 0, ..., W_{k,t} > 0)pW_{j,t}^{\alpha}}{p\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + q\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})} \right) =: \mathbb{E}(\bigwedge_{j=1}^{k} A_j).$$

Similarly, we have

$$\mathbb{E}(B) = \mathbb{E}\left(\bigwedge_{j=1}^{k} \frac{\mathbb{I}(W_{1,t} < 0, ..., W_{k,t} < 0)q(-W_{j,t})^{\alpha}}{p\mathbb{E}(\mathbb{I}(W_{j,t} > 0)W_{j,t}^{\alpha}) + q\mathbb{E}(\mathbb{I}(W_{j,t} < 0)(-W_{j,t})^{\alpha})}\right) =: \mathbb{E}(\bigwedge_{j=1}^{k} B_j).$$

By the law of total expectation we have

$$\mathbb{E}(A) = \sum_{j=1}^{k} \mathbb{E}(A_j | \bigwedge_{i=1}^{k} A_i \ = A_j)\mathrm{P}(\bigwedge_{i=1}^{k} A_i \ = A_j), \tag{44}$$

$$\mathbb{E}(B) = \sum_{j=1}^{k} \mathbb{E}(B_j | \bigwedge_{i=1}^{k} B_i \ = B_j)\mathrm{P}(\bigwedge_{i=1}^{k} B_i \ = B_j). \tag{45}$$

Note that the event $\bigwedge_{i=1}^{k} A_i = A_j$ is equivalent to the event $\bigwedge_{i=1}^{k} c_i W_{i,t} = c_j W_{j,t}$, where c_i is given in (42). Likewise, the event $\bigwedge_{i=1}^{k} B_i = B_j$ is equivalent to the event $\bigvee_{i=1}^{k} c_i W_{i,t} = c_j W_{j,t}$. Therefore, we can simplify (44) and (45) to get (41). $\square$

4 Simulation Study

In this section we perform a simulation study in order to verify the result in (35) and to examine the potential determinants that influence the tail dependence between the OLS estimators. For this purpose, we consider the simple regression model

$$Y_t = X_{1,t} + X_{2,t} + \eta_t, \qquad \text{for } t = 1, 2, ..., n, \tag{46}$$

where we assume that $(X_{1,t}, X_{2,t})$ is i.i.d bivariate normal with mean zero, both standard deviations are equal to 1/5 and the correlation coefficient is equal to ρ. In addition, we assume that η_t is i.i.d. Student-t with mean zero, unit variance, α degrees of freedom and independent of $(X_{1,t}, X_{2,t})$. In this model we have $\beta_1 = \beta_2 = 1$. Under these conditions, (46) is a specific version of the model (6) because the Student-t distribution is regularly varying. Specifically, the Student-t distribution with α degrees of freedom satisfies Definition 3 with tail exponent α and $p = q = 0.5$ due to its symmetry.

We start with showing the scatter plot of the OLS estimators under the model (46). To demonstrate the impact of the heavy-tailed error terms, we provide a parallel scatter plot of the OLS estimators under an identical model but with the exception that the errors follow standard normal distributions. From the two plots in Figure 1 we observe that both the marginal and dependence behaviour of the OLS estimators under heavy-tailed errors are different from that under normally distributed errors. The OLS estimators inherit the marginal heavy-tailed behaviour of the errors in the sense that extreme values are more likely to occur. For example, the observed range for both estimators under heavy-tailed errors is roughly two times greater than under normal errors. In addition, the dependence behaviour of the OLS estimators is different from that under normally distributed errors: extreme values are more concentrated along the diagonals, indicating the presence of tail dependence.

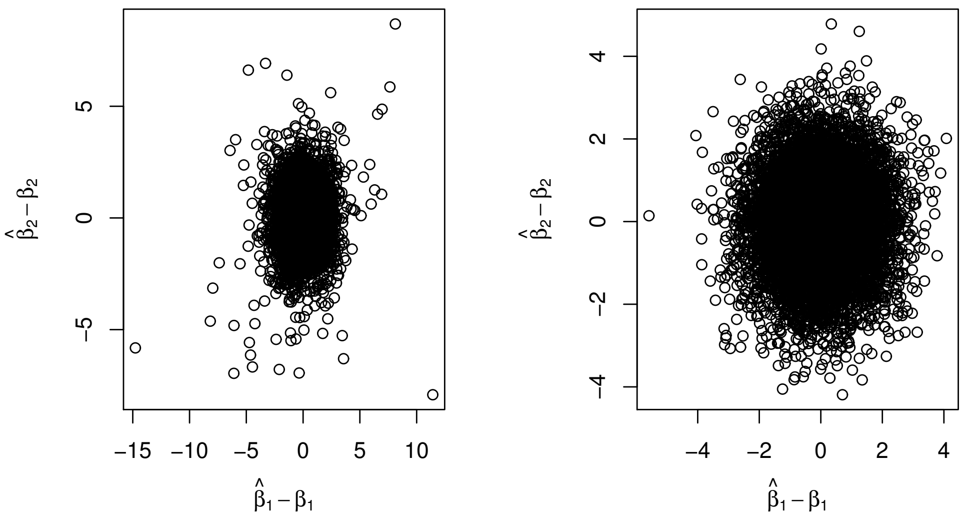


Figure 1: Scatter plots of $\hat{\beta}_1 - \beta_1$ against $\hat{\beta}_2 - \beta_2$, obtained by 10 000 runs of model (46) using $n = 25$ and $\rho = 0$ under rescaled Student-t(3) (left) and standard normally distributed errors (right).

Subsequently, we evaluate the tail dependence between the OLS estimators by two methods. Firstly, we provide an estimate using the result in Theorem (5). Secondly, we use a non-parametric estimator using multivariate EVT.

Under the conditions specified above we can simplify the tail dependence to

$$\lambda_{\hat{\beta}_1,\hat{\beta}_2} = \frac{\mathbb{E}(\mathbb{I}(C_t > 0, D_t > 0)C_t^{\alpha}|C_t < D_t)}{\mathbb{E}(\mathbb{I}(C_t > 0)C_t^{\alpha})}, \tag{47}$$

because C_t and D_t are symmetrically and identically distributed. Based on the formula in (47), we can estimate the tail dependence by its empirical analogue as

$$\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}(n,m) = \frac{n * m}{\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}(C_{ij}>0)C_{ij}^{\alpha}} \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}(C_{ij}>0,D_{ij}>0)\mathbb{I}(C_{ij}<D_{ij})C_{ij}^{\alpha}}{\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{I}(C_{ij}<D_{ij})},$$

where m is the number of simulation runs of model (46) with n observations. Consequently, the estimator $\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}$ pools the n pairs $(C_{1j},D_{1j})...(C_{nj},D_{nj})$ from each of the m simulation runs to calculate the tail dependence. Note that each of the m simulation runs also gives us a pair $(\hat{\beta}_{1,j},\hat{\beta}_{2,j})$. Therefore, we can also simply count the instances that the OLS estimators simultaneously exceed high quantiles using the estimator

$$I_{\hat{\beta}_1,\hat{\beta}_2}(m,k) = \frac{\sum_{j=1}^{m}\mathbb{I}(\hat{\beta}_{1,j}-\beta_1>\hat{\beta}_1^{(m-k)},\ \hat{\beta}_{2,j}-\beta_2>\hat{\beta}_2^{(m-k)})}{k},$$

where $\hat{\beta}_i^{(m-k)}$ is the order statistic of the m estimates $\hat{\beta}_{i,1},...,\hat{\beta}_{i,m}$ such that only k estimates are larger. Note that $I_{\hat{\beta}_1,\hat{\beta}_2}$ is expected to be a less accurate estimator than $\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}$ because its speed of convergence is$\sqrt{k}$.

Table 1: Pairs of tail dependence estimates $(I_{\hat{\beta}_1,\hat{\beta}_2},\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2})$ by taking $m = 1\,000\,000$ and $k = 1000$, for different values of α, ρ and n.

$\boldsymbol{\rho = 0}$	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$	$\boldsymbol{\rho = 0.3}$	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
$n = 25$	0.18, 0.18	0.09, 0.12	0.04, 0.08	0.03, 0.05		0.09, 0.10	0.05, 0.05	0.02, 0.03	0.01, 0.02
$n = 50$	0.19, 0.18	0.11, 0.12	0.03, 0.08	0.01, 0,05		0.08, 0.10	0.03, 0.05	0.00, 0.03	0.00, 0.02
$n = 100$	0.17, 0.18	0.09, 0.12	0.02, 0.08	0.01, 0.05		0.09, 0.10	0.04, 0.05	0.01, 0.03	0.00, 0.02
$n = 200$	0.17, 0.18	0.07, 0.12	0.02, 0.08	0.00, 0.05		0.10, 0.10	0.03, 0.05	0.01, 0.03	0.00, 0.02
$\boldsymbol{\rho = -0.3}$					$\boldsymbol{\rho = -0.9}$				
$n = 25$	0.27, 0.29	0.16, 0.21	0.11, 0.16	0.07, 0.12		0.70, 0.72	0.68, 0.67	0.57, 0.63	0.55, 0.59
$n = 50$	0.30, 0.29	0.18, 0.22	0.08, 0.16	0.05, 0.12		0.70, 0.72	0.63, 0.67	0.55, 0.63	0.50, 0.60
$n = 100$	0.28, 0.29	0.16, 0.22	0.07, 0.16	0.04, 0.12		0.72, 0.72	0.62, 0.67	0.53, 0.63	0.48, 0.59
$n = 200$	0.27, 0.29	0.15, 0.22	0.05, 0.16	0.04, 0.12		0.69, 0.72	0.61, 0.67	0.48, 0.63	0.48, 0.59

Table 1 presents values of the estimators $I_{\hat{\beta}_1,\hat{\beta}_2}$ and $\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}$ for different levels of the tail exponent α, the correlation between the regressors, ρ, and the number of observations, n. In absolute terms, we observe that the estimators coincide for all combinations of α, ρ and n. Therefore, we can conclude that the result in Theorem 5 is in line with the simulation results. We also observe that some differences occur between the two estimators $I_{\hat{\beta}_1,\hat{\beta}_2}$ and $\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}$ for higher values of α: for example, for $\alpha = 4$, $\rho = 0$ and $n \geq 50$, the two estimators differ in excess of 50% . This might be a consequence of the fixed choice of m

and k. Notice that as the tail of the regression errors becomes lighter, we need a higher m and k with a lower k/m ratio to get an equally accurate estimate of the tail dependence with $I_{\hat{\beta}_1,\hat{\beta}_2}$.

In addition, we observe from Table 1 that $\hat{\lambda}_{\hat{\beta}_1,\hat{\beta}_2}$ and $I_{\hat{\beta}_1,\hat{\beta}_2}$ decrease as α increases. Thus, tail dependence between the OLS estimators increases as the tail of the regression errors becomes heavier. As α increases, the results obtained from the Student-t distribution with α degrees of freedom become increasingly similar to that from the normal distribution. Table 1 also illustrates that lower values of ρ lead to an increase in the tail dependence between the OLS estimators. This effect occurs because $\mathrm{P}(C_t > 0, D_t > 0)$ increases as ρ decreases.

5 The Impact on the F-statistic

In this section, we first discuss the robustness of the F-test for joint tests on regression coefficients to different distributional assumptions on the regression errors. In particular, we illustrate by means of a simulation experiment that the F-test is robust to the assumption of regularly varying regression errors — despite the tail dependence between the OLS estimators that arises from this assumption. In addition, we show that the two main components of the F-statistic, the fitted sum of squares (FSS) and residual sum of squares (RSS), are tail dependent under regularly varying regression errors.

5.1 Literature Review

Under the classical OLS assumptions, the F-statistic corresponding to the null hypothesis $\beta_i = 0$, $\forall i \in \mathrm{I}$ in model (40) follows an F distribution for any non-empty index set $\mathrm{I} \subset \{1, ..., k\}$. Without loss of generality, here and in subsequent subsections, we consider the null hypothesis $\beta = 0$ such that

$$\mathrm{F} = \frac{\eta'\,\Sigma\,\eta}{\eta'(\mathrm{I} - \Sigma)\eta}\frac{n-k}{k} \sim \mathrm{F}(k, n-k), \tag{48}$$

where $\Sigma = X(X'X)^{-1}X'$. The numerator and denominator of the F-distribution are independent chi-square distributed random variables divided by their corresponding degrees of freedom. A random variable has a chi-square distribution with k degrees of freedom if it is the sum of k squared standard-normally distributed variables. The latter statement can be reconciled with (48) by noting that Σ and I are of rank k and n respectively. In regression analysis, the F-statistic is calculated by means of the intuitively more appealing form

$$F = \frac{(X\hat{\beta})'(X\hat{\beta})}{(Y - X\hat{\beta})'(Y - X\hat{\beta})} \frac{n-k}{k},$$

where $(X\hat{\beta})'(X\hat{\beta})$ and $(Y - X\hat{\beta})'(Y - X\hat{\beta})$ are the fitted sum of squares (FSS) and the residual sum of squares (RSS) respectively. We then reject the null hypothesis $\beta = 0$ if the FSS is too large relative to the RSS.

The study by Zellner (1976) is one of the first to document the robustness of the F-test. He studies the linear multiple regression model with Student-t distributed errors and finds that the F-test is still valid under Student-t distributed errors. This result follows from the Gaussian scale mixture representation of a Student-t random variable. The Gaussian scale mixture representation of a Student-t random variable with v degrees of freedom is given by $h^{-1/2}\eta_t$, where η_t is standard normally distributed and the scalar h is such that vh has a chi-square distribution with v degrees of freedom. Breusch et al. (1997) uses this representation to show that any scale-free function of Student-t distributed errors has the same distribution as under normally distributed errors. For example, if we let $\eta = h^{-1/2}\eta$ in (48), we see that the scalar h drops out, which implies that the F-test is still valid under Student-t distributed errors.

Robustness of the F-test against Student-t distributed errors suggests that the F-test may be unaffected by the presence of heavy tails. In fact, Qin & Wan (2004) shows that the F-test is valid under the class of elliptically symmetric distributed errors. This class includes the heavy-tailed multivariate Student-t and the symmetric multivariate stable distributions for example.

5.2 Some Simulation Results

The previous subsection illustrated that the F-test is robust to different distributional assumptions on the errors, which in part subsume the semi-parametric assumption of regularly varying tails. Here, we perform a simulation experiment to confirm the robustness of the F-test and to examine the dependence between the FSS and RSS.

Consider the regression model (46) of Section 4 under the null hypothesis that $\beta_1 = \beta_2 = 0$,

$$Y_t = \eta_t, \qquad \text{for } t = 1, ..., n. \tag{49}$$

We investigate the tail of the F-statistic corresponding to the null hypothesis $\beta_1 = \beta_2 = 0$, under Student-t distributed errors with 1 degree of freedom (or Cauchy errors) and standard normally distributed errors. In addition, we consider an error distribution derived from a Generalized Pareto distributed random variable, G, with shape parameter (α) equal to 1 and location parameter equal to 0. Its CDF is given by $F(x) = 0.4\text{P}(G < x) + 0.6\text{P}(G > -x)$. Unlike the Cauchy and normally distributed errors, the latter distribution is not part of the class of elliptically symmetric distributions. Note that since the F-statistic is a scale-free function of η_t, we do not have to ensure the existence of the second moment of the error distributions for rescaling purposes.

Table 2: Quantiles of the F-statistic under different error distributions, obtained by $m = 1\,000\,000$ simulations of model (49) and calculating the corresponding m F-statistics, with $n = 12$ and $\rho = 0$.

	Error Type			
F-Quantile	Generalized Pareto	Cauchy	Normal	**Theoretical $\mathbf{F(2,10)}$**
90%	2.92	2.92	2.93	**2.92**
95%	4.10	4.09	4.11	**4.10**
99%	7.58	7.53	7.57	**7.56**
99.9%	14.88	14.85	15.03	**14.91**

Table 2 shows the quantiles of the F-statistic under the different error distributions. As expected, the quantiles of the F-statistic under Cauchy and normally distributed errors correspond closely to the theoretical quantiles of the F-distribution. Furthermore, the F-test also appears to remain valid when the regression errors are generated with the Generalized Pareto errors.

In the light of (48), regularly varying errors make the marginal behaviour of the FSS and RSS more extreme under the null hypothesis. Since the F-test is still correct under the heavy-tailed Cauchy and Generalized Pareto distributions, this implies that the FSS and RSS can no longer be independent. Otherwise, larger values of the FSS would have mapped into larger values of F-statistic, resulting in higher quantiles than the ones observed in Table 2. This change in dependency is reflected in Figure 2, which shows that large values of the RSS and FSS are strongly positively dependent under Cauchy distributed errors, but not under normally distributed errors. The positive dependence between the RSS and FSS effectively prohibits the ratio FSS/RSS, and in turn the F-statistic, from becoming larger

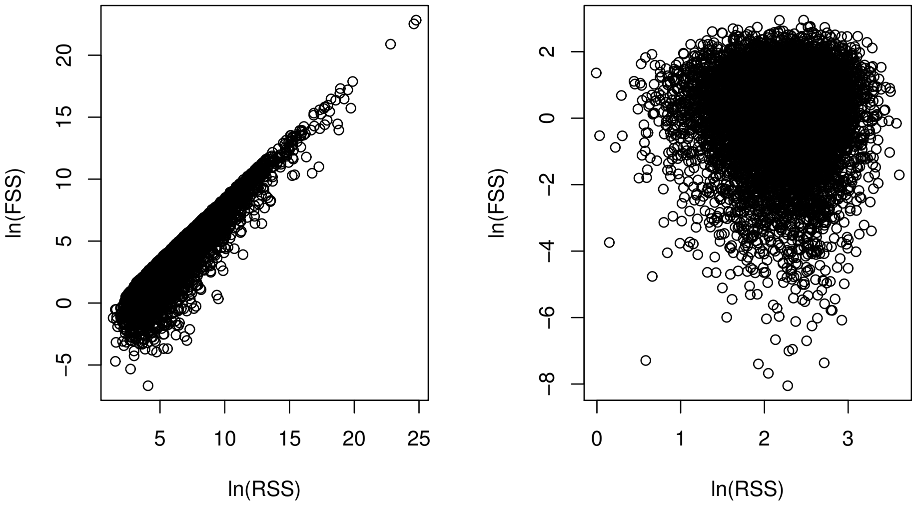


Figure 2: Logarithmic scatter plots of FSS against RSS under Cauchy (left) and normally distributed errors (right), obtained by 10 000 runs of model (49), with $n = 12$ and $\rho = 0$.

more often than under normally distributed errors. Furthermore, the diagonal pattern in the scatter plot suggests that the RSS and FSS are tail dependent. A proof of this conjecture will be given in the next subsection.

5.3 Tail Dependence between the FSS and RSS

From (48), we define the tail dependence between the RSS and FSS as

$$\lambda_{\mathrm{FSS,RSS}} = \lim_{u\to 1} \frac{\mathrm{P}(\eta'\ \Sigma\ \eta > Q_u(\eta'\ \Sigma\ \eta),\ \eta'\ (\mathrm{I}-\Sigma)\ \eta > Q_u(\eta'\ (\mathrm{I}-\Sigma)\ \eta))}{1-u}. \tag{50}$$

Equivalently, we can write

$$\lambda_{\mathrm{FSS,RSS}} = \mathbb{E}\left(\lim_{u\to 1} \frac{\bar{\mathrm{P}}(AT \in B_u)}{1-u}\right), \tag{51}$$

where $B_u = (Q_u(\eta'\ \Sigma\ \eta),\ \infty]\times(Q_u(\eta'\ (\mathrm{I}-\Sigma)\ \eta),\ \infty]$, T is a $\frac{n(n+1)}{2} \times 1$ random vector $T = (T_0,\ T_1)' = (\eta_1^2, \eta_2^2, ..., \eta_n^2,\ \eta_1\eta_2, ..., \eta_{n-1}\eta_n)'$ and A is a $2 \times \frac{n(n+1)}{2}$ matrix $A = (A_0\ A_1)$ such that

$$A_0T_0' = (\sum_{t=1}^{n} L_{tt}\eta_t^2,\ \sum_{t=1}^{n}(1-L_{tt})\eta_t^2)', \qquad A_1T_1' = (2\sum_{t<j} L_{tj}\eta_t\eta_j,\ -2\sum_{t<j} L_{tj}\eta_t\eta_j)',$$

where the random variables L_{tj} are the elements of Σ in the t-th row and j-th column. In regression analysis, the diagonal element L_{tt} reflects the "leverage" of observation t and satisfies $0 \leq L_{tt} \leq 1$.

To deal with the quadratic combinations of the errors, we use that the cross-products $\eta_t\eta_j$ are regularly varying with tail index α (Embrechts & Goldie, 1980), whereas the squared terms η_t^2 are regularly varying with tail index $\alpha/2$. Accordingly, we show that the contribution of A_1T_1 to the joint conditional probability $\bar{\mathrm{P}}(AT \in B)$ in (50) is negligible as $u \to 1$. The result then follows by applying Theorem 5.

Theorem 7. *Assume there exists a regression model with k regressors as*

$$Y_t = X_{1,t}\beta_1 + ... + X_{k,t}\beta_k + \eta_t, \qquad \textit{for } \ t = 1, ..., n,$$

where (η_t) is an i.i.d. error sequence of regularly varying random variables satisfying $p, q > 0$, $(X_{1,t}, ..., X_{k,t})$ is an i.i.d. sequence of k-dimensional random vectors containing the explanatory variables, independent of the error sequence (η_t). Then, under the null hypothesis $\beta_1 = ... = \beta_k = 0$, the tail dependence between the FSS and RSS is given by

$$\lambda_{FSS,RSS} = \frac{\mathbb{E}(L_{tt}^{\alpha/2} | L_{tt} < \gamma(1-L_{tt}))}{\mathbb{E}(L_{tt}^{\alpha/2})}\mathrm{P}(L_{tt} < \gamma(1-L_{tt})) + \frac{\mathbb{E}((1-L_{tt})^{\alpha/2} | L_{tt} > \gamma(1-L_{tt}))}{\mathbb{E}((1-L_{tt})^{\alpha/2})}\mathrm{P}(L_{tt} > \gamma(1-L_{tt})),$$
$$\textit{where } \gamma \textit{ is the positive scalar } \ \gamma = \Big(\mathbb{E}(L_{tt}^{\alpha/2})/\mathbb{E}((1-L_{tt})^{\alpha/2})\Big)^{2/\alpha}. \tag{52}$$

Proof of Theorem 7. Recall that $T = (T_0,\ T_1)' = (\eta_1^2, \eta_2^2, ..., \eta_n^2,\ \eta_1\eta_2, ..., \eta_{n-1}\eta_n)'$. We will prove that T is multivariate regularly varying and that its limiting measure μ only assigns mass to the first n dimensions of T. Consequently, this implies that AT is multivariate regularly varying and that its limiting measure only assigns mass to A_0T_0. This allows us to ignore all cross-products $\eta_t\eta_j$ in (51).

Denote the vector $\boldsymbol{x} = (x_1, ..., x_n)' \in \bar{\mathbb{R}}_+^n \setminus \{\boldsymbol{0}\}$ and the vector $\boldsymbol{z} = (z_1, ..., z_{n(n-1)/2})' \in \bar{\mathbb{R}}^{n(n-1)/2} \setminus \{-\boldsymbol{\infty}\}$. First, we will show that for all sets of the form $B = ([\boldsymbol{0}, \boldsymbol{x}] \times [-\boldsymbol{\infty}, \boldsymbol{z}])^{\mathsf{c}}$, T satisfies

$$\lim_{y\to\infty} \frac{\mathrm{P}(T \in yB)}{\mathrm{P}(\eta_t^2 > y)} = \mu(B), \tag{53}$$

where the measure μ only concentrates its mass on the first n dimensions as

$$\mu\left(([\mathbf{0},\boldsymbol{x}]\times[-\boldsymbol{\infty},\boldsymbol{z}])^{\mathrm{c}}\right)=\mu\left([\mathbf{0},\boldsymbol{x}]^{\mathrm{c}}\times\bar{\mathbb{R}}^{n(n-1)/2}\right),\quad \text{for all values of the vector } \boldsymbol{z}.$$

We will prove the limit in (53) by providing upper and lower bounds of the probability $\mathrm{P}(T\in yB)$. By excluding the possibility that the cross-products $\eta_t\eta_j$ exceed high quantiles, we have a lower bound for $\mathrm{P}(T\in yB)$ as

$$\mathrm{P}(\eta_1^2>yx_1\cup...\cup\eta_n^2>yx_n\cup\ \eta_1\eta_2>yz_1\cup...\cup\eta_{n-1}\eta_n>yz_{n(n-1)/2})\geq\mathrm{P}(\eta_1^2>yx_1\cup...\cup\eta_n^2>yx_n). \tag{54}$$

By the independence of $\eta_1^2,...,\eta_n^2$ we obtain as $y\to\infty$,

$$\mathrm{P}(\eta_1^2>yx_1\cup...\cup\eta_n^2>yx_n)\sim\sum_{t=1}^{n}\mathrm{P}(\eta_t^2>yx_t). \tag{55}$$

Upon dividing (54) by $\mathrm{P}(\eta_t^2>y)$, we have by (55) and regular variation of η_t^2,

$$\liminf_{y\to\infty}\frac{\mathrm{P}(\eta_1^2>yx_1\cup...\cup\eta_n^2>yx_n\cup\ \eta_1\eta_2>yz_1\cup...\cup\eta_{n-1}\eta_n>yz_{n(n-1)/2})}{\mathrm{P}(\eta_t^2>y)}\geq\sum_{t=1}^{n}x_t^{-\alpha/2}. \tag{56}$$

Next, by excluding the possibility that any subset of random variables in T jointly exceeds their corresponding high quantiles, we obtain an upper bound for $\mathrm{P}(T\in yB)$ as

$$\mathrm{P}(\eta_1^2>yx_1\cup...\cup\eta_n^2>yx_n\cup\ \eta_1\eta_2>yz_1\cup...\cup\eta_{n-1}\eta_n>yz_{n(n-1)/2})\leq\sum_{t=1}^{n}\mathrm{P}(\eta_t^2>yx_t)+\sum_{t<j}\mathrm{P}(\eta_t\eta_j>yz_{t+j-2}). \tag{57}$$

Note that η_t^2 has a lower tail index than any cross-product $\eta_i\eta_j$. Therefore, we obtain

$$\lim_{y\to\infty}\frac{\mathrm{P}(\eta_i\eta_j>yz_{i+j-2})}{\mathrm{P}(\eta_t^2>y)}=\lim_{y\to\infty}\frac{y^{-\alpha/2}cL_{ij}(yz_{i+j-2})}{L_t(y)}=0,\ \text{for any } i\neq j, \tag{58}$$

because c is a positive constant and $cL_{ij}(yz_{i+j-2})/L_t(y)$ is a slowly varying function. Upon dividing (57) by $\mathrm{P}(\eta_t^2 > y)$, we have by (58) and regular variation of η_t^2

$$\limsup_{y\to\infty} \frac{\mathrm{P}\left(\eta_1^2 > yx_1 \cup ... \cup \eta_n^2 > yx_n \cup\ \eta_1\eta_2 > yz_1 \cup ... \cup \eta_{n-1}\eta_n > yz_{n(n-1)/2}\right)}{\mathrm{P}(\eta_t^2 > y)} \leq \sum_{t=1}^{n} x_t^{-\alpha/2}. \quad (59)$$

Combining the upper and lower bounds in (56) and (59) gives us

$$\lim_{y\to\infty} \frac{\mathrm{P}(T \in yB)}{\mathrm{P}(\eta_t^2 > y)} = \sum_{t=1}^{n} x_t^{-\alpha/2}.$$

Hence, we have proved that the relation in (53) holds for all sets of the form $([\mathbf{0}, \boldsymbol{x}] \times [-\boldsymbol{\infty}, \boldsymbol{z}])^{\mathsf{c}}$, with the limit measure satisfying $\mu\left(([\mathbf{0}, \boldsymbol{x}] \times [-\boldsymbol{\infty}, \boldsymbol{z}])^{\mathsf{c}}\right) = \mu\left([\mathbf{0}, \boldsymbol{x}]^{\mathsf{c}} \times \bar{\mathbb{R}}^{n(n-1)/2}\right)$ for all values of the vector $\boldsymbol{z}$. Note that the sigma-algebra generated by the complements of the rectangles

$$\left([\mathbf{0}, \boldsymbol{x}] \times [-\boldsymbol{\infty}, \boldsymbol{z}] : \boldsymbol{x} \in \bar{\mathbb{R}}_+^n \setminus \{\mathbf{0}\}, \boldsymbol{z} \in \mathbb{R}^{n(n-1)/2} \setminus \{-\boldsymbol{\infty}\}\right),$$

contains all Borel sets $B \subset \bar{\mathbb{R}}_+^n \times \bar{\mathbb{R}}^{n(n-1)/2}$. That is, for all Borel sets $B \subset \bar{\mathbb{R}}_+^n \times \bar{\mathbb{R}}^{n(n-1)/2}$,

$$\lim_{y\to\infty} \frac{\mathrm{P}(T \in yB)}{\mathrm{P}(\eta_t^2 > y)} = \mu(B) = \lim_{y\to\infty} \frac{\mathrm{P}(T_0 \in yB_0)}{\mathrm{P}(\eta_t^2 > y)},$$

where B_0 is the projection of B onto the first n dimensions. Hence, T is multivariate regularly varying with limit measure μ. Consequently, we can calculate the tail dependence as

$$\lambda_{\mathrm{FSS,RSS}} = \mathbb{E}\left(\lim_{u\to 1} \frac{\bar{\mathrm{P}}(\sum_{i=1}^n L_{ii}\eta_i^2 > Q_u(FSS),\ \sum_{i=1}^n (1 - L_{ii})\eta_i^2 > Q_u(RSS))}{1-u}\right).$$

By denoting $C_t = L_{tt}$, $D_t = 1 - L_{tt}$ and taking $\eta_t = \eta_t^2$ with tail index $\alpha/2$ in Theorem 5, we obtain the result in (52). □

Table 3 displays the simulated tail dependence between the RSS and FSS under a similar setup as in Section 4. We observe in Table 3 that the tail dependence between the FSS and RSS decreases as α increases. In addition, the two estimators $\hat{\lambda}_{\mathrm{FSS,RSS}}$ and $I_{\mathrm{FSS,RSS}}$ coincide closely except for the largest value of α. The most striking difference between Table 1 and Table 3 is that the tail dependence between the FSS and RSS does not dependent on the correlation between the regressors, ρ. Furthermore, the tail dependence between the FSS and RSS is substantially stronger than the tail dependence between the

Table 3: Tail dependence estimates between the FSS and RSS by taking $m = 10\,000\,000$ and $k = 1000$, for different values of α and ρ

$n = 25,\ \rho = 0$

	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
$\hat{\lambda}_{\text{FSS,RSS}}$	0.62	0.49	0.39	0.31
$I_{\text{FSS,RSS}}$	0.63	0.48	0.36	0.23

$n = 25,\ \alpha = 3$

	$\rho = -0.9$	$\rho = -0.3$	$\rho = 0.3$	$\rho = 0.9$
$\hat{\lambda}_{\text{FSS,RSS}}$	0.49	0.49	0.49	0.49
$I_{\text{FSS,RSS}}$	0.48	0.47	0.49	0.47

OLS estimators in the base case $\rho = 0$. This can partly be explained by the fact that the quadratic forms of the FSS and RSS are more heavy-tailed, with tail index $\alpha/2$, and by the fact that the weights L_{ii} and $(1 - L_{ii})$ are always non-negative.

6 Conclusion

This thesis has studied the joint tail behaviour of the OLS estimators under heavy-tailed regression errors. We have shown that the OLS estimators are tail dependent under the class of regularly varying error distributions in a multiple regression model. Consequently, the OLS estimators are more likely to simultaneously attain their corresponding extremes than in a standard regression model with normally distributed errors. However, one can be reassured that ambiguity about the heavy-tailed nature of the regression errors can not materialize negatively in the context of joint tests on the regression coefficients: applying OLS to heavy-tailed data does not invalidate the F-test for the regression coefficients, even though the FSS and RSS are tail dependent.

In the simulation study, we demonstrated an estimator obtained from multivariate EVT and an estimator derived from the theory in Section 3 for obtaining the tail dependence between the OLS estimators. The former estimator is non-parametric and relies on running the regression m times, while effectively using only a small fraction of the corresponding m OLS estimates. The latter estimator is semi-parametric, exploits the parametric setup of the regression and consequently can be applied to a single regression. In practice, without having multiple realizations of the OLS estimator from a given regression model, one can only use the semi-parametric estimator to infer the tail dependence between the OLS estimators. The asymptotic properties of this estimator are left for future research.

References

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1989). *Regular variation* (Vol. 27). Cambridge university press.

Blattberg, R., & Sargent, T. (1971). Regression with non-gaussian stable disturbances: Some sampling results. *Econometrica*, 501–510.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, *10*(2), 323–331.

Breusch, T., Robertson, J., & Welsh, A. (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica*, *51*(3), 269–286.

Donoho, D. L., & Huber, P. J. (1983). The notion of breakdown point. *A festschrift for Erich L. Lehmann*, *157184*.

Embrechts, P., & Goldie, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society*, *29*(2), 243–256.

Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis. *The Journal of Business*, *36*(4), 420–429.

Feller, W. (1971). *Introduction to probability theory and its applications* (Vol. ii). John Wiley & Sons.

Gopikrishnan, P., Meyer, M., Amaral, L. N., & Stanley, H. E. (1998). Inverse cubic law for the distribution of stock price variations. *The European Physical Journal B-Condensed Matter and Complex Systems*, *3*(2), 139–140.

He, X., Jurečková, J., Koenker, R., & Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica*, 1195–1214.

Heij, C., De Boer, P., Franses, P. H., Kloek, T., Van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics.* OUP Oxford.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools.* Princeton university press.

Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory.* Eindhoven University of Technology.

Mikosch, T., & de Vries, C. G. (2013). Heavy tails of OLS. *Journal of Econometrics*, *172*(2), 205–221.

Phani Kumar, V. (2008). *Indexes fall hard on bloody friday.* Retrieved from `http://www.marketwatch.com`

Qin, H., & Wan, A. T. (2004). On the properties of the t-and f-ratios in linear regressions with nonnormal errors. *Econometric Theory*, *20*(4), 690–700.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer Science & Business Media.

van Oordt, M. (2013). On extreme events in banking and finance, PhD Thesis No. 41644 , Tinbergen Institute.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, *71*(354), 400–405.