TAIL DEPENDENCE OF OLS

JOCHEM OORSCHOT
Erasmus University Rotterdam
Tinbergen Institute

CHEN ZHOU
Erasmus University Rotterdam
De Nederlandsche Bank
Tinbergen Institute

This paper shows that if the errors in a multiple regression model are heavy-tailed, the ordinary least squares (OLS) estimators for the regression coefficients are tail-dependent. The tail dependence arises, because the OLS estimators are stochastic linear combinations of heavy-tailed random variables. Moreover, tail dependence also exists between the fitted sum of squares (FSS) and the residual sum of squares (RSS), because they are stochastic quadratic combinations of heavy-tailed random variables.

1. INTRODUCTION

Financial data are often heavy-tailed. For instance, Jansen and de Vries (1991) investigate the tail behavior of stock returns and find tail index estimates consistent with a finite variance, but with infinite higher moments. The heavy-tailed nature of financial data can distort statistical analysis, especially in small samples. Mikosch and de Vries (2013) show that in an ordinary least squares (OLS) regression—an econometric method often applied to financial data—the estimator for the regression coefficient is heavy-tailed if the errors are heavy-tailed. As a follow-up study, we explore the joint tail behavior of the OLS estimators. In particular, we establish that the OLS estimators in a multiple regression model with heavy-tailed errors are tail-dependent.

This study contributes to the literature of regression analysis with heavy-tailed data. There are two streams of literature dealing with heavy-tails in regression: one stream aims at finding more robust, or more efficient, estimation methods than OLS, and the other aims at characterizing the behavior of the OLS estimator(s) when the errors are heavy-tailed.

The first stream of literature addresses the concern of efficiency loss of the OLS estimator when applied to severely heavy-tailed data with infinite variance. Blattberg and Sargent (1971) examine the performance of the minimum sum of absolute errors (MSAE) estimator and a class of best linear unbiased estimators

in a regression model with stable Paretian distributed errors. They find that the turning point in efficiency lies at a tail index of approximately 1.5: for lower values of the tail index, the MSAE estimator performs considerably better than the OLS estimator, while the OLS estimator performs only slightly better for higher values of the tail index. Others have studied the OLS estimator in terms of robustness. For example, He et al. (1990) investigate the robustness of regression estimators by means of their "tail performance." They show that when the errors follow heavy-tailed distributions, the tail performance of the OLS estimator does not depend on the sample size and is effectively as worse as with a single observation only. By contrast, the MSAE, or least absolute deviations, estimator and least median of squares estimators have considerably better tail performance and are therefore also not expected to exhibit tail dependence.

The second stream of literature aims at describing the (tail) behavior of the OLS estimators under heavy-tailed errors. Mikosch and de Vries (2013) is the only study to provide small sample analytical results on the distribution of the OLS estimator when the errors follow heavy-tailed distributions. Specifically, they find that the OLS estimator is heavy-tailed in the case of a simple linear regression model with additive or multiplicative errors. Consequently, approximations prescribed by the central limit theorem do not suffice in small samples. Our study fits into the second stream of literature. Compared to Mikosch and de Vries (2013), we go beyond the simple regression model and study the tail dependence among the OLS estimators in a multiple regression model with heavy-tailed errors. Furthermore, we study the tail dependence between the regression fitted sum of squares (FSS) and the residual sum of squares (RSS).

To establish the theoretical result in this study, we express the OLS estimators for the regression coefficients as stochastically weighted sums of the errors. van Oordt (2013) shows that positive deterministic linear combinations of positive heavy-tailed random variables exhibit tail dependence. We extend this result in three ways: (i) allowing for real-valued weights, (ii) allowing for stochastic weights, and (iii) allowing for real-valued heavy-tailed random variables, of which the second extension is the most technically involved one. The tail dependence between the OLS estimators then follows from this extended result. Furthermore, the FSS and the RSS are stochastically weighted sums of squares and cross products of the errors. Consequently, their tail dependence can be established in a similar way.

In Section 2, we will define the multiple regression model and present the main result on the tail dependence between the OLS estimators and that between the FSS and the RSS. Section 3 illustrates the theoretical results by means of a simulation study. Proofs are deferred to Section 4.

2. THEORY

2.1. Model Setup

Consider the multiple regression model

$$Y_t = X_{1,t}\beta_1 + \cdots + X_{k,t}\beta_k + \eta_t, \quad \text{for } t = 1, 2, .., n, \tag{1}$$

under the conditions that (i) $\{\eta_t\}_{t=1}^n$ is an identically and independently distributed (i.i.d) sequence of random variables, and (ii) $\{(X_{1,t}, \ldots, X_{k,t})\}_{t=1}^n$ is an i.i.d. sequence of k-dimensional random vectors containing explanatory variables, independent of the error sequence $\{\eta_t\}_{t=1}^n$. We assume that η_t follows a heavy-tailed distribution, i.e., its distribution function is regularly varying. More specifically, a random variable, or its distribution function F, is said to be 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 $x \to \infty$,

$$F(-x) + 1 - F(x) \sim x^{-\alpha} L(x),$$

$$\lim_{x \to \infty} \frac{F(-x)}{F(-x) + 1 - F(x)} = q, \quad \lim_{x \to \infty} \frac{1 - F(x)}{F(-x) + 1 - F(x)} = p. \tag{2}$$

Here, a slowly varying function L is defined by having the property $L(\lambda x) \sim L(x)$ as $x \to \infty$, for any $\lambda > 0$ (see Bingham, Goldie, and Teugels, 1989). Condition (2) is referred to as a tail balance condition. Note that this assumption on F is semiparametric in the sense that it does not impose any restrictions on the moderate levels of the distribution. We do not restrict the range of α: when $\alpha < 1$, η_t does not have a finite mean.

Denote the vectors $Y = (Y_1, \ldots, Y_n)'$, $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,n})'$, $\eta = (\eta_1, \eta_2, \ldots, \eta_n)'$, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$, and the matrix $X = (X_1, X_2, \ldots, X_k)$. The OLS estimator for β is then given by $\hat{\beta} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'\eta$. Observe that the difference between the OLS estimator and the corresponding true value is a stochastic linear combination of η. For example, if $k = 2$, we can 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

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

$$D_t = \frac{X_{2,t} \sum_{j=1}^n X_{1,j}^2 - X_{1,t} \sum_{j=1}^n X_{1,j} X_{2,j}}{\sum_{j=1}^n X_{1,j}^2 \sum_{j=1}^n X_{2,j}^2 - \left(\sum_{j=1}^n X_{1,j} X_{2,j}\right)^2}. \tag{3}$$

Here, $\left\{(C_t, D_t)\right\}_{t=1}^n$ is an identically distributed, but not independent, sequence of random vectors. Consequently, we omit subscripts and write (C, D) whenever only the distribution of (C_t, D_t) is involved.

We aim at calculating the tail dependence between the OLS estimators. We define tail dependence between any two random variables Z_1 and Z_2 as

$$\lambda_{Z_1, Z_2} := \lim_{u \to 1} \mathbb{P}(Z_1 > Q_u(Z_1) | Z_2 > Q_u(Z_2)),$$

where $Q_u(Z) = \inf\{l : F_Z(l) \geq u\}$ denotes the quantile of a random variable Z at probability level u. Similarly, we define multivariate tail dependence between Z_i and the other random variables as

$$\lambda_{Z_i | Z_{-i}} := \lim_{u \to 1} \mathbb{P}\left(Z_i > Q_u(Z_i) \mid Z_j > Q_u(Z_j) \; \forall j \neq i\right),$$

where Z_{-i} is a shorthand notation for denoting all random variables $Z_1, \ldots, Z_n$ except Z_i.

2.2. Tail Dependence Between the OLS Estimators

Consider model (1) with $k = 2$. We then have the following result for the tail dependence between the OLS estimators.

THEOREM 1. *Assume that the errors $\{\eta_t\}_{t=1}^n$ in model (1), for $k = 2$, follow a heavy-tailed distribution—as defined in (2)—with $p, q > 0$ and tail index α. In addition, assume that the weights (C, D) satisfy $\mathbb{E}|C|^{\alpha+\epsilon}, \mathbb{E}|D|^{\alpha+\epsilon} < \infty$, for some $\epsilon > 0$, where (C, D) is defined as in (3). Then, the tail dependence between the OLS estimators for β_1 and β_2 is given by*

$$
\lambda_{\hat{\beta}_1, \hat{\beta}_2} = \frac{p\mathbb{E}(\mathbb{I}_{A \cap \{D > 0\}}(C^+)^\alpha) + q\mathbb{E}(\mathbb{I}_{A^c \cap \{D < 0\}}(C^-)^\alpha)}{p\mathbb{E}(C^+)^\alpha + q\mathbb{E}(C^-)^\alpha}
$$
$$
+ \frac{p\mathbb{E}(\mathbb{I}_{A^c \cap \{C > 0\}}(D^+)^\alpha) + q\mathbb{E}(\mathbb{I}_{A \cap \{C < 0\}}(D^-)^\alpha)}{p\mathbb{E}(D^+)^\alpha + q\mathbb{E}(D^-)^\alpha}, \tag{4}
$$

where $\mathbb{I}_{\{.\}}$ is the indicator function, $C^+ := \max\{C, 0\}$, $C^- := \max\{-C, 0\}$, the variables D^+ and D^- are defined similarly, and A is the event

$$
A = \left\{ \frac{|C|^\alpha}{|D|^\alpha} < \frac{p\mathbb{E}(C^+)^\alpha + q\mathbb{E}(C^-)^\alpha}{p\mathbb{E}(D^+)^\alpha + q\mathbb{E}(D^-)^\alpha} \right\}.
$$

Mathematically, Theorem 1 can be interpreted as showing that stochastically weighted sums of regularly varying random variables are tail-dependent. This also follows directly from the multivariate version of Breiman's lemma (see Proposition A.1 in Basrak, Davis, and Mikosch, 2002). Our result, however, is more specific in that we have an explicit formula relating the level of tail dependence to the distribution of the regressors. With the explicit formula, one can estimate the tail dependence using the observed regressors (see Section 3 below). Following (4), the level of tail dependence is determined by the tail index, α, and the joint distribution of the stochastic weights (C, D). We resort to a simulation study in Section 3 to illustrate the effect of these determinants.

One high-level condition required in Theorem 1 is $\mathbb{E}|C|^{\alpha+\epsilon}, \mathbb{E}|D|^{\alpha+\epsilon} < \infty$. In Proposition 1 below, we provide sufficient conditions on the distribution of the regressors which guarantee such a requirement.

PROPOSITION 1. *Assume that the following conditions are satisfied:*

1. *For $i = 1, 2$, either.*
   a. *the common marginal distribution of $X_{i,t}$ satisfies that, for some $\gamma > \frac{2\alpha}{n}$, $c > 0$, and $x_0 > 0$, such that, for all $x < x_0$,*

   $$
   \mathbb{P}(|X_{i,t}| < x) \leq cx^\gamma,
   $$

or

    b. *the common density function of $X_{i,t}$ is bounded in the neighborhood of zero and $n > 2\alpha$.*

2. *There exist some $c_1 > 0$, $\gamma_1 > 2\alpha$, and $0 < x_1 < 1$ such that, for all $0 < x < x_1$,*

$$\mathbb{P}\left(\frac{\left(\sum_{j=1}^{n} X_{2,j}X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} > 1 - x\right) \le c_1 x^{\gamma_1}.$$

Then, there exists some $\epsilon > 0$ such that $\mathbb{E}|C_t|^{\alpha+\epsilon} < \infty$ and $\mathbb{E}|D_t|^{\alpha+\epsilon} < \infty$.

We remark that the first condition in Proposition 1 is a requirement on the marginal distribution of the regressors. It is similar to the condition in Lemma 3.6 of Mikosch and de Vries (2013). Note that for financial data, condition 1(b) may not be likely to hold (see, e.g., Han, Cho, and Phillips, 2011). The second condition in Proposition 1 requires that the probability of having two linearly dependent regressors is sufficiently low. Finally, as an example, these conditions hold when the regressors $(X_{1,t}, X_{2,t})$ follow a bivariate normal distribution with correlation parameter $|\rho| < 1$ and $n > 4\alpha + 2$. Proofs for Proposition 1 and the example are postponed to Section 4.

In the general case with k regressors, we may still write $\hat{\beta}_i = \beta_i + W^{(i)}\eta$, where the vector $W^{(i)} = (W_{i,1}, \ldots, W_{i,n})$ is a function of the regressor matrix X. The weights $\left\{(W_{1,t}, W_{2,t}, \ldots, W_{k,t})\right\}_{t=1}^{n}$ form an identically distributed, but not independent, sequence of k-dimensional random vectors. Therefore, we omit the time subscripts and continue to write $(W_1, W_2, \ldots, W_k)$. Consequently, we have the following extension of Theorem 1.

THEOREM 2. *Assume that the errors $\{\eta_t\}_{t=1}^{n}$ in (1) follow a heavy-tailed distribution—as defined in (2)—with $p, q > 0$ and tail index α. In addition, assume that the weights $(W_1, W_2, \ldots, W_k)$ satisfy $\mathbb{E}(|W_i|^{\alpha+\epsilon}) < \infty$, for some $\epsilon > 0$, for $i = 1, \ldots, k$. Denote, for $z = 1, 2$,*

$$\lambda_{\hat{\beta}_z, \ldots, \hat{\beta}_k}$$
$$= \sum_{j=z}^{k} \frac{p\mathbb{E}(\mathbb{I}_{A_{j,z}} \cap \{W_z, \ldots, W_k > 0\}\left(W_j^+\right)^\alpha) + q\mathbb{E}(\mathbb{I}_{B_{j,z}} \cap \{W_z, \ldots, W_k < 0\}(W_j^-)^\alpha)}{p\mathbb{E}(W_j^+)^\alpha + q\mathbb{E}(W_j^-)^\alpha},$$

where the events $A_{j,z}$ and $B_{j,z}$ are given by

$$A_{j,z} = \left\{j = \arg\max_{z \le i \le k} \frac{p\left(W_i^+\right)^\alpha}{p\mathbb{E}\left(W_i^+\right)^\alpha + q\mathbb{E}(W_i^-)^\alpha}\right\},$$

$$B_{j,z} = \left\{j = \arg\min_{z \le i \le k} \frac{q\left(W_i^-\right)^\alpha}{p\mathbb{E}\left(W_i^+\right)^\alpha + q\mathbb{E}\left(W_i^-\right)^\alpha}\right\}.$$

If $\lambda_{\hat{\beta}_2,\dots,\hat{\beta}_k} \neq 0$, then the multivariate tail dependence between $\hat{\beta}_1$ and $\hat{\beta}_2,\dots,\hat{\beta}_k$ is given by

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

2.3. Tail Dependence Between the FSS and the RSS

In regression analysis, two important quantities are the FSS and the RSS. These quantities are often used to test the explanatory power of the regression model via the F-test. That is, to test $\beta_1 = \cdots = \beta_k = 0$ in the regression model given by (1). For example, under the assumption that the errors $\{\eta_t\}_{t=1}^n$ in (1) are normally distributed, the F-statistic is defined as

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

where the terms $\left(X\hat{\beta}\right)'\left(X\hat{\beta}\right)$ and $\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)$ are the FSS and the RSS, respectively. Under the null hypothesis $\beta_1 = \cdots = \beta_k = 0$, we can write the FSS and the RSS as

$$FSS = \eta'\Sigma\eta, \quad RSS = \eta'(I - \Sigma)\eta, \quad \text{with } \Sigma = X(X'X)^{-1}X'. \tag{5}$$

Similar to the OLS estimators, the FSS and the RSS are weighted quadratic combinations of the errors, where the weights are contained in Σ. Denote the tth diagonal element of Σ as L_{tt}. Then, $\{L_{tt}\}_{t=1}^n$ are identically distributed random variables. We omit the subscripts and use L whenever only the distribution of L_{tt} is involved. In the following theorem, we show that the FSS and the RSS are tail-dependent when the errors are heavy-tailed.

THEOREM 3. *Assume that the errors $\{\eta_t\}_{t=1}^n$ in (1) follow a heavy-tailed distribution with $p, q > 0$ and tail index α. Then, under the null hypothesis $\beta_1 = \cdots = \beta_k = 0$, the tail dependence between the FSS and the RSS is given by*

$$\lambda_{FSS,RSS} = \frac{\mathbb{E}\left(\mathbb{I}_B L^{\alpha/2}\right)}{\mathbb{E}\left(L^{\alpha/2}\right)} + \frac{\mathbb{E}\left(\mathbb{I}_{B^c}(1-L)^{\alpha/2}\right)}{\mathbb{E}\left((1-L)^{\alpha/2}\right)}$$

and B is the event

$$B = \left\{ \frac{L^{\alpha/2}}{(1-L)^{\alpha/2}} < \frac{\mathbb{E}\left(L^{\alpha/2}\right)}{\mathbb{E}\left((1-L)^{\alpha/2}\right)} \right\}.$$

We have two remarks about Theorem 3. First, for a given α, $\lambda_{FSS,RSS}$ is solely determined by the distribution of L—the distribution of a diagonal element of Σ. In regression analysis, a diagonal element of Σ is known as the leverage of an

observation and indicates to what extent the observation "determines" the OLS estimate. Note that $0 \leq L_{tt} \leq 1$, which guarantees the existence of all moments of L. Therefore, no moment condition is needed in Theorem 3.

Second, under normally distributed errors, the FSS and the RSS are independent. By contrast, we show that under heavy-tailed errors, the two are at least tail-dependent. Consequently, the F-statistic may not follow an F-distribution, if errors are heavy-tailed. Nevertheless, if the regressors are deterministic and the errors are in the class of multivariate elliptical distributions, the F-statistic still follows an F-distribution (see, e.g., Qin and Wan, 2004). Multivariate elliptical distributions include heavy-tailed distributions, such as the Cauchy distribution and the multivariate Student t distribution (recall the Gaussian scale mixture representation; see Breusch, Robertson, and Welsh, 1997), but do not include the i.i.d. case we are considering.

Although the null hypothesis in Theorem 3 appears to be specific, the same conclusion is valid for a broader null where only a subset of the coefficients are set to zero, i.e., a nested model test. Write $X = [X^{(0)}, X^{(1)}]$ and $\beta = [(\beta^{(0)})^T, (\beta^{(1)})^T]^T$, where the first term with $\cdot^{(0)}$ contains the first k_0 dimensions and the second term with $\cdot^{(1)}$ contains the other $k - k_0$ dimensions. Suppose we are testing the null that the first k_0 dimensions of the coefficients are zero, i.e., $\beta^{(0)} = 0_{k_0}$. By the Frisch–Waugh–Lovell theorem, the OLS estimator for $\beta^{(0)}$ is the same as the OLS estimator obtained when regressing $M_1 Y$ on $M_1 X^{(0)}$, where $M_1 = I - X^{(1)}((X^{(1)})^T (X^{(1)}))^{-1} (X^{(1)})^T$. Moreover, the residuals in both regressions are the same. Hence, testing $\beta^{(0)} = 0_{k_0}$ in the full model is equivalent to testing the explanatory power in the regression after eliminating the impact of $X^{(1)}$. Therefore, the same conclusion as in Theorem 3 holds.

3. SIMULATION STUDY

The main results in Section 2 show that the tail dependence between the OLS estimators depends on the tail index α and the joint distribution of the regressors. Here, we perform a simulation study to validate this result and demonstrate to what extent these determinants affect the level of tail dependence.

3.1. Bivariate Tail Dependence Between the OLS Estimators

We simulate data from the regression model

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

where $\{(X_{1,t}, X_{2,t})\}_{t=1}^{n}$ are i.i.d observations drawn from a bivariate normal distribution with mean zero, standard deviation 1/5, and correlation ρ. In addition, $\{\eta_t\}_{t=1}^{n}$ is an i.i.d. sequence of errors, independent of $\{(X_{1,t}, X_{2,t})\}_{t=1}^{n}$. For the heavy-tailed case, we draw these errors from a Student t distribution with mean zero and α degrees of freedom. In this model, we have that $\beta_1 = \beta_2 = 1$. Under these conditions, the regression model (6) satisfies the conditions of Theorems 1

and 3, because the Student t distribution with α degrees of freedom is regularly varying with tail index α and $p = q = 0.5$ due to its symmetry. To compare with the heavy-tailed case, we also draw errors from a standard normal distribution. In all simulation studies, we consider the regression model with $n \in \{25, 50, 100, 200\}$ and repeat the simulation m times for each n.

Figure 1 shows the scatter plots of the (transformed) OLS estimators under the heavy-tailed and normally distributed errors in the left and right panels, respectively. For the sake of comparison, we transform the estimates to having uniform marginal distributions using the empirical distribution functions estimated from the m runs. The top two plots show the result of $m = 2,000$ runs. We observe that the dependence structure of the OLS estimators under heavy-tailed errors is different from that under normally distributed errors: observations are more concentrated along the outer diagonals, suggesting the presence of tail dependence in the case of heavy-tailed errors and the potential tail independence under normally distributed errors. In order to further demonstrate the tail dependence structure in the right tails, in the two bottom plots, we show the results of $m = 10,000$ runs with zooming in to the joint tail regions when both transformed estimators are above 0.95. Under (tail) independence, one would expect to find $10,000 * 0.05 * 0.05 = 25$ such pairs in this area. We observe that under the heavy-tailed errors, there are 152 such pairs in this region, about five times more than the expected number under (tail) independence. However, under the normally distributed errors, there are 32 such pairs, close to the expected number.

Subsequently, we evaluate the tail dependence between the OLS estimators by two methods. The first method involves estimating the tail dependence from simulated regressors using the result in Theorem 1. Note that under the distribution of $\{\eta_t\}_{t=1}^n$ specified above, we simplify the tail dependence in Theorem 1 to

$$\lambda_{\hat{\beta}_1, \hat{\beta}_2} = \frac{\mathbb{E}\left(\mathbb{I}_{\{D>0\}}\left(C^+\right)^{\alpha} \mid C < D\right)}{\mathbb{E}\left(C^+\right)^{\alpha}}, \tag{7}$$

because C and D are symmetrically and identically distributed. Based on (7), we can estimate the tail dependence by its empirical analog using the observed regressors as

$$\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2} = \frac{n}{\sum_{i=1}^n \left(C_i^+\right)^{\alpha}} \frac{\sum_{i=1}^n \mathbb{I}_{\{D_i > 0\}} \mathbb{I}_{\{C_i < D_i\}} \left(C_i^+\right)^{\alpha}}{\sum_{i=1}^n \mathbb{I}_{\{C_i < D_i\}}}. \tag{8}$$

With the m simulation runs, we obtain m estimates of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$. From these estimates, we calculate the mean and standard deviation of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ and report it in Table 1 as the lower entry for each given set of parameters (α, n, ρ). For the case $\rho = 0$, we consider a wider range of values for α from 1 to 5. In this case, we draw the errors from the standard Student $t(\alpha)$ distribution. Notice that, for $\alpha = 1$, the errors have no finite mean, and for $\alpha = 2$, the errors have no finite variance. For the other cases where $\rho \neq 0$, we only consider $\alpha > 2$, because it is often assumed in regression models that the errors have a finite variance. In these cases, we draw the errors from
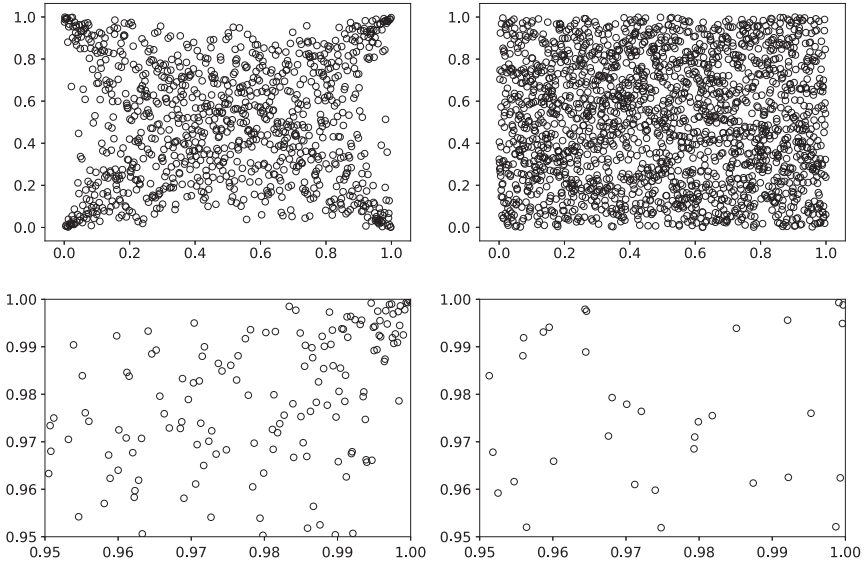
FIGURE 1. Scatter plot displaying marginally transformed $\hat{\beta}_1, \hat{\beta}_2$. *Notes:* the top two figures are obtained by transforming the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ to have marginal uniform distributions, using $m = 2,000$ runs of model (6), $n = 25$, and $\rho = 0$ with errors following the standard Student $t(1)$ distribution (left) and the standard normal distribution (right). The bottom two figures are displaying the tail regions obtained under the same conditions except that $m = 10,000$ runs have been used.

the standard Student t distribution with α degrees of freedom and unit variance. This choice facilitates a better comparison across different values of α. Note that in this simulation study, we refrain from estimating α and consider α to be known in (8). In practice, one needs to estimate α from the observations $\{Y_t\}_{t=1}^n$, which may potentially inflate the variance of the estimator.

As a second method, we use a nonparametric estimator from multivariate extreme value theory to estimate the tail dependence. Note that the m simulation runs result in m pairs of estimates $(\hat{\beta}_1, \hat{\beta}_2)$. Therefore, we can use a nonparametric estimator for the tail dependence between $\hat{\beta}_1$ and $\hat{\beta}_2$ as follows:

$$I_{\hat{\beta}_1, \hat{\beta}_2}(m, k) = \frac{\sum_{j=1}^m \mathbb{I}\left(\hat{\beta}_{1,j} > \hat{\beta}_1^{(m-k)}, \ \hat{\beta}_{2,j} > \hat{\beta}_2^{(m-k)}\right)}{k}, \tag{9}$$

where $\hat{\beta}_i^{(1)} \leq \hat{\beta}_i^{(2)} \leq \cdots \leq \hat{\beta}_i^{(m)}$ are the order statistics of the m estimates $\hat{\beta}_{i,1}, \ldots, \hat{\beta}_{i,m}$, for $i = 1, 2$, and k is an intermediate sequence such that $k/m \rightarrow 0$ as $(k, m) \rightarrow \infty$. For each given set of parameters (α, n, ρ), we report the estimate $I_{\hat{\beta}_1, \hat{\beta}_2}$ in Table 1 as the upper entry. All estimates in Table 1 are obtained by fixing $m = 1,000,000$ and $k = 1,000$.

TABLE 1. Tail dependence estimates: $I_{\hat{\beta}_1, \hat{\beta}_2}$ and the mean and standard deviation of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$

	$\rho = 0$				
	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
$n = 25$	0.28	0.18	0.09	0.04	0.03
	0.30 (0.19)	0.20 (0.21)	0.15 (0.22)	0.12 (0.23)	0.11 (0.25)
$n = 50$	0.31	0.19	0.11	0.04	0.01
	0.30 (0.13)	0.19 (0.14)	0.13 (0.14)	0.10 (0.15)	0.08 (0.16)
$n = 100$	0.32	0.17	0.09	0.02	0.01
	0.29 (0.09)	0.19 (0.10)	0.12 (0.10)	0.09 (0.10)	0.07 (0.10)
$n = 200$	0.32	0.17	0.07	0.01	0.00
	0.29 (0.07)	0.18 (0.07)	0.07 (0.12)	0.08 (0.07)	0.06 (0.07)

	$\rho = 0.9$				$\rho = 0.3$		
	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 5$		$\alpha = 2.01$	$\alpha = 3$	$\alpha = 5$
$n = 25$	0.00	0.00	0.00		0.09	0.05	0.01
	0.01 (0.01)	0.00 (0.01)	0.00 (0.00)		0.11 (0.14)	0.07 (0.13)	0.04 (0.14)
$n = 50$	0.01	0.00	0.00		0.08	0.03	0.00
	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)		0.11 (0.09)	0.06 (0.08)	0.03 (0.08)
$n = 100$	0.01	0.00	0.00		0.09	0.04	0.00
	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)		0.10 (0.06)	0.06 (0.05)	0.02 (0.04)
$n = 200$	0.00	0.00	0.00		0.10	0.03	0.00
	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)		0.10 (0.04)	0.06 (0.03)	0.02 (0.03)

	$\rho = -0.3$				$\rho = -0.9$		
	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 5$		$\alpha = 2.01$	$\alpha = 3$	$\alpha = 5$
$n = 25$	0.27	0.16	0.07		0.70	0.68	0.55
	0.32 (0.27)	0.26 (0.31)	0.21 (0.37)		0.73 (0.41)	0.70 (0.51)	0.67 (0.64)
$n = 50$	0.30	0.18	0.05		0.70	0.63	0.50
	0.30 (0.19)	0.24 (0.22)	0.17 (0.27)		0.72 (0.29)	0.69 (0.38)	0.65 (0.53)
$n = 100$	0.28	0.16	0.04		0.72	0.62	0.48
	0.30 (0.13)	0.23 (0.15)	0.15 (0.19)		0.72 (0.21)	0.68 (0.28)	0.63 (0.43)
$n = 200$	0.27	0.15	0.04		0.69	0.61	0.48
	0.30 (0.09)	0.22 (0.11)	0.14 (0.13)		0.72 (0.15)	0.67 (0.21)	0.61 (0.33)

Note: For each set of parameters (α, n, ρ), the upper entry presents an estimate of $I_{\hat{\beta}_1, \hat{\beta}_2}$ as in (9) and the lower entry presents the mean and standard deviation of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ (in parentheses) obtained from (8).

We have the following observations from the simulation. First, the result in Theorem 1 is in line with the simulations. The average of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ coincides closely with $I_{\hat{\beta}_1, \hat{\beta}_2}$ for most combinations of α, ρ, and n. Some differences occur between the two estimators $I_{\hat{\beta}_1, \hat{\beta}_2}$ and $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ for high values of α: for example, for $\alpha = 5$, $n \geq 25$, and $\rho = 0$, the average of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ and $I_{\hat{\beta}_1, \hat{\beta}_2}$ differ in excess of 50%. This result might be a consequence of the fixed choice of k in the estimator $I_{\hat{\beta}_1, \hat{\beta}_2}$.

Second, for given α and ρ, the mean of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ is generally invariant as n increases. For given n and ρ, both $I_{\hat{\beta}_1, \hat{\beta}_2}$ and $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ decrease as α increases. As α increases, the Student t distribution with α degrees of freedom becomes more similar to the normal distribution. Consequently, tail dependence between the OLS estimators decreases as the tail of the errors becomes lighter. Table 1 also illustrates that more negative correlation corresponds to higher tail dependence between the OLS estimators. This is due to the increase in $\mathbb{P}(C > 0, D > 0)$ as ρ decreases.

Finally, we remark that the standard deviation of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ can be relatively high for small n. This suggests that one should be cautious when applying the estimator in (8) for regressions with limited number of observations. In addition, the standard deviation of $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ increases as the correlation ρ decreases, but is relatively insensitive to the level of α except in the case $\rho = -0.9$.

3.2. Tail Dependence Between the FSS and the RSS

Similar to the previous subsection, we estimate the tail dependence between the FSS and the RSS from the simulated regression model. We use an estimator based on Theorem 3 and a nonparametric estimator, denoted as $\hat{\lambda}_{\text{FSS,RSS}}$ and $I_{\text{FSS,RSS}}$, respectively. These estimators are constructed in a similar way as $\hat{\lambda}_{\hat{\beta}_1, \hat{\beta}_2}$ and $I_{\hat{\beta}_1, \hat{\beta}_2}$. We omit the details.

Table 2 displays the tail dependence estimates $\hat{\lambda}_{\text{FSS,RSS}}$ and $I_{\text{FSS,RSS}}$ under model (6), with $m = 10,000,000$ and $k = 1,000$. We only present the results for $n = 25$, since results for other values of n remain qualitatively the same.[1] The average of $\hat{\lambda}_{\text{FSS,RSS}}$ and $I_{\text{FSS,RSS}}$ coincide closely except for the highest value of α. Furthermore, we observe that the tail dependence between the FSS and the RSS decreases as α increases. In contrast with the results in Table 1, the level of tail dependence between the FSS and the RSS does not depend on the correlation between the regressors, ρ. In addition, the tail dependence between the FSS and the RSS is stronger than the tail dependence between the OLS estimators in the base case $\rho = 0$. This can partly be explained by the fact that the quadratic forms of the FSS and the RSS are more heavy-tailed, with tail index $\alpha/2$, and by the fact that the weights L_{tt} and $(1 - L_{tt})$ are always nonnegative.

---

[1]The results are available upon request.

TABLE 2. Tail dependence estimates between the FSS and the RSS: $I_{\text{FSS, RSS}}$ and $\hat{\lambda}_{\text{FSS, RSS}}$

	$\rho = 0$			$\alpha = 3$			
	$\alpha = 2.01$	$\alpha = 3$	$\alpha = 5$	$\rho = 0.9$	$\rho = 0.3$	$\rho = -0.3$	$\rho = -0.9$
$n = 25$	0.63	0.48	0.23	0.48	0.47	0.49	0.47
	0.62 (0.05)	0.49 (0.06)	0.31 (0.07)	0.49 (0.06)	0.49 (0.06)	0.49 (0.06)	0.49 (0.06)

Note: For each set of parameters (α, n, ρ), the upper entry presents an estimate of $I_{\text{FSS, RSS}}$ and the lower entry presents the mean and standard deviation of $\hat{\lambda}_{\text{FSS, RSS}}$ (in parentheses).

4. PROOFS

To calculate the tail dependence between the OLS estimators, we will analyze the joint probability that the estimators exceed high thresholds. Since the estimators $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are stochastically weighted sums of regularly varying errors, we start by analyzing nonstochastic linear combinations of regularly varying random variables.

4.1. Extension of the Feller Theorem

Theorem 4 provides an approximation to the joint probability that two real-valued linear combinations of real-valued regularly varying random variables exceed high thresholds.

THEOREM 4. *Let $\eta_1, \eta_2, \ldots, \eta_n$ be real-valued regularly varying variables. Assume that they are independent. In addition, assume that c_i, d_i are real-valued coefficients, for $i = 1, 2, \ldots, n$. In addition, assume that $\frac{x}{y} \to \kappa$ as $x \to \infty$, with $\kappa > 0$. Then, it holds that, as $x \to \infty$,*

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

$$+ \sum_{i=1}^{n} \mathbb{I}\{c_i < 0, d_i < 0\} \mathbb{P}\left(\eta_i < \frac{x}{c_i} \wedge \frac{y}{d_i}\right). \quad (10)$$

The relation in (10) has a similar interpretation as the Feller theorem (see Feller, 1971, p. 278); for high x and y, the events $\{\sum_{i=1}^{n} c_i \eta_i > x\}$ and $\{\sum_{i=1}^{n} d_i \eta_i > y\}$ occur simultaneously solely due to a sufficiently extreme value of one of the $\eta_1, \ldots, \eta_n$ if their corresponding coefficients have the same sign.

As a corollary of Theorem 4, we consider the joint tail probability of k linear combinations of n real-valued regularly varying random variables.

COROLLARY 1. *Let $\eta_1, \eta_2, \ldots, \eta_n$ be real-valued regularly varying random variables. Assume that they are independent. In addition, assume that $w_{1,i}, \ldots, w_{k,i}$*

are real-valued coefficients, for $i = 1, 2, \ldots, n$. In addition, assume that $\frac{x_1}{x_j} \to \kappa_j$ as $x_1 \to \infty$, with $\kappa_j > 0$, for $j = 2, \ldots, k$. Then, it holds that, as $x_1 \to \infty$,

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

To prove Theorem 4, we first deal with n nonnegative regularly varying random variables in Lemma 1. In addition, this lemma does not require the assumption of independence. Instead, it only requires the weaker assumption that each pair of random variables are either independent or satisfying $\mathbb{P}\left(\eta_i > 0, \ \eta_j > 0\right) = 0$.

LEMMA 1. *Let $\eta_1, \eta_2, \ldots, \eta_n$ be nonnegative regularly varying variables. Assume that, for all $1 \le i \le n$, η_i is independent of $\{\eta_j : \mathbb{P}(\eta_i > 0, \eta_j > 0) > 0\}$. In addition, assume that c_i, d_i are real-valued coefficients, for $i = 1, 2, \ldots, n$. In addition, assume that $\frac{x}{y} \to \kappa$ as $x \to \infty$, with $\kappa > 0$. Then, it holds that, as $x \to \infty$,*

$$
\mathbb{P}\left(\sum_{i=1}^{n} c_i\eta_i > x, \ \sum_{i=1}^{n} d_i\eta_i > y\right) \sim \sum_{i=1}^{n} \mathbb{I}\{c_i > 0, d_i > 0\}\mathbb{P}\left(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}\right). \tag{11}
$$

If $\sum_{i=1}^{n} \mathbb{I}\{c_i > 0, d_i > 0\} = 0$, the above relation should be read as

$$
\mathbb{P}\left(\sum_{i=1}^{n} c_i\eta_i > x, \ \sum_{i=1}^{n} d_i\eta_i > y\right) = o\left(\sum_{i=1}^{n} \mathbb{P}\left(\eta_i > x\right)\right).
$$

Proof of Lemma 1. We prove Lemma 1 by mathematical induction. We start by proving the Lemma for $n = 2$. Here, we only deal with $c_1 < 0$ and $c_2, d_1, d_2 > 0$. The other cases, including the case $\sum_{i=1}^{2} \mathbb{I}\{c_i > 0, d_i > 0\} = 0$, are similar or simpler. We handle the joint probability on the left-hand side of (11) by providing its upper and lower bounds using set manipulation.

Note that, for any $0 < \delta < 1/2$, we have that

$$
\{c_1\eta_1 + c_2\eta_2 > x, \ d_1\eta_1 + d_2\eta_2 > y\} \supset \left\{\eta_1 < \frac{\delta x}{-c_1}, \ \eta_2 > (1+\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2}\right)\right\},
$$
$$
\tag{12}
$$

$$
\{c_1\eta_1 + c_2\eta_2 > x, \ d_1\eta_1 + d_2\eta_2 > y\} \subset \left\{\eta_2 > (1-\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2}\right)\right\}
$$

$$\cup \left\{ \eta_1 > \frac{\delta y}{d_1}, \, \eta_2 > \frac{x}{c_2} \right\}. \tag{13}$$

From (12), we get that the lower bound on the joint probability is

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

If $\mathbb{P}\left(\eta_1 > 0, \eta_2 > 0 \right) = 0$, we have that

$$\mathbb{P}\left(\eta_1 < \frac{\delta x}{-c_1}, \eta_2 > (1+\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2} \right) \right) = \mathbb{P}\left(\eta_2 > (1+\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2} \right) \right),$$

and if η_1 and η_2 are independent, we have

$$\mathbb{P}\left(\eta_1 < \frac{\delta x}{-c_1}, \eta_2 > (1+\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2} \right) \right)$$
$$= \mathbb{P}\left(\eta_1 < \frac{\delta x}{-c_1} \right) \mathbb{P}\left(\eta_2 > (1+\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2} \right) \right).$$

Note that $\mathbb{P}\left(\eta_1 < \frac{\delta x}{-c_1} \right) \to 1$ as $x \to \infty$. By first letting $x \to \infty$ and then letting $\delta \to 0$, we have that

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

From (13), we get that the upper bound on the joint probability is

$$\mathbb{P}\left(c_1 \eta_1 + c_2 \eta_2 > x, \, d_1 \eta_1 + d_2 \eta_2 > y \right)$$
$$\leq \mathbb{P}\left(\eta_2 > (1-\delta)\left(\frac{x}{c_2} \vee \frac{y}{d_2} \right) \right) + \mathbb{P}\left(\eta_1 > \frac{\delta y}{d_1}, \, \eta_2 > \frac{x}{c_2} \right). \tag{14}$$

If η_1 and η_2 are independent, the last term on the right-hand side of (14) is then $\mathbb{P}\left(\eta_1 > \frac{\delta y}{d_1} \right) \mathbb{P}\left(\eta_2 > \frac{x}{c_2} \right)$, which is of higher order than the first term as $x \to \infty$. If $\mathbb{P}\left(\eta_1 > 0, \, \eta_2 > 0 \right) = 0$, then the last term on the right-hand side of (14) is zero. Therefore, by letting $\delta \to 0$, we have that

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

Combining the upper and lower bounds gives the result for $n = 2$.

Assuming that the lemma holds for $n-1$ nonnegative regularly varying random variables, we prove that it also holds for n. First, we deal with the case $\sum_{i=1}^{n} \mathbb{I}\{c_i > 0, d_i > 0\} \neq 0$. Then, there is at least one positive pair (c_i, d_i). Without loss of generality (w.l.o.g.), assume that $c_n, d_n > 0$. Note that, for any $0 < \delta < 1/2$, we

have that

$$\left\{ \sum_{i=1}^{n} c_i \eta_i > x, \ \sum_{i=1}^{n} d_i \eta_i > y \right\} \supset \left\{ \sum_{i=1}^{n-1} c_i \eta_i > x, \ \sum_{i=1}^{n-1} d_i \eta_i > y \right\}$$
$$\cup \left\{ \eta_n > (1+\delta) \left(\frac{x}{c_n} \vee \frac{y}{d_n} \right), \ \sum_{i=1}^{n-1} c_i \eta_i > -\delta x, \ \sum_{i=1}^{n-1} d_i \eta_i > -\delta y \right\}, \tag{15}$$

and

$$\left\{ \sum_{i=1}^{n} c_i \eta_i > x, \ \sum_{i=1}^{n} d_i \eta_i > y \right\}$$
$$\subset \left\{ \sum_{i=1}^{n-1} c_i \eta_i > x(1-\delta), \ \sum_{i=1}^{n-1} d_i \eta_i > y(1-\delta) \right\} \cup \left\{ \eta_n > (1-\delta) \left(\frac{x}{c_n} \vee \frac{y}{d_n} \right) \right\}$$
$$\cup \left(\left\{ \sum_{i=1}^{n-1} c_i \eta_i > \delta x \ \text{or} \ \sum_{i=1}^{n-1} d_i \eta_i > \delta y \right\} \cap \left\{ \eta_n > \delta \left(\frac{x}{c_n} \wedge \frac{y}{d_n} \right) \right\} \right). \tag{16}$$

By (15), we have that

$$\mathbb{P} \left(\sum_{i=1}^{n} c_i \eta_i > x, \ \sum_{i=1}^{n} d_i \eta_i > y \right)$$
$$\geq \mathbb{P} \left(\sum_{i=1}^{n-1} c_i \eta_i > x, \ \sum_{i=1}^{n-1} d_i \eta_i > y \right)$$
$$+ \mathbb{P} \left(\eta_n > (1+\delta) \left(\frac{x}{c_n} \vee \frac{y}{d_n} \right), \ \sum_{i=1}^{n-1} c_i \eta_i > -\delta x, \ \sum_{i=1}^{n-1} d_i \eta_i > -\delta y \right)$$
$$- \mathbb{P} \left(\sum_{i=1}^{n-1} c_i \eta_i > x, \ \sum_{i=1}^{n-1} d_i \eta_i > y, \eta_n > (1+\delta) \left(\frac{x}{c_n} \vee \frac{y}{d_n} \right) \right)$$
$$:= \mathbb{I}_1 + \mathbb{I}_2 - \mathbb{I}_3.$$

If $\sum_{i=1}^{n-1} \mathbb{I}\{c_i > 0, d_i > 0\} > 0$, we apply the induction hypothesis on $\mathbb{I}_1$ to get that

$$\liminf_{x \to \infty} \frac{\mathbb{I}_1}{\sum_{i=1}^{n-1} \mathbb{I}\{c_i > 0, d_i > 0\} \mathbb{P} \left(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i} \right)} = 1.$$

Denote $I_{ind} = \{\eta_i : \mathbb{P}(\eta_n > 0, \eta_i > 0) > 0\}$. Since η_n is independent of $(\eta_i, i \in I_{ind})$ and $\mathbb{P}(\eta_n > 0, \eta_i > 0) = 0$, for $i \in I_{ind}^c$, it follows that

$$\mathbb{I}_2 = \mathbb{P} \left(\eta_n > (1+\delta) \left(\frac{x}{c_n} \vee \frac{y}{d_n} \right), \ \sum_{i \in I_{ind}} c_i \eta_i > -\delta x, \ \sum_{i \in I_{ind}} d_i \eta_i > -\delta y \right)$$

$$= \mathbb{P}\left(\eta_n > (1+\delta)\left(\frac{x}{c_n} \vee \frac{y}{d_n}\right)\right)\mathbb{P}\left(\sum_{i \in I_{ind}}^{n-1} c_i\eta_i > -\delta x, \sum_{i \in I_{ind}}^{n-1} d_i\eta_i > -\delta y\right),$$

which leads to

$$\liminf_{x \to \infty} \frac{\mathbb{I}_2}{\mathbb{P}\left(\eta_n > \left(\frac{x}{c_n} \vee \frac{y}{d_n}\right)\right)} \geq (1+\delta)^{-\alpha}.$$

The term $\mathbb{I}_3$ is of a higher order than $\sum_{i=1}^n \mathbb{P}(\eta_i > x)$: the proof follows a similar way as handling $\mathbb{I}_2$ by splitting variables into the set I_{ind} and its complement. Combining all three terms and further taking $\delta \to 0$, we obtain the lower bound that

$$\liminf_{x \to \infty} \frac{\mathbb{P}\left(\sum_{i=1}^n c_i\eta_i > x, \sum_{i=1}^n d_i\eta_i > y\right)}{\sum_{i=1}^n \mathbb{I}\{c_i > 0, d_i > 0\}\mathbb{P}\left(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}\right)} \geq 1.$$

For the upper bound, we get from (16) that

$$\mathbb{P}\left(\sum_{i=1}^n c_i\eta_i > x, \sum_{i=1}^n d_i\eta_i > y\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{n-1} c_i\eta_i > x(1-\delta), \sum_{i=1}^{n-1} d_i\eta_i > y(1-\delta)\right)$$

$$+ \mathbb{P}\left(\eta_n > (1-\delta)\left(\frac{x}{c_n} \vee \frac{y}{d_n}\right)\right)$$

$$+ \mathbb{P}\left(\left\{\sum_{i=1}^{n-1} c_i\eta_i > \delta x \text{ or } \sum_{i=1}^{n-1} d_i\eta_i > \delta y\right\} \cap \left\{\eta_n > \delta\left(\frac{x}{c_n} \wedge \frac{y}{d_n}\right)\right\}\right)$$

$$:= \mathbb{L}_1 + \mathbb{L}_2 + \mathbb{L}_3.$$

The proof follows a similar way as handling the lower bound: the term $\mathbb{L}_1$ can be handled using the induction hypothesis. The term $\mathbb{L}_2$ can be handled using the regular variation. Finally, similar to $\mathbb{I}_3$, the $\mathbb{L}_3$ term is of a higher order than $\sum_{i=1}^n \mathbb{P}(\eta_i > x)$. Combining the three terms together and further taking $\delta \to 0$, we obtain the upper bound as

$$\limsup_{x \to \infty} \frac{\mathbb{P}\left(\sum_{i=1}^n c_i\eta_i > x, \sum_{i=1}^n d_i\eta_i > y\right)}{\sum_{i=1}^n \mathbb{I}\{c_i > 0, d_i > 0\}\mathbb{P}\left(\eta_i > \frac{x}{c_i} \vee \frac{y}{d_i}\right)} \leq 1.$$

The combination of the lower and upper bounds yields the result for n.

If $\sum_{i=1}^{n-1} \mathbb{I}\{c_i > 0, d_i > 0\} = 0$, the proof is similar and simpler, because the terms $\mathbb{I}_1$ and $\mathbb{L}_1$ are of a higher order than $\sum_{i=1}^n \mathbb{P}(\eta_i > x)$.

Next, we handle the statement regarding the case $\sum_{i=1}^n \mathbb{I}\{c_i > 0, d_i > 0\} = 0$ by mathematical induction. The proof for $n = 2$ is trivial. Assume that the result holds for $n - 1$.

If all c_i and all d_i are nonpositive, the result for n holds trivially. Suppose at least one c_i or d_i is positive. w.l.o.g., assume that $c_n > 0$ and $d_n \leq 0$. Notice that

$$\left\{ \sum_{i=1}^{n} c_i \eta_i > x, \ \sum_{i=1}^{n} d_i \eta_i > y \right\}$$

$$\subset \left\{ \sum_{i=1}^{n-1} c_i \eta_i > x(1-\delta), \ \sum_{i=1}^{n-1} d_i \eta_i > y \right\} \cup \left\{ \eta_n > (1-\delta)\frac{x}{c_n}, \ \sum_{i=1}^{n-1} d_i \eta_i > y \right\}$$

$$\cup \left\{ \sum_{i=1}^{n-1} c_i \eta_i > \delta x, \ \eta_n > \delta \frac{x}{c_n} \right\}.$$

The statement is proved for n provided that the probability of each set on the right-hand side is of a higher order than $\sum_{i=1}^{n} \mathbb{P}(\eta_i > x)$. The first set can be handled by the induction hypothesis. The second and third sets can be handled again by splitting the variables into the set I_{ind} and its complement. ∎

To prove Theorem 4, we separate the positive and negative parts of a real-valued regularly varying random variable and "absorb" the negative part into the corresponding coefficient. The result then follows by Lemma 1.

Proof of Theorem 4. W.l.o.g., consider the case that $n = 2$ and let $\eta_i^+ = \max\{\eta_i, 0\}$ and $\eta_i^- = \max\{-\eta_i, 0\}$. In addition, let $e_i = -c_i$ and $z_i = -d_i$. By definition, it holds that $\mathbb{P}(\eta_i^+ > 0, \ \eta_i^- > 0) = 0$. In addition, by definition, it holds that

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

Since the random variables $\{\eta_1^+, \eta_1^-, \eta_2^+, \eta_2^-\}$ satisfy the conditions of Lemma 1, the theorem is proved by applying the lemma. ∎

4.2. Proofs of Theorems 1 and 2

We can apply the results in Section 4.1 to the OLS estimators by a conditioning argument. Denoting $\mathbb{P}(...|X)$ by $\tilde{\mathbb{P}}(...)$, we have that

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

In the Appendix, we show that we may interchange the limit and the expectation, which leads to

$$\lambda_{\hat{\beta}_1, \hat{\beta}_2} = \mathbb{E} \left(\lim_{u \to 1} \frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1 \right), \ \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2 \right) \right)}{1-u} \right). \quad (17)$$

Then, the "conditional tail dependence," the term within the expectation in (17), can be calculated with the aid of Theorem 4.

We start with obtaining an approximation for $\mathbb{P}\left(\eta_t > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)$—the probability which needs to be dealt with after we apply Theorem 4 to the conditional tail dependence. To obtain the required approximation, we need to deal with the unconditional marginal distribution of $\hat{\beta}_1 - \beta_1$. For that purpose, we state here Lemma 3.4 of Mikosch and de Vries (2013), which extends Breiman's result (see Proposition 3 in Breiman, 1965) to a multivariate regularly varying random vector ξ and another random vector ζ of the same dimension. We define a random vector $\xi = (\xi_1, \ldots, \xi_d)'$ to be multivariate regularly varying with index $\alpha > 0$ if there exists a Radon measure μ on $\mathbb{R}^d \setminus (-\epsilon, \epsilon)^d$, for some $\epsilon > 0$, such that for every Borel set $E \subset \mathbb{R}^d \setminus (-\epsilon, \epsilon)^d$ satisfying that $\mu(\partial E) = 0$,

$$\lim_{x \to \infty} \frac{\mathbb{P}\left(\xi \in xE\right)}{\mathbb{P}\left(|\xi| > x\right)} = \mu(E),$$

with the homogeneity condition $\mu(xE) = x^{-\alpha}\mu(E)$, see Theorem 6.1 in Resnick (2007). A special case of multivariate regular variation of ξ is when $\xi_1, \ldots, \xi_d$ are independent regularly varying random variables (Resnick, 2007, Sect. 6.5.1).

LEMMA 2. *Assume that ξ is multivariate regularly varying in $\mathbb{R}^d$ with index $\alpha > 0$ and is independent of the random vector ζ. Furthermore, it holds that $\mathbb{E}|\zeta|^{\alpha+\epsilon} < \infty$, for some $\epsilon > 0$, where $|\cdot|$ denotes any given norm on $\mathbb{R}^d$. Then, the scalar product $\psi = \xi'\zeta$ is regularly varying with index α. Moreover, if ξ has independent components, then as $x \to \infty$,*

$$P(\psi > x) \sim \mathbb{P}\left(|\xi| > x\right)\left(\sum_{i=1}^{d} r_i^+ \mathbb{E}(\zeta_i^+)^\alpha + \sum_{i=1}^{d} r_i^- \mathbb{E}(\zeta_i^-)^\alpha\right), \text{ with}$$

$$r_i^+ = \lim_{x \to \infty} \frac{\mathbb{P}(\xi_i > x)}{\mathbb{P}(|\xi| > x)}, \ r_i^- = \lim_{x \to \infty} \frac{\mathbb{P}(\xi_i < -x)}{\mathbb{P}(|\xi| > x)}.$$

We now present an approximation for $\mathbb{P}(\eta_t > Q_u(\hat{\beta}_1 - \beta_1))$ in the following lemma.

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

$$\mathbb{P}\left(\eta_t > Q_u\left(\sum_{j=1}^{n} C_j\eta_j\right)\right) \sim \frac{p(1-u)}{np\mathbb{E}\left(C_t^+\right)^\alpha + nq\mathbb{E}\left(C_t^-\right)^\alpha}, \tag{18}$$

$$\mathbb{P}\left(\eta_t < -Q_u\left(\sum_{j=1}^{n} C_j \eta_j\right)\right) \sim \frac{q(1-u)}{np\mathbb{E}\left(C_t^+\right)^\alpha + nq\mathbb{E}\left(C_t^-\right)^\alpha}. \tag{18}$$

Proof of Lemma 3. Define $|\eta|_\infty := \max_t(|\eta_t|)$. By regular variation of η_t and the tail balance condition (2), we have that, as $x \to \infty$,

$$\mathbb{P}(|\eta|_\infty > x) \sim n\mathbb{P}(|\eta_t| > x) \sim \frac{n}{p}\mathbb{P}(\eta_t > x).$$

Recall that, by definition, $\mathbb{P}\left(\sum_{t=1}^{n} C_t \eta_t > Q_u\left(\sum_{j=1}^{n} C_j \eta_j\right)\right) = 1 - u$. Equation (18) follows immediately from Lemma 2 and the tail balance condition. ∎

Using Lemma 3, we can now prove Theorem 1.

Proof of Theorem 1. We start by deriving the conditional tail dependence between the OLS estimators, which is the limit appearing in (17).

First, we show that $Q_u(\hat{\beta}_1 - \beta_1)/Q_u(\hat{\beta}_2 - \beta_2) \to \kappa > 0$ as $u \to 1$, with

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

From Lemma 3, we get that, as $u \to 1$,

$$\frac{\mathbb{P}\left(\eta > Q_u(\hat{\beta}_1 - \beta_1)\right)}{\mathbb{P}\left(\eta > Q_u(\hat{\beta}_2 - \beta_2)\right)} \to \frac{p\mathbb{E}\left(D_t^+\right)^\alpha + q\mathbb{E}\left(D_t^-\right)^\alpha}{p\mathbb{E}\left(C_t^+\right)^\alpha + q\mathbb{E}\left(C_t^-\right)^\alpha}.$$

The statement thus follows from the fact that η has a regularly varying tail.

We can then take $x = Q_u\left(\hat{\beta}_1 - \beta_1\right)$ and $y = Q_u\left(\hat{\beta}_2 - \beta_2\right)$ in Theorem 4 to obtain that, as $u \to 1$,

$$\frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1\right), \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2\right)\right)}{1-u} \sim \frac{M_u^+}{1-u} + \frac{M_u^-}{1-u}, \tag{19}$$

where M_u^+ and M_u^- are given by

$$M_u^+ = \sum_{t=1}^{n} \mathbb{I}(C_t > 0, D_t > 0)$$

$$\times \left[\tilde{\mathbb{P}}\left(\eta_t > \frac{Q_u\left(\hat{\beta}_1 - \beta_1\right)}{C_t}\right) \wedge \tilde{\mathbb{P}}\left(\eta_t > \frac{Q_u\left(\hat{\beta}_2 - \beta_2\right)}{D_t}\right)\right],$$

$$M_u^- = \sum_{t=1}^{n} \mathbb{I}(C_t < 0, D_t < 0)$$

$$\times \left[\tilde{\mathbb{P}} \left(\eta_t < \frac{Q_u \left(\hat{\beta}_1 - \beta_1 \right)}{C_t} \right) \wedge \tilde{\mathbb{P}} \left(\eta_t < \frac{Q_u \left(\hat{\beta}_2 - \beta_2 \right)}{D_t} \right) \right].$$

We first handle the term M_u^+. Note that $\tilde{\mathbb{P}}(\eta_t > Q_u(\hat{\beta}_j - \beta_j)) = \mathbb{P}(\eta_t > Q_u(\hat{\beta}_j - \beta_j))$, for $j = 1, 2$. Therefore, by Lemma 3 and the regular variation of η_t, we have, as $u \to 1$,

$$\tilde{\mathbb{P}} \left(\eta_t > \frac{Q_u \left(\hat{\beta}_1 - \beta_1 \right)}{C_t} \right) \sim \frac{p(1-u)C_t^\alpha}{np\mathbb{E} \left(C_t^+ \right)^\alpha + nq\mathbb{E} \left(C_t^- \right)^\alpha} \qquad \text{if } C_t > 0,$$

$$\tilde{\mathbb{P}} \left(\eta_t > \frac{Q_u \left(\hat{\beta}_2 - \beta_2 \right)}{D_t} \right) \sim \frac{p(1-u)D_t^\alpha}{np\mathbb{E} \left(D_t^+ \right)^\alpha + nq\mathbb{E} \left(D_t^- \right)^\alpha} \qquad \text{if } D_t > 0.$$

The indicator functions in the expression of M_u^+ allow us to consider only $C_t > 0$ 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 that

$$\lim_{u \to 1} \frac{M_u^+}{1-u} = \sum_{t=1}^n \frac{p\mathbb{I}_{\{D_t > 0\}} \left(C_t^+ \right)^\alpha}{np\mathbb{E} \left(C_t^+ \right)^\alpha + nq\mathbb{E} \left(C_t^- \right)^\alpha} \wedge \frac{p\mathbb{I}_{\{C_t > 0\}} \left(D_t^+ \right)^\alpha}{np\mathbb{E} \left(D_t^+ \right)^\alpha + nq\mathbb{E} \left(D_t^- \right)^\alpha} := M^+.$$

Similarly, we have that

$$\lim_{u \to 1} \frac{M_u^-}{1-u} = \sum_{t=1}^n \frac{q\mathbb{I}_{\{D_t < 0\}} \left(C_t^- \right)^\alpha}{np\mathbb{E} \left(C_t^+ \right)^\alpha + nq\mathbb{E} \left(C_t^{-\alpha} \right)} \wedge \frac{q\mathbb{I}_{\{C_t < 0\}} \left(D_t^- \right)^\alpha}{np\mathbb{E} \left(D_t^+ \right)^\alpha + nq\mathbb{E} \left(D_t^- \right)^\alpha} := M^-.$$

By relation (19), we have that

$$\lim_{u \to 1} \frac{\tilde{\mathbb{P}} \left(\hat{\beta}_1 - \beta_1 > Q_u \left(\hat{\beta}_1 - \beta_1 \right), \hat{\beta}_2 - \beta_2 > Q_u \left(\hat{\beta}_2 - \beta_2 \right) \right)}{1-u} = M^+ + M^-.$$

It follows that the unconditional tail dependence between the OLS estimators is given by $\mathbb{E} \left(M^+ + M^- \right)$. Since $\{(C_t, D_t)\}_{t=1}^n$ are identically distributed random vectors, we obtain that

$$\mathbb{E}(M^+) = \mathbb{E} \left(\frac{p\mathbb{I}_{\{D_t > 0\}} \left(C_t^+ \right)^\alpha}{p\mathbb{E} \left(C_t^+ \right)^\alpha + q\mathbb{E} \left(C_t^- \right)^\alpha} \wedge \frac{p\mathbb{I}_{\{C_t > 0\}} \left(D_t^+ \right)^\alpha}{p\mathbb{E} \left(D_t^+ \right)^\alpha + q\mathbb{E} \left(D_t^- \right)^\alpha} \right) := \mathbb{E} \left(M_1^+ \wedge M_2^+ \right).$$

$$\mathbb{E}(M^-) = \mathbb{E} \left(\frac{q\mathbb{I}_{\{D_t < 0\}} \left(C_t^- \right)^\alpha}{p\mathbb{E} \left(C_t^+ \right)^\alpha + q\mathbb{E} \left(C_t^- \right)^\alpha} \wedge \frac{q\mathbb{I}_{\{C_t < 0\}} \left(D_t^- \right)^\alpha}{p\mathbb{E} \left(D_t^+ \right)^\alpha + q\mathbb{E} \left(D_t^- \right)^\alpha} \right) := \mathbb{E} \left(M_1^- \wedge M_2^- \right).$$

By the law of total expectation, we obtain the right-hand side of (4), which yields the theorem by (17). ∎

The proof of Theorem 2 follows similar steps as that of Theorem 1 and is therefore omitted.

4.3. Proof of Theorem 3

Recall from (5) that under the null hypothesis $\beta = 0$, the FSS and the RSS can be written as $\eta' \Sigma \eta$ and $\eta' (I - \Sigma) \eta$, respectively, where $\Sigma = X (X'X)^{-1} X'$. Therefore, the tail dependence between the FSS and the RSS is given by

$$\lambda_{\text{FSS}, \text{RSS}} = \lim_{u \to 1} \frac{\mathbb{P}\left(\eta' \, \Sigma \eta > Q_u\left(\eta' \Sigma \eta\right), \eta' \left(I - \Sigma\right) \eta > Q_u\left(\eta' \left(I - \Sigma\right) \eta\right)\right)}{1 - u}.$$

We shall interchange the expectation and the limit. Because Σ contains bounded entries only, the justification of the interchanging can be obtained in a similar manner as in the situation of (17) (see the Appendix). Consequently, we get that

$$\lambda_{\text{FSS}, \text{RSS}} = \mathbb{E}\left(\lim_{u \to 1} \frac{\tilde{\mathbb{P}}\left(\eta' \, \Sigma \eta > Q_u\left(\eta' \Sigma \eta\right), \eta' \left(I - \Sigma\right) \eta > Q_u\left(\eta' \left(I - \Sigma\right) \eta\right)\right)}{1 - u}\right).$$

$$\textbf{(20)}$$

Note that under the $\tilde{\mathbb{P}}$ measure, the matrix Σ can be regarded as nonstochastic. In regression analysis, the elements on the diagonal of Σ, L_{tt}, reflect the "leverage" of observation Y_t and satisfy $0 \le L_{tt} \le 1$, whereas the elements on the off-diagonal, L_{tj}, satisfy $-1/2 \le L_{tj} \le 1/2$, for $t \ne j$.

In the proof, we use the fact that for the cross products $\eta_t \eta_j$ with $t \ne j$, both the left and right tails are regularly varying with tail index α (Embrechts and Goldie, 1980), whereas the squared terms η_t^2 are regularly varying with tail index $\alpha/2$.

Proof of Theorem 3. For any $x > 0$, define the following sets:

$$A_0(x) = \left\{\sum_{t=1}^{n} L_{tt} \eta_t^2 > x\right\} \text{ and } A_1(x) = \left\{\sum_{t=1}^{n} (1 - L_{tt}) \eta_t^2 > x\right\},$$

$$B_0(x) = \left\{\exists t < j, \text{ s.t. } 2 L_{tj} \eta_t \eta_j > x\right\} \text{ and } B_1(x) = \left\{\exists t < j, \text{ s.t. } - 2 L_{tj} \eta_t \eta_j > x\right\},$$

$$C_0(x) = \left\{\sum_{t=1}^{n} L_{tt} \eta_t^2 + \sum_{t<j} 2 L_{tj} \eta_t \eta_j > x\right\} \text{ and }$$

$$C_1(x) = \left\{\sum_{t=1}^{n} (1 - L_{tt}) \eta_t^2 - \sum_{t<j} 2 L_{tj} \eta_t \eta_j > x\right\}.$$

We study the probability of these sets under the measure $\tilde{\mathbb{P}}$.

First, we handle the sets $A_i(x)$, for $i = 0, 1$. From the Feller theorem, we have that

$$\lim_{x \to \infty} \frac{\tilde{\mathbb{P}}(A_0(x))}{\mathbb{P}(\eta_t^2 > x)} = \sum_{t=1}^{n} L_{tt}^{\alpha/2} \quad \text{and} \quad \lim_{x \to \infty} \frac{\tilde{\mathbb{P}}(A_1(x))}{\mathbb{P}(\eta_t^2 > x)} = \sum_{t=1}^{n} (1 - L_{tt})^{\alpha/2}. \tag{21}$$

Here, in the denominators, we use the fact that $\tilde{\mathbb{P}}(\eta_t^2 > x) = \mathbb{P}(\eta_t^2 > x)$ due to independence between the error terms and the regressors.

Next, we handle the sets $B_i(x)$, for $i = 0, 1$. Note that $B_0(x) = \bigcup_{t<j} \{2L_{tj}\eta_t\eta_j > x\}$ and $B_1(x) = \bigcup_{t<j} \{2L_{tj}\eta_t\eta_j < -x\}$. For any $t < j$, since $|\eta_t\eta_j|$ has a lower tail index than η_t^2, we get that, as $x \to \infty$, $\tilde{\mathbb{P}}(2L_{tj}|\eta_t\eta_j| > x) = o(\mathbb{P}(\eta_t^2 > x))$, which implies that, for both $i = 0, 1$,

$$\lim_{x \to \infty} \frac{\tilde{\mathbb{P}}(B_i(x))}{\mathbb{P}(\eta_t^2 > x)} = 0. \tag{22}$$

Finally, we handle the sets $C_i(x)$, for $i = 0, 1$. Note that, for any $0 < \epsilon < 1/2$,

$$A_0(x(1+\epsilon)) \bigcap B_1^c \left(\frac{\epsilon}{n(n-1)/2} x \right) \subset C_0(x) \subset A_0(x(1-\epsilon)) \bigcup B_0 \left(\frac{\epsilon}{n(n-1)/2} x \right).$$

Hence, we get the upper bound for $\tilde{\mathbb{P}}(C_0(x))$ as

$$\limsup_{x \to \infty} \frac{\tilde{\mathbb{P}}(C_0(x))}{\mathbb{P}(\eta_t^2 > x)} \le \limsup_{x \to \infty} \frac{\tilde{\mathbb{P}}(A_0(x(1-\epsilon))) + \tilde{\mathbb{P}}\left(B_0\left(\frac{\epsilon}{n(n-1)/2}x\right)\right)}{\mathbb{P}(\eta_t^2 > x)}$$

$$= (1-\epsilon)^{-\alpha/2} \sum_{t=1}^{n} L_{tt}^{\alpha/2},$$

where, in the last step, we apply (21), (22), and the fact that $\mathbb{P}(\eta_t^2 > x)$ is a regularly varying function with index $-\alpha/2$. Similarly, we have the lower bound for $\tilde{\mathbb{P}}(C_0(x))$ as

$$\liminf_{x \to \infty} \frac{\tilde{\mathbb{P}}(C_0(x))}{\mathbb{P}(\eta_t^2 > x)} \ge \limsup_{x \to \infty} \frac{\tilde{\mathbb{P}}(A_0(x(1+\epsilon))) - \tilde{\mathbb{P}}\left(B_1\left(\frac{\epsilon}{n(n-1)/2}x\right)\right)}{\mathbb{P}(\eta_t^2 > x)}$$

$$= (1+\epsilon)^{-\alpha/2} \sum_{t=1}^{n} L_{tt}^{\alpha/2}.$$

By taking $\epsilon \to 0$, the upper and lower bounds coincide, which yields the limit of $\tilde{\mathbb{P}}(C_0(x))$. A limit result for $\tilde{\mathbb{P}}(C_1(x))$ can also be obtained in a similar way. We present both limit results as follows:

$$\lim_{x \to \infty} \frac{\tilde{\mathbb{P}}(C_0(x))}{\mathbb{P}(\eta_t^2 > x)} = \sum_{t=1}^{n} L_{tt}^{\alpha/2} \quad \text{and} \quad \lim_{x \to \infty} \frac{\tilde{\mathbb{P}}(C_1(x))}{\mathbb{P}(\eta_t^2 > x)} = \sum_{t=1}^{n} (1 - L_{tt})^{\alpha/2}.$$

By taking expectation on both sides of the two limit relations, we get that

$$\lim_{x \to \infty} \frac{\mathbb{P}(C_0(x))}{\mathbb{P}(\eta_t^2 > x)} = n\mathbb{E}(L^{\alpha/2}) \text{ and } \lim_{x \to \infty} \frac{\mathbb{P}(C_1(x))}{\mathbb{P}(\eta_t^2 > x)} = n\mathbb{E}(1 - L)^{\alpha/2}.$$

Recall that $C_0(x) = \{\eta' \Sigma \eta > x\}$, from the properties of regularly varying function, we get that, as $u \to 1$,

$$\lim_{u \to 1} \frac{Q_u(\eta' \Sigma \eta)}{Q_u(\eta_t^2)} = \left(n\mathbb{E}(L^{\alpha/2})\right)^{\frac{2}{\alpha}} \text{ and } \lim_{u \to 1} \frac{Q_u(\eta'(I - \Sigma)\eta)}{Q_u(\eta_t^2)} = \left(n\mathbb{E}(1 - L)^{\alpha/2}\right)^{\frac{2}{\alpha}},$$

which further yields

$$\lim_{u \to 1} \frac{Q_u(\eta' \Sigma \eta)}{Q_u(\eta'(I - \Sigma)\eta)} = \left(\frac{\mathbb{E}(L^{\alpha/2})}{\mathbb{E}(1 - L)^{\alpha/2}}\right)^{\frac{2}{\alpha}}. \tag{23}$$

We are now ready to calculate the limit inside the expectation in (20). Choose specific values $x = Q_u(\eta' \Sigma \eta)$ and $y = Q_u(\eta'(I - \Sigma)\eta)$. Using the notation of the C_i sets, we can write the set in the numerator as $C_0(x) \bigcap C_1(y)$. Similar to the way we handled C_i individually, for any $0 < \epsilon < 1$, we can get an upper bound of this set as

$$C_0(x) \bigcap C_1(y) \subset \left(A_0\left(x(1 - \epsilon)\right) \bigcup B_0\left(\frac{\epsilon}{n(n-1)/2}x\right)\right)$$
$$\bigcap \left(A_1\left(y(1 - \epsilon)\right) \bigcup B_1\left(\frac{\epsilon}{n(n-1)/2}y\right)\right).$$

Clearly, the upper bound can be expanded as the union set of four sets in the form of $A_0 \bigcap A_1, B_0 \bigcap A_1, A_0 \bigcap B_1$, and $B_0 \bigcap B_1$. Under the $\tilde{\mathbb{P}}$ measure, the limit relation (22) ensures that the probability of the last three sets involving B_i sets tends to zero faster than $1 - u$ as $u \to 1$. Hence, we only need to handle the first set in the form of $A_0 \bigcap A_1$.

The limit in (23) shows that x/y converges to a positive constant as $u \to 1$. Since $\{\eta_t^2\}_{t=1}^n$ are n i.i.d. heavy-tailed random variables, we can apply Lemma 1 to the series $\{\eta_t^2\}_{t=1}^n$, which leads to an upper bound for $\tilde{\mathbb{P}}\left(C_0(x) \bigcap C_1(y)\right)$ as

$$\limsup_{u \to 1} \frac{\tilde{\mathbb{P}}\left(C_0(x) \bigcap C_1(y)\right)}{1 - u} \leq \lim_{u \to 1} \frac{\tilde{\mathbb{P}}\left(A_0\left(x(1 - \epsilon)\right) \bigcap A_1\left(y(1 - \epsilon)\right)\right)}{1 - u}$$
$$= \lim_{u \to 1} \frac{\sum_{t=1}^n \tilde{\mathbb{P}}\left(\eta_t^2 > (1 - \epsilon)\left(\frac{x}{L_{tt}} \vee \frac{y}{1 - L_{tt}}\right)\right)}{1 - u}$$
$$= (1 - \varepsilon)^{-\alpha/2} \sum_{t=1}^n \frac{L_{tt}^{\alpha/2}}{n\mathbb{E}(L^{\alpha/2})} \wedge \frac{(1 - L_{tt})^{\alpha/2}}{n\mathbb{E}((1 - L)^{\alpha/2})}. \tag{24}$$

Similarly, to obtain a lower bound for $\tilde{\mathbb{P}}\left(C_0(x) \bigcap C_1(y)\right)$, we use the fact that

$$C_0(x) \bigcap C_1(y) \supset \left(A_0\left(x(1 + \epsilon)\right) \bigcap B_1^c\left(\frac{\epsilon}{n(n-1)/2}x\right)\right)$$

$$\bigcap \left(A_1\left(y(1+\epsilon)\right) \bigcap B_0^c\left(\frac{\epsilon}{n(n-1)/2}y \right) \right)$$
$$= \left(A_0\left(x(1+\epsilon)\right) \bigcap A_1\left(y(1+\epsilon)\right) \right) \bigcap$$
$$\times \left(B_1\left(\frac{\epsilon}{n(n-1)/2}x \right) \bigcup B_0\left(\frac{\epsilon}{n(n-1)/2}y \right) \right)^c.$$

Under the $\tilde{\mathbb{P}}$ measure, the probability of the set $B_1\left(\frac{\epsilon}{n(n-1)/2}x \right) \bigcup B_0\left(\frac{\epsilon}{n(n-1)/2}y \right)$ tends to zero faster than $1-u$ as $u \to 1$, which yields that

$$\liminf_{u \to 1} \frac{\tilde{\mathbb{P}}\left(C_0(x) \bigcap C_1(y) \right)}{1-u} \geq \lim_{u \to 1} \frac{\tilde{\mathbb{P}}\left(A_0\left(x(1+\epsilon)\right) \bigcap A_1\left(y(1+\epsilon)\right) \right)}{1-u}$$
$$= (1+\varepsilon)^{-\alpha/2} \sum_{t=1}^{n} \frac{L_{tt}^{\alpha/2}}{n\mathbb{E}(L^{\alpha/2})} \wedge \frac{(1-L_{tt})^{\alpha/2}}{n\mathbb{E}((1-L)^{\alpha/2})}, \qquad \textbf{(25)}$$

where the last step is derived in a similar way as for the upper bound. Combining (24) and (25), together with taking $\epsilon \to 0$, we obtain that

$$\lim_{u \to 1} \frac{\tilde{\mathbb{P}}\left(C_0(x) \bigcap C_1(y) \right)}{1-u} = \sum_{t=1}^{n} \frac{L_{tt}^{\alpha/2}}{n\mathbb{E}(L^{\alpha/2})} \wedge \frac{(1-L_{tt})^{\alpha/2}}{n\mathbb{E}((1-L)^{\alpha/2})}.$$

The theorem follows from taking expectation on both sides of this limit relation, combined with interchanging the expectation and the limit, justified in the Appendix. ∎

4.4. Proof of Proposition 1

Proof of Proposition 1. Since $1 - \frac{\left(\sum_{j=1}^{n} X_{2,j}X_{1,j} \right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} > 0$, we can write $|C_t|$ as

$$|C_t| = \frac{\left| \frac{X_{1,t}}{\sum_{j=1}^{n} X_{1,j}^2} - \frac{X_{2,t}\sum_{j=1}^{n} X_{1,j}X_{2,j}}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right|}{1 - \frac{\left(\sum_{j=1}^{n} X_{2,j}X_{1,j} \right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2}}$$
$$\leq \frac{\frac{|X_{1,t}|}{\sum_{j=1}^{n} X_{1,j}^2} + \frac{|X_{2,t}| \max_{1 \leq j \leq n} |X_{2,j}| \sum_{j=1}^{n} |X_{1,j}|}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2}}{1 - \frac{\left(\sum_{j=1}^{n} X_{2,j}X_{1,j} \right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2}} \leq \frac{\frac{|X_{1,t}|}{\sum_{j=1}^{n} X_{1,j}^2} + \frac{\sum_{j=1}^{n} |X_{1,j}|}{\sum_{j=1}^{n} X_{1,j}^2}}{1 - \frac{\left(\sum_{j=1}^{n} X_{2,j}X_{1,j} \right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2}}.$$

Since $\sum_{j=1}^{n} X_{1,j}^2 \geq n(\sum_{j=1}^{n} |X_{1,j}|)^2 > \frac{n}{4}(\sum_{j=1}^{n} |X_{1,j}| + |X_{1,t}|)^2$, we get that, for any $0 < K < \frac{n}{4}$,

$$|C_t|^{-2} > K \sum_{j=1}^{n} X_{1,j}^2 \left(1 - \frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right)^2 := K\tilde{C}_t.$$

Therefore, to prove $\mathbb{E}|C_t|^{\alpha+\epsilon} < +\infty$, we only need to verify that $\mathbb{E}\tilde{C}_t^{-(\alpha+\epsilon)/2} < +\infty$.

By the Cauchy–Schwarz inequality, we get that

$$\mathbb{E}\tilde{C}_t^{-(\alpha+\epsilon)/2} \leq \mathbb{E} \left(\sum_{j=1}^{n} X_{1,j}^2 \right)^{-\alpha-\epsilon} \mathbb{E} \left(1 - \frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right)^{-2(\alpha+\epsilon)}.$$

By Lemma 3.6 in Mikosch and de Vries (2013), Condition 1 implies that $\mathbb{E}\left(\sum_{j=1}^{n} X_{1,j}^2\right)^{-\alpha-\epsilon} < +\infty$.

Finally, we show that Condition 2 implies that $\mathbb{E}\left(1 - \frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right)^{-2(\alpha+\epsilon)} < +\infty$. Since

$$\mathbb{E}\left(1 - \frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right)^{-2(\alpha+\epsilon)}$$

$$= \int_1^\infty \mathbb{P}\left(\frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} > 1 - x^{-\frac{1}{2(\alpha+\epsilon)}} \right) dx,$$

Condition 2 implies that, for $x > x_1^{-2(\alpha+\epsilon)}$,

$$\mathbb{P}\left(\frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} > 1 - x^{-\frac{1}{2(\alpha+\epsilon)}} \right) \leq c_1 x^{-\frac{\gamma_1}{2(\alpha+\epsilon)}}.$$

Since $\gamma_1 > 2\alpha$, by choosing ϵ sufficiently small such that $\frac{\gamma_1}{2(\alpha+\epsilon)} > 1$, we obtain that $\mathbb{E}\left(1 - \frac{\left(\sum_{j=1}^{n} X_{2,j} X_{1,j}\right)^2}{\sum_{j=1}^{n} X_{1,j}^2 \sum_{j=1}^{n} X_{2,j}^2} \right)^{-2(\alpha+\epsilon)} < +\infty$. Proposition 1 thus follows.

As an example, we verify that if the regressors $(X_{1,t}, X_{2,t})$ follow a bivariate normal distribution with correlation parameter $|\rho| < 1$, then the two sufficient conditions hold. Condition 1(b) regarding the marginal distribution holds trivially. We only check that Condition 2 holds.

W.l.o.g., we assume that marginal normal distribution has both mean zero and unit variance. Hotelling ([1953](#)) provides the density function of the sample correlation coefficient R_n, where $R_n^2 = \frac{(\sum_{t=1}^n X_{2,t} X_{1,t})^2}{\sum_{t=1}^n X_{1,t}^2 \sum_{t=1}^n X_{2,t}^2}$, as follows:

$$g_n(r, \rho) = \frac{n-2}{\sqrt{2\pi}} \frac{\Gamma(n-1)}{\Gamma(n-1/2)} (1-\rho^2)^{(n-1)/2} (1-r^2)^{(n-4)/2} (1-\rho r)^{-n-1/2} \times$$
$$F(1/2, 1/2, n-1/2; (1+\rho r)/2),$$

where $F(.,.,.;.)$ is the Gaussian hypergeometric function. Condition 2 can be rewritten as there exists x_1 such that, for all $x < x_1$,

$$\int_0^x g_n(1-z, \rho)dz \le c_1 x^{\gamma_1} \text{ and } \int_0^x g_n(z-1, \rho)dz \le c_1 x^{\gamma_1},$$

for some $c_1 > 0$ and $\gamma_1 > 2\alpha$. This condition is met if $g_n(x-1, \rho) \le c_1 \gamma_1 x^{\gamma_1 - 1}$ and $g_n(1-x, \rho) \le c_1 \gamma_1 x^{\gamma_1 - 1}$ for x sufficiently close to zero.

Notice that, as $x \to 0$, $g_n(x-1, \rho) = O(x^{(n-4)/2})$ and $g_n(1-x, \rho) = O(x^{(n-4)/2})$. Hence, the inequality holds for $\gamma_1 - 1 < \frac{n-4}{2}$. Therefore, Condition 2 holds if $n > 4\alpha + 2$. ∎

5. CONCLUSION

In this paper, we study the tail dependence between OLS estimators for the regression coefficients when the error terms in the regression are heavy-tailed. We show the presence of tail dependence and provide an explicit formula relating the level of tail dependence to the distribution of the regressors. We also show that the FSS and the RSS are tail-dependent by providing an explicit formula for calculating the level of tail dependence. Simulation studies confirm our theoretical findings.

In practice, the error terms in a regression may possess heavy tails, for instance, when regressing a heavy-tailed dependent variable, such as a financial variable, on a set of thin-tailed independent variables. If, in addition, the number of observations is low, then the OLS estimators for the regression coefficients can be heavy-tailed and tail-dependent. That means if the estimator for one coefficient deviates largely from its true value, the estimator for another coefficient may also deviate from its true value substantially. A practitioner has to be cautious in interpreting the significance of the results from such a regression.

APPENDIX

Interchanging the Expectation and the Limit

We show that we may interchange the expectation and the limit in ([17](#)). The proof follows similar steps as in the proof of Breiman's lemma (see Breiman, [1965](#)).

In particular, we will show that

$$\lim_{u \to 1} \frac{\mathbb{P}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1\right), \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2\right)\right)}{1 - u}$$

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

Recall that $\eta = (\eta_1, \ldots, \eta_n)'$, define $\mathbf{C} = (C_1, \ldots, C_n)'$, and let $||.||$ denote the euclidean norm. We have

$$\frac{\mathbb{P}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1\right), \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2\right)\right)}{1 - u}$$

$$= \mathbb{E}\left(\frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1), \hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2)\right)\mathbb{I}\{||\mathbf{C}||^{\alpha+\epsilon} \geq 1/(1-u)\}}{1 - u}\right)$$

$$+ \mathbb{E}\left(\frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u(\hat{\beta}_1 - \beta_1), \hat{\beta}_2 - \beta_2 > Q_u(\hat{\beta}_2 - \beta_2)\right)\mathbb{I}\{||\mathbf{C}||^{\alpha+\epsilon} < 1/(1-u)\}}{1 - u}\right)$$

$$:= I_1^u + I_2^u.$$

We first handle I_1^u. Note that the conditional probability $\tilde{\mathbb{P}}(\cdots)$ is bounded by 1. Therefore, as $u \to 1$,

$$I_1^u \leq \mathbb{E}\left(||\mathbf{C}||^{\alpha+\epsilon}\mathbb{I}\{||\mathbf{C}||^{\alpha+\epsilon} \geq 1/(1-u)\}\right) \to 0,$$

because $\mathbb{E}|C|^{\alpha+\epsilon} < \infty$ implies that $\mathbb{E}||\mathbf{C}||^{\alpha+\epsilon} < \infty$ for sufficiently small ϵ.

To deal with the second term, I_2^u, we will find an integrable function dominating its integrand such that we can apply the dominated convergence theorem. First, by applying the Cauchy–Schwarz inequality to $\mathbf{C}'\eta = \hat{\beta}_1 - \beta_1$, we obtain

$$\frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1\right), \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2\right)\right)}{1 - u} \leq \frac{\tilde{\mathbb{P}}\left(||\mathbf{C}||||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)}{1 - u}. \tag{26}$$

Since $||\eta||$ possesses a tail index α (see, e.g., Lemma 3.4 in Mikosch and de Vries (2013)), Lemma 3 implies that $\mathbb{P}\left(||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right) = O(1-u)$ as $u \to 1$. Therefore, there exists $K > 0$ such that for u close enough to 1,

$$\frac{\tilde{\mathbb{P}}\left(||\mathbf{C}||||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)}{1 - u} \leq K\frac{\tilde{\mathbb{P}}\left(||\mathbf{C}||||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)}{\mathbb{P}\left(||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)}. \tag{27}$$

Note that on the event $\{||\mathbf{C}||^{\alpha+\epsilon} < 1/(1-u)\}$, we have $Q_u\left(\hat{\beta}_1 - \beta_1\right)/||\mathbf{C}|| \to \infty$ as $u \to 1$. Therefore, by Potter's bounds for regularly varying functions (Potter, 1942), we have that

on the event $\{||\mathbf{C}||^{\alpha+\epsilon} < 1/(1-u)\}$, for any $\delta > 0$, as u is close enough to 1,

$$\frac{\tilde{\mathbb{P}}\left(||\mathbf{C}||||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)}{\mathbb{P}\left(||\eta|| > Q_u\left(\hat{\beta}_1 - \beta_1\right)\right)} \leq (1+\delta)||\mathbf{C}||^\alpha \max\left(||\mathbf{C}||^\delta, ||\mathbf{C}||^{-\delta}\right). \tag{28}$$

We obtain by combining (26)–(28) that, as u is close enough to 1,

$$\frac{\tilde{\mathbb{P}}\left(\hat{\beta}_1 - \beta_1 > Q_u\left(\hat{\beta}_1 - \beta_1\right), \hat{\beta}_2 - \beta_2 > Q_u\left(\hat{\beta}_2 - \beta_2\right)\right)}{1-u}$$
$$\leq K(1+\delta)||\mathbf{C}||^\alpha \max\left(||\mathbf{C}||^\delta, ||\mathbf{C}||^{-\delta}\right).$$

Hence, by choosing δ sufficiently small, we have that the integrand in I_2^u is bounded by the integrable function $K(1+\delta)||\mathbf{C}||^\alpha \max\left(||\mathbf{C}||^\delta, ||\mathbf{C}||^{-\delta}\right)$. The statement for interchanging the limit and the expectation follows from applying the dominated convergence theorem.

REFERENCES

Basrak, B., R. A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99(1), 95–115.

Bingham, N. H., C. M. Goldie, & J. L. Teugels (1989) *Regular Variation*, Vol. 27. Cambridge University Press.

Blattberg, R. & T. Sargent (1971) Regression with non-Gaussian stable disturbances: Some sampling results. *Econometrica* 39(3), 501–510.

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

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

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

Feller, W. (1971) *Introduction to Probability Theory and Its Applications*, Vol. II. Wiley.

Han, C., J. S. Cho, & P. C. Phillips (2011) Infinite density at the median and the typical shape of stock return distributions. *Journal of Business & Economic Statistics* 29(2), 282–294.

He, X., J. Jurečková, R. Koenker, & S. Portnoy (1990) Tail behavior of regression estimators and their breakdown points. *Econometrica* 58(5), 1195–1214.

Hotelling, H. (1953) New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)* 15(2), 193–232.

Jansen, D. W. & C. G. de Vrie (1991) On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics* 73(1), 18–24.

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

Potter, H. (1942) The mean values of certain Dirichlet series, II. *Proceedings of the London Mathematical Society* 2(1), 1–19.

Qin, H. & A. T. Wan (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, Tinbergen Institute.