TI 2006-085/2
Tinbergen Institute Discussion Paper

Tail Probabilities for Regression Estimators

Thomas Mikosch[1]
Casper G. de Vries[2]

1 University of Copenhagen,
2 Erasmus Universiteit Rotterdam, and Tinbergen Institute.

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

TAIL PROBABILITIES FOR REGRESSION ESTIMATORS

THOMAS MIKOSCH[1] AND CASPER G. DE VRIES[2]

Abstract. Estimators of regression coefficients are known to be asymptotically normally distributed, provided certain regularity conditions are satisfied. In small samples and if the noise is not normally distributed, this can be a poor guide to the quality of the estimators. The paper addresses this problem for small and medium sized samples and heavy tailed noise. In particular, we assume that the noise is regularly varying, i.e., the tails of the noise distribution exhibit power law behavior. Then the distributions of the regression estimators are heavy tailed themselves. This is relevant for regressions involving financial data which are typically heavy tailed. In medium sized samples and with some dependency in the noise structure, the regression coefficient estimators can deviate considerably from their true values. The relevance of the theory is demonstrated for the highly variable cross country estimates of the expectations coefficient in yield curve regressions.

1. Some motivation

Estimators of the coefficients in equations of regression type which involve financial data are often found to vary considerably across different samples. This observation pertains to finance models like the CAPM beta regression, the forward premium equation and the yield curve regression. In economics, macro models like the monetary model of the foreign exchange rate often yield regression coefficients which significantly deviate from the unitary coefficient on money which is based on the theoretical assumption that money is neutral. The uncertainty in CAPM regressions was reviewed in Campbell et al. (1997, Chapter 5) and Cochrane (2001, Chapter 15). Lettau and Ludvigson (2001) explicitly model the time variation in beta. Starting with Bilson (1981), Hodrick (1987) and Lewis (1995) report wildly different estimates for the Fisher coefficient in forward premium regressions; see Fisher (1930, p. 39). Moreover, typical estimates of the expectation coefficient in yield curve regressions reported by Fama (1976), Mankiw and Miron (1986), Campbell and Shiller (1991) show substantial variation over time and appear to be downward biased; Campbell et al. (1997, Chapter 10.2) provide a lucid review. The coefficient of the relative money supply in the regression of the exchange rate on the variables of the monetary model of the foreign exchange rate varies considerably around its theoretical unitary value; see for example Frenkel (1993, Chapter 4). In monetary economics parameter uncertainty is sometimes explicitly taken into account when it comes to policy decisions, see Brainard (1967) and, more recently, Sack (2000). In a random coefficient model, see for example Feige and Swamy (1974), the (regression) coefficients themselves are subject to randomness and therefore fluctuate about some fixed values. Estimation of the random coefficient models is reviewed in Maddala (1977) and Chow (1984).

In this paper the considerable variation in regression coefficients across different samples is explained from the heavy tailed nature of the distribution of the innovations, which can be either additive or multiplicative. The standard $\sqrt{n}$-rates of convergence for the parameter estimators,

---

Date: April 6, 2006.

[1] This research is supported by a Danish Research Council (SNF) Grant.

[2] Corresponding author.

We like to thank Martijn van Harten and Olaf van Veen for their valuable research assistance.

Both authors are grateful to the stimulating research environments provided at EURANDOM in Eindhoven (The Netherlands) and at the Newton Institute in Cambridge (UK) in July 2001 during the thematic period on Managing Risk and Uncertainty, where early parts of this paper were written.

needed for the construction of asymptotic confidence bands, are based on the validity of the central limit theorem. The latter result is a large sample implication which breaks down in samples of small and medium size when the distribution of the noise in the regression equations has considerably thicker tails than the normal one. In such cases the error term has much more influence on the parameter estimates, i.e., large values of the noise cause significant deviations from the classical central limit theory for the regression coefficient estimator. This occurs in particular when the noise structure exhibits some time dependency (like ARMA or ARCH).

A major motivation for our approach is that financial data do not satisfy the normality condition which is often assumed in classical regression estimation. There is in fact considerable statistical evidence that the distribution of returns on asset prices and exchange rates as well as interest rates is suitably modeled by leptokurtic and heavy tailed distributions; see the discussion in Embrechts et al. (1997, Chapter 6), Campbell et al. (1997), or the recent paper by Mikosch (2003). For example, this property appears naturally when the innovations exhibit an ARCH structure.

The finite sample properties of regression coefficient estimators with a deterministic regressor are well understood if the underlying noise variables are normally distributed. The central limit theorem yields a large sample theory for the estimators in regressions with random regressor with possibly non-normal additive noise. However, financial data are unlikely to come from a normal distribution. This concerns, for example, the regressions in the CAPM model for estimating beta. Strategic decisions and rapidly changing business environments imply that one usually works with a relatively short data window. Similarly, as soon as some macro variables are part of the regression, one is compelled to use low frequency data and hence a small sample. Thus a possible cause for the variation in the coefficient estimates are the moderate sample sizes in combination with heavy tailed noise. Therefore the finite sample properties of the estimators under these conditions deserve considerable interest.

Thus our purpose is to study the contribution of heavy tailed noise distributions to the small sample variation of regression estimators. It is shown that, even when standard moment conditions such as the existence of the variance, a finite fourth moment, etc., are satisfied, the usual benchmarks based on central limit theory can be misleading. We give expressions for the tail of the distribution of the regression estimators for a fixed sample size. Small sample results for the distribution of regression estimators are rare and exact results are difficult to obtain, but if the tails of the noise are regularly varying, the tail asymptotic behavior of the regression estimator can be determined. The results hinge on relatively weak assumptions regarding the stochastic nature of the explanatory variable. Under additive uncertainty, we require that the joint density of the explanatory variables is bounded in some neighborhood of the origin. A restriction is the condition that the regressor be exogenous, but the regressor is not assumed to be fixed. We allow for the possibility that the multiplicative noise component is correlated with the additive noise term, and in this sense there can be correlation between the economic explanatory part and the additive noise structure. Moreover, both the noise and the regressor are allowed to be time dependent. The time dependency has a major effect and sets the heavy tailed innovations based regression coefficients apart from their normal based counterparts.

The theoretical results are first illustrated by means of a simulation experiment. The Monte Carlo study demonstrates the importance of dependency in the noise structure. Subsequently, we investigate the relevance of the theory for the expectations hypothesis coefficients in yield curve regressions. These coefficients are well known to exhibit downward 'bias' and exhibit high variability. The downward bias is generally viewed as being due to a risk premium and liquidity premium, to which this paper has little to add. But the theory is relevant for explaining the considerable variability of the coefficient estimates.

2. The model

We study the regression model:

$$Y_t = (\beta + \varepsilon_t)X_t + \varphi_t \,, \tag{2.1}$$

where $((\varepsilon_t, \varphi_t))$ is a strictly stationary noise sequence of 2-dimensional random vectors, and (X_t) is a sequence of explanatory variables, independent of the noise. In what follows, we write φ, ε, etc., for generic elements of the strictly stationary sequences (φ_t), (ε_t), etc. The coefficient β is a fixed parameter to be estimated by regression, but it can also be of interest to determine distributional characteristics of ε_t such as its variance. The model (2.1) comprises a large variety of different economic models since it allows for both additive and multiplicative uncertainty. If the noises ε_t and φ_t have zero mean, then, conditionally on the information at time $t-1$, the model (2.1) captures the structure of many of the rational expectations finance models such as the CAPM.

In what follows, we assume that the right tail of the marginal distributions $F_\varepsilon(x)$ and $F_\varphi(x)$ of ε and φ, respectively, is regularly varying with index $\alpha > 0$. This means that the limits

$$\lim_{x\to\infty} \frac{1-F(xs)}{1-F(x)} = s^{-\alpha} \quad \text{for all } s > 0, \tag{2.2}$$

exist for $F \in \{F_\varepsilon, F_\varphi\}$. Regular variation entails that $(\alpha+\delta)$th moments of F are infinite for $\delta > 0$, supporting the intuition on the notion of heavy tailed distribution. Some prominent members of the class of distributions with regularly varying tails are the Student-t, F-, Fréchet, infinite variance stable and Pareto distributions. First order approximations to the tails of these distribution functions F are comparable to the tail $c\,x^{-\alpha}$ of a Pareto distribution for some $c, \alpha > 0$, i.e.,

$$\lim_{x\to\infty} \frac{1-F(x)}{c\,x^{-\alpha}} = 1\,.$$

The power like decay in the right tail area implies the lack of moments higher than α. There are other distributions which have fatter tails than the normal distribution, such as the exponential or lognormal distributions. But these distributions possess all power moments. They are less suitable for capturing the very large positive and very small negative values observed in financial data sets.

Independent positive random variables $A_1, \ldots, A_n$ with regularly varying right tails (possibly with different indices) satisfy a well known additivity property of their convolutions; see for example Feller (1971). This means that

$$\lim_{x\to\infty} \frac{\sum_{i=1}^n P(A_i > x)}{P\left(\sum_{i=1}^n A_i > x\right)} = 1\,. \tag{2.3}$$

This is a useful fact when it comes to evaluating the distributional tail of (weighted) sums of random variables with regularly varying tails. The *ordinary least squares* (OLS) estimator of β is comprised of such sums, but also involves products and ratios of random variables. In particular, the OLS estimator $\widehat{\beta}$ of β in model (2.1), given by

$$\widehat{\beta} = \frac{\sum_{i=1}^n X_t Y_t}{\sum_{i=1}^n X_t^2} = \beta + \rho_{n,\varepsilon} + \rho_{n,\varphi}\,, \tag{2.4}$$

involves the quantities

$$\rho_{n,\varepsilon} := \frac{\Sigma_{i=1}^n \varepsilon_t X_t^2}{\Sigma_{t=1}^n X_t^2} \quad \text{and} \quad \rho_{n,\varphi} := \frac{\Sigma_{i=1}^n \varphi_t X_t}{\Sigma_{t=1}^n X_t^2}. \tag{2.5}$$

Thus, in the case of fixed regressors X_t and with noise $(\varepsilon_t, \varphi_t)$ whose components have distributions with regularly varying tails, one can rely on the additivity property (2.3). But if the regressors are stochastic, we face a more complicated problem for which we derive new results.

This paper aims at investigating the finite sample variability of the regression coefficient estimator in models with additive noise and random coefficients when the noise comes from a heavy tailed distribution. In Section 3 we derive the finite sample tail properties of the distribution of the OLS estimator of β in model (2.1) when the noise has a distribution with regularly varying tails; see (2.2). A small simulation study in Section 4 conveys the relevance of the theory. Section 5 applies the theory to the distribution of the expectations coefficient in yield curve estimation. Some proofs are relegated to the Appendix.

3. Theory

In this section we derive the finite sample tail properties of the distribution of the OLS regression coefficient estimator in the model (2.1) when the noise distribution has regularly varying tails. To this end we first recall in Section 3.1.1 the definitions of regular and slow variation as well as the basic scaling property for convolutions of random variables with regularly varying distributions. Subsequently, we obtain the regular variation properties for inner products of those vectors of random variables that appear in the OLS estimator of β. The joint distribution of these inner products is multivariate regularly varying. In Section 3.1.2 we give conditions for the finiteness of moments of quotients of random variables. Finally, we derive the asymptotic tail behavior of the distribution of the OLS estimator of β by combining the previous results. We present the main results on the finite sample tail behavior of $\widehat{\beta}$ for iid regularly varying noise (Section 3.2.1), for regularly varying linearly dependent noise (Section 3.2.2) and give some comments on the case of general regularly varying noise (Section 3.2.3). In Section 3.2.4 the finite sample results are contrasted with the large sample results typically used in standard econometric test procedures.

3.1. **Preliminaries.**

3.1.1. *Regular variation.* A positive measurable function L on $[0,\infty)$ is said to be *slowly varying* if

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1 \quad \text{for all } c > 0.$$

The function $g(x) = x^\alpha L(x)$ for some $\alpha \in \mathbb{R}$ is then said to be *regularly varying with index* α. We say that the *random variable* X and *its distribution* F (we use the same symbol F for its distribution function) are *regularly varying with* (*tail*) *index* $\alpha \ge 0$ if there exist $p, q \ge 0$ with $p + q = 1$ and a slowly varying function L such that

$$F(-x) = q\, x^{-\alpha}\, L(x)\,(1+o(1)) \quad \text{and } \overline{F}(x) := 1 - F(x) = p\, x^{-\alpha}\, L(x)\,(1+o(1))\,, \quad x \to \infty\,. \tag{3.1}$$

Condition (3.1) is usually referred to as a *tail balance condition.* For an encyclopedic treatment of regular variation, see Bingham et al. (1987).

In what follows, $a(x) \sim b(x)$ for positive functions a and b means that $a(x)/b(x) \to 1$, usually as $x \to \infty$. We start with an auxiliary result which is a slight extension of Lemma 2.1 in Davis and Resnick (1996) where this result was proved for non-negative random variables. The proof in the general case is analogous and therefore omitted.

Lemma 3.1. *Let F be a distribution function concentrated on $(0,\infty)$ satisfying (3.1) with $p, q > 0$. Assume $Z_1, \ldots, Z_n$ are random variables such that $F(x) = P(|Z_1| \le x)$ and*

$$\lim_{x\to\infty} \frac{P(Z_i > x)}{\overline{F}(x)} = c_i^+ \quad \textit{and} \quad \lim_{x\to\infty} \frac{P(Z_i \le -x)}{\overline{F}(x)} = c_i^-\,, \quad i = 1, \ldots, n\,, \tag{3.2}$$

for some non-negative numbers $c_i^\pm$ and

$$\lim_{x\to\infty} \frac{P(Z_i > x\,, Z_j > x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{P(Z_i \le -x\,, Z_j > x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{P(Z_i \le -x\,, Z_j \le -x)}{\overline{F}(x)} = 0\,, \quad i \ne j\,.$$

Then

$$\lim_{x\to\infty} \frac{P(Z_1 + \cdots + Z_n > x)}{\overline{F}(x)} = c_1^+ + \cdots + c_n^+$$

and

$$\lim_{x\to\infty} \frac{P(Z_1+\cdots+Z_n \le -x)}{\overline{F}(x)} = c_1^- + \cdots + c_n^- .$$

The following result is a consequence of this lemma.

Lemma 3.2. *Suppose Z_i are regularly varying random variables with tail index $\alpha_i > 0$, $i = 1,\dots,n$. Assume that one of the following conditions holds.*

1. *The Z_i's are independent and satisfy (3.2) with $\overline{F}(x) = P(|Z_1| > x)$, $x > 0$.*
2. *The Z_i's are non-negative and independent.*
3. *Z_1 and Z_2 are regularly varying with indices $0 < \alpha_1 < \alpha_2$ and the parameters p_1, q_1 in the tail balance condition (3.2) for Z_1 are positive.*

Then under (1) or (2) the relations

$$\begin{aligned} P(Z_1+\cdots+Z_n > x) &\sim P(Z_1 > x) + \cdots + P(Z_n > x)\,, \\ P(Z_1+\cdots+Z_n \le -x) &\sim P(Z_1 \le -x) + \cdots + P(Z_n \le -x) \end{aligned}$$

hold as $x \to \infty$. If condition (3) applies, as $x \to \infty$,

$$P(Z_1+Z_2 > x) \sim P(Z_1 > x) \quad \text{and} \quad P(Z_1+Z_2 \le -x) \sim P(Z_1 \le -x)$$

The proof is given in the Appendix.

Recall the definition of a regularly varying random vector $\mathbf{X}$ with values in $\mathbb{R}^d$; see for example de Haan and Resnick (1977), Resnick (1986,1987). In what follows, $\mathbb{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$ with respect to a (given) norm $|\cdot|$ and $\stackrel{v}{\to}$ refers to *vague convergence* on the Borel σ-field of $\mathbb{S}^{d-1}$; see Resnick (1986,1987) for details.

Definition 3.3. *The random vector $\mathbf{X}$ with values in $\mathbb{R}^d$ and its distribution are said to be* regularly varying with index α and spectral measure $P_{\mathbf{\Theta}}$ *if there exists a random vector $\mathbf{\Theta}$ with values in $\mathbb{S}^{d-1}$ and distribution $P_{\mathbf{\Theta}}$ such that the following limit exists for all $t > 0$:*

$$\frac{P(|\mathbf{X}| > tx\,, \mathbf{X}/|\mathbf{X}| \in \cdot)}{P(|\mathbf{X}| > x)} \stackrel{v}{\to} t^{-\alpha} P_{\mathbf{\Theta}}(\cdot)\,, \quad x \to \infty\,. \tag{3.3}$$

The vague convergence in (3.3) means that

$$\frac{P\left(|\mathbf{X}| > tx\,, \mathbf{X}/\,|\mathbf{X}| \in S\right)}{P(|\mathbf{X}| > x)} \to t^{-\alpha} P_{\mathbf{\Theta}}(S)\,,$$

for all Borel sets $S \subset \mathbb{S}^{d-1}$ such that $P_{\mathbf{\Theta}}(\partial(S)) = 0$, where $\partial(S)$ denotes the boundary of S. Alternatively, (3.3) is equivalent to the totality of the relations

$$\frac{P(\mathbf{X} \in x\,A)}{P(|\mathbf{X}| > x)} \to \mu(A)\,.$$

Here μ is a measure on the Borel σ-field of $\overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ with property $\mu(tA) = t^{-\alpha}\mu(A)$, $t > 0$, for any Borel set $A \subset \overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$, bounded away from zero and such that $\mu(\partial(A)) = 0$.

We have the following result.

Lemma 3.4. *Assume that $\mathbf{X} = (X_1, ..., X_d)'$ is regularly varying in $\mathbb{R}^d$ with index $\alpha > 0$ and is independent of the random vector $\mathbf{Y} = (Y_1, ..., Y_d)'$ which satisfies $E|\mathbf{Y}|^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$. Then the scalar product $Z = \mathbf{X}'\mathbf{Y}$ is regularly varying with index α. Moreover, if $\mathbf{X}$ has independent* *components, then as $x \to \infty$,*

$$
\begin{aligned}
P(Z > x) &\sim P(|\mathbf{X}| > x) \left[\sum_{i=1}^d c_i^+ \, E[Y_i^\alpha I_{\{Y_i>0\}}] + \sum_{i=1}^d c_i^- \, E[|Y_i|^\alpha I_{\{Y_i<0\}}] \right] , \\
P(Z \le -x) &\sim P(|\mathbf{X}| > x) \left[\sum_{i=1}^d c_i^- \, E[Y_i^\alpha I_{\{Y_i>0\}}] + \sum_{i=1}^d c_i^+ E[|Y_i|^\alpha I_{\{Y_i<0\}}] \right] ,
\end{aligned} \tag{3.4}
$$

where

$$c_i^+ = \lim_{x\to\infty} \frac{P(X_i > x)}{P(|\mathbf{X}| > x)} \quad \text{and} \quad c_i^- = \lim_{x\to\infty} \frac{P(X_i \le -x)}{P(|\mathbf{X}| > x)} \,.$$

The proof is given in the Appendix.

Remark 3.5. To give some intuition on Lemma 3.4, consider the case $d = 1$, i.e., $Z = X_1 Y_1$, and assume for simplicity that X_1 and Y_1 are positive random variables. Then the lemma says that

$$P(X_1 Y_1 > x) \sim EY_1^\alpha \, P(X_1 > x) \,, \quad x \to \infty \,. \tag{3.5}$$

The latter relation is easily seen if one further specifies that $P(X_1 > x) = cx^{-\alpha}$, $x \ge c^{1/\alpha}$. Then a conditioning argument immediately yields for large x,

$$
\begin{aligned}
P(Z > x) &= E[P(X_1 Y_1 > x \mid Y_1)] \\
&= \int_0^{c^{-1/\alpha}x} P(X_1 > x/y) \, dP(Y_1 \le y) + P(Y_1 > c^{-1/\alpha}x) \\
&= \int_0^{c^{-1/\alpha}x} cx^{-\alpha} \, y^\alpha \, dP(Y_1 \le y) + P(Y_1 > c^{-1/\alpha}x) \\
&= P(X_1 > x) \int_0^{c^{-1/\alpha}x} y^\alpha \, dP(Y_1 \le y) + o(P(X_1 > x)) \\
&= P(X_1 > x) \, EY_1^\alpha \, (1 + o(1)) \,.
\end{aligned}
$$

Relation (3.5) is usually referred to as *Breiman's result*; see Breiman (1965). A generalization to matrix-valued $\mathbf{Y}$ and vectors $\mathbf{X}$ can be found in Basrak et al. (2002).

Remark 3.6. The lemma can be extended to infinite series $\sum_i X_i Y_i$ if (X_i) and (Y_i) are independent, the finite-dimensional distributions of (X_i) are regularly varying with index α, $E\,|Y_i|^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$ and some further technical conditions are satisfied.

3.1.2. *Finiteness of moments.* In this section we give conditions under which the moments of the random variables

$$\widetilde{X}_t = \frac{X_t}{\sum_{s=1}^n X_s^2} \, I_{\{X_t \ne 0\}} , \quad t = 1, ..., n,$$

are finite, where (X_t) is a sequence of random variables. First note that

$$|\widetilde{X}_t|^{-2} \ge I_{\{X_t \ne 0\}} \sum_{s=1}^n X_s^2 =: Y_t \,, \quad t = 1, ..., n, \tag{3.6}$$

It will be convenient to work with the sequence (Y_t).

Lemma 3.7. *Let α be a positive number. Assume that one of the following conditions is satisfied:*

(1) $X_1, ..., X_n$ *are iid,* $P(|X| \le x) \le cx^\gamma$ *for some* $\gamma, c > 0$ *and all* $x \le x_0$*, and* $n\gamma > \alpha$.
(2) $(X_1, ..., X_n)$ *has a bounded density* f_n *in some neighborhood of the origin and* $n > \alpha$.

Then $EY_t^{-\alpha/2} < \infty$ *and, hence,* $E|\widetilde{X}_t|^\alpha < \infty$ *for* $t = 1, ..., n$.

The proof is given in the Appendix. Condition (2) is for example satisfied if $(X_1, ..., X_n)$ is Gaussian and $n > \alpha$.

3.2. **Tail asymptotics for regression coefficient estimators.**

3.2.1. *IID noise.* In this section we consider three sequences of random variables satisfying the following basic **Assumptions:**

(1) (X_t) is a sequence of random variables with $X_t \neq 0$ a.s. for every t.
(2) (ε_t) is iid, and ε is regularly varying with index $\alpha_\varepsilon > 0$.
(3) (φ_t) is iid, and φ is regularly varying with index $\alpha_\varphi > 0$.
(4) (X_t) is independent of $((\varepsilon_t, \varphi_t))$.

We investigate the distributional tails of the quantities $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ defined in (2.5). Recall from (2.4) that the latter quantities are closely related to the OLS estimator $\widehat{\beta}$ of β in the regression model (2.1) with multiplicative and additive noise.

Proposition 3.8. *Assume conditions* (1), (4) *and fix* $n \geq 2$.

(1) *If* (2) *holds, then as* $x \to \infty$,

$$\begin{cases} P(\rho_{n,\varepsilon} > x) & \sim \quad P(\varepsilon > x) \displaystyle\sum_{i=1}^n E\left[\frac{X_i^2}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varepsilon}, \\ P(\rho_{n,\varepsilon} \leq -x) & \sim \quad P(\varepsilon \leq -x) \displaystyle\sum_{i=1}^n E\left[\frac{X_i^2}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varepsilon}. \end{cases} \tag{3.7}$$

(2) *If* (3) *and, in addition,*

$$E\left(\sum_{i=1}^n X_i^2\right)^{-\frac{\alpha_\varphi}{2}+\delta} < \infty \quad \textit{for some} \quad \delta > 0, \tag{3.8}$$

hold, then, as $x \to \infty$,

$$P(\rho_{n,\varphi} > x) \;\sim\; P(\varphi > x) \sum_{i=1}^n E\left[\frac{X_i I_{\{X_i>0\}}}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varphi} + P(\varphi \leq -x) \sum_{i=1}^n E\left[\frac{|X_i| I_{\{X_i<0\}}}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varphi},$$

$$P(\rho_{n,\varphi} \leq -x) \;\sim\; P(\varphi \leq -x) \sum_{i=1}^n E\left[\frac{X_i I_{\{X_i>0\}}}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varphi} + P(\varphi > x) \sum_{i=1}^n E\left[\frac{|X_i| I_{\{X_i<0\}}}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varphi}.$$

Proof. Since (X_t) and (ε_t) are independent and $X_t^2/\sum_{s=1}^n X_s^2 \leq 1$, statement (1) follows from Lemma 3.4.

If (X_t) and (φ_t) are independent and $E|X_t/\sum_{s=1}^n X_s^2|^{\alpha_\varphi+\epsilon} < \infty$ for some $\epsilon > 0$, one can apply Lemma 3.4 to the tails of $\rho_{n,\varphi}$. An appeal to (3.6) and (3.8) ensures that this condition is satisfied, and therefore (2) follows. □

Remark 3.9. Sufficient conditions for condition (3.8) are given in Lemma 3.7.

Remark 3.10. Various expressions in Proposition 3.8 can be simplified if one assumes that $X_1, ..., X_n$ are *weakly exchangeable*, i.e., the distribution of $X_{\pi(1)}, \ldots, X_{\pi(n)}$ remains unchanged for any permutation $\pi(1), \ldots, \pi(n)$ of the integers $1, \ldots, n$. This condition is satisfied if (X_t) is an *exchangeable sequence.* This means that (X_n) is conditionally iid. If $X_1,...,X_n$ are weakly exchangeable, then, for example, (3.7) turns into

$$P(\rho_{n,\varepsilon}>x)\ \sim\ n\,P(\varepsilon>x)\,E\left[\frac{X_1^2}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varepsilon},$$

$$P(\rho_{n,\varepsilon}\le -x)\ \sim\ n\,P(\varepsilon\le -x)\,E\left[\frac{X_1^2}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varepsilon}.$$

Remark 3.11. If we assume in addition that (X_t) is stationary and ergodic, the strong law of large numbers applies to $(|X_t|^p)$ for any $p>0$ with $E|X|^p<\infty$. This, together with a dominated convergence argument, allows one to determine the asymptotic order of the tail balance parameters in Proposition 3.8. We restrict ourselves to $\rho_{n,\varepsilon}$; the quantities $\rho_{n,\varphi}$ can be treated analogously. Consider

$$m_n(\alpha_\varepsilon):=\sum_{i=1}^n E\left[\frac{X_t^2}{\sum_{s=1}^n X_s^2}\right]^{\alpha_\varepsilon}=E\left[\frac{n^{-\alpha_\varepsilon}\sum_{t=1}^n |X_t|^{2\alpha_\varepsilon}}{(n^{-1}\sum_{s=1}^n X_s^2)^{\alpha_\varepsilon}}\right]$$

Assume that $E\,|X|^{2\max(1,\alpha_\varepsilon)}<\infty$. Then the law of large numbers and uniform integrability imply that as $n\to\infty$,

$$m_n(\alpha_\varepsilon)\begin{cases}\to & 0 \quad \text{if } \alpha_\varepsilon>1,\\ = & 1 \quad \text{if } \alpha_\varepsilon=1,\\ \to & \infty \quad \text{if } \alpha_\varepsilon<1.\end{cases}$$

Proposition 3.8 provides sufficient conditions for regular variation of $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$. From this property we can derive our main result on the tails of the OLS estimator $\widehat{\beta}=\beta+\rho_{n,\varepsilon}+\rho_{n,\varphi}$ of the regression coefficient β.

Corollary 3.12. *If the conditions of Proposition 3.8 hold, then $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ are regularly varying with corresponding indices α_ε and α_φ. Moreover, if we assume that $\alpha_\varepsilon\ne\alpha_\varphi$ or that (ε_t) is independent of (φ_t) and $\alpha_\varepsilon=\alpha_\varphi$, then, as $x\to\infty$,*

$$P(\rho_{n,\varepsilon}+\rho_{n,\varphi}>x)\ \sim\ P(\rho_{n,\varepsilon}>x)+P(\rho_{n,\varphi}>x)\,,$$

$$P(\rho_{n,\varepsilon}+\rho_{n,\varphi}\le -x)\ \sim\ P(\rho_{n,\varepsilon}\le -x)+P(\rho_{n,\varphi}\le -x)\,,$$

where the corresponding asymptotic expressions for the tails of $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ are given in Proposition 3.8.

Proof. The regular variation of $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ is immediate from Proposition 3.8. If $\alpha_\varepsilon\ne\alpha_\varphi$, then the statement follows from the third part of Lemma 3.2. If $\alpha=\alpha_\varepsilon=\alpha_\varphi$ and (ε_t) and (φ_t) are independent, then the vector $(\varepsilon_1,...,\varepsilon_n,\varphi_1,...,\varphi_n)$ is regularly varying in $\mathbb{R}^{2n}$ with index α. Now a direct application of Lemma 3.4 yields the statement. □

3.2.2. *The noise is a linear process.* In the applications we also consider sequences (ε_t) and (φ_t) of dependent random variables. We assume that (ε_t) is a linear process, i.e., there exist real coefficients ψ_j and an iid sequence (Z_t) such that ε_t has representation

$$\varepsilon_t=\sum_{j=0}^\infty \psi_j Z_{t-j}=\sum_{j=-\infty}^t \psi_{t-j}Z_j\,,\quad t\in\mathbb{Z}\,. \tag{3.9}$$

The best known examples of (causal) linear processes are the ARMA and FARIMA processes; see for example Brockwell and Davis (1991). Throughout we assume that $Z=Z_1$ is regularly varying with index $\alpha_Z>0$ satisfying the tail balance condition

$$P(Z>x)=p\,\frac{L(x)}{x^{\alpha_Z}}\,(1+o(1))\quad\text{and}\quad P(Z\le -x)=q\,\frac{L(x)}{x^{\alpha_Z}}\,(1+o(1))\,,\quad x\to\infty\,, \tag{3.10}$$

for some $p, q \geq 0$ and $p + q = 1$. If the additional conditions

$$\begin{cases} \sum_{i=0}^{\infty} \psi_i^2 < \infty & \text{for some } \alpha_Z > 2\,, \\ \sum_{i=0}^{\infty} |\psi_i|^{\alpha_Z - \epsilon} < \infty & \text{for some } \alpha_Z \leq 2, \text{ some } \epsilon > 0\,, \\ EZ = 0 & \text{for } \alpha_Z > 1\,, \end{cases} \tag{3.11}$$

on the coefficients ψ_j and the distribution of Z hold, then (see Mikosch and Samorodnitsky (2000)) ε_t is regularly varying with index $\alpha_\varepsilon = \alpha_Z$ satisfying the relations

$$P(\varepsilon > x) = (1 + o(1))\, P(|Z| > x) \sum_{j=0}^{\infty} \Big[p\,(\psi_j^+)^{\alpha_Z} + q\,(\psi_j^-)^{\alpha_Z} \Big]\,, \quad x \to \infty\,, \tag{3.12}$$

and

$$P(\varepsilon \leq -x) = (1 + o(1))\, P(|Z| > x) \sum_{j=0}^{\infty} \Big[q\,(\psi_j^+)^{\alpha_Z} + p\,(\psi_j^-)^{\alpha_Z} \Big]\,, \quad x \to \infty\,, \tag{3.13}$$

where x^+ and x^- denote the positive and negative parts of the real number x. This means that ε_t is regularly varying with index $\alpha_\varepsilon = \alpha_Z$, and it is not difficult to show that the finite-dimensional distributions of (ε_t) are also regularly varying with the same index α_ε.

For further discussion we also assume that φ_t is a linear process with representation

$$\varphi_t = \sum_{j=-\infty}^{t} c_{t-j}\,\gamma_j\,, \quad t \in \mathbb{Z}\,, \tag{3.14}$$

where (γ_t) is an iid regularly varying sequence with index $\alpha_\gamma > 0$. Assuming (3.11) for (c_j) instead of (ψ_j) and the tail balance condition (3.10) for γ instead of Z, it follows that the finite-dimensional distributions of (φ_t) are regularly varying with index $\alpha_\varphi = \alpha_\gamma$ and the relations analogous to (3.12) and (3.13) hold for the left and right tails of φ_t.

Next, we investigate the tail behavior of the quantities $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ under the assumption that (ε_t) and (φ_t) are regularly varying linear processes. We give our basic **Assumptions:**

(1) (X_t) is a sequence of random variables with $X_t \neq 0$ a.s. for every t.
(2) (ε_t) is a linear process with representation (3.9), iid regularly varying noise (Z_t) with index $\alpha_\varepsilon > 0$ and coefficients ψ_j satisfying (3.10) and (3.11).
(3) (φ_t) is a linear process with representation (3.14), iid regularly varying noise (γ_t) with index $\alpha_\varphi > 0$ and coefficients c_j satisfying (3.10) (with ψ_j replaced by c_j) and (3.11) (with Z_j replaced by γ_j).
(4) (X_t) is independent of $((\varepsilon_t, \varphi_t))$.

The following result shows that $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ are regularly varying. Compare this result with Proposition 3.8 in the case of iid noise.

Proposition 3.13. *Assume that conditions* (1), (4) *hold. Fix* $n \geq 2$.

(1) *If* (2) *holds, as* $x \to \infty$, *then*

$$P(\rho_{n,\varepsilon} > x) \sim P(Z > x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2}{\sum_{t=1}^{n} X_t^2}\right]^{+}\right]^{\alpha_\varepsilon} + P(Z \le -x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2}{\sum_{t=1}^{n} X_t^2}\right]^{-}\right]^{\alpha_\varepsilon},$$

$$P(\rho_{n,\varepsilon} \le -x) \sim P(Z > x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2}{\sum_{t=1}^{n} X_t^2}\right]^{-}\right]^{\alpha_\varepsilon} + P(Z \le -x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2}{\sum_{t=1}^{n} X_t^2}\right]^{+}\right]^{\alpha_\varepsilon}.$$

(2) *If* (3) *and, in addition,* (3.8) *hold, then*

$$P(\rho_{n,\varphi} > x) \sim P(\gamma > x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} c_{t-j} X_t}{\sum_{t=1}^{n} X_t^2}\right]^{+}\right]^{\alpha_\varphi} + P(\gamma \le -x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} c_{t-j} X_t}{\sum_{t=1}^{n} X_t^2}\right]^{-}\right]^{\alpha_\varphi},$$

$$P(\rho_{n,\varphi} \le -x) \sim P(\gamma > x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} c_{t-j} X_t}{\sum_{t=1}^{n} X_t^2}\right]^{-}\right]^{\alpha_\varphi} + P(\gamma \le -x) \sum_{j=-\infty}^{n} E\left[\left[\frac{\sum_{t=\max(1,j)}^{n} c_{t-j} X_t}{\sum_{t=1}^{n} X_t^2}\right]^{+}\right]^{\alpha_\varphi}.$$

The following corollary gives our main result about the tail behavior of the OLS estimator $\widehat{\beta}$ of β in the case when both (ε_t) and (γ_t) constitute linear processes.

Corollary 3.14. *If the conditions of Proposition 3.13 hold, then $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ are regularly varying with corresponding indices α_ε and α_φ. Moreover, if we assume that $\alpha_\varepsilon \neq \alpha_\varphi$ or that (ε_t) is independent of (φ_t) and $\alpha_\varepsilon = \alpha_\varphi$, then, as $x \to \infty$,*

$$P(\rho_{n,\varepsilon} + \rho_{n,\varphi} > x) \sim P(\rho_{n,\varepsilon} > x) + P(\rho_{n,\varphi} > x),$$

$$P(\rho_{n,\varepsilon} + \rho_{n,\varphi} \le -x) \sim P(\rho_{n,\varepsilon} \le -x) + P(\rho_{n,\varphi} \le -x),$$

where the corresponding asymptotic expressions for the tails of $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ are given in Proposition 3.13.

3.2.3. *More general dependent noise.* In Section 3.2.2 we demonstrated how the results of Section 3.2.1 change under linear dependence. We focused on the linear process case because we were able to obtain explicit expressions for the asymptotic tail behaviors of $\rho_{n,\varepsilon}$, $\rho_{n,\varphi}$ and $\widehat{\beta}$. For more complicated dependence structures, the regular variation of these quantities follows by an application of Lemma 3.4, if the finite-dimensional distributions of the noise sequences (ε_t) and (φ_t) are regularly varying.

For example, if we assume that both (ε_t) and (φ_t) constitute GARCH processes, then the finite-dimensional distributions of these processes are regularly varying with positive indices, provided some mild conditions on the noise sequences of the GARCH processes hold. We refer to Basrak et al. (2002) for the corresponding theory on regular variation of GARCH processes. Alternatively, one can choose (ε_t) as a GARCH process with regularly varying finite-dimensional distributions and (φ_t) as a linear process (e.g. ARMA) with regularly varying finite-dimensional distributions, or vice versa. The index of regular variation of a GARCH process is a known function of the GARCH parameters and the distribution of the noise. For GARCH processes the asymptotic behavior of the tails of $\widehat{\beta}$ cannot be given in explicit form as for linear processes; see Proposition 3.13. However, we know from Lemma 3.4 that $\widehat{\beta}$ inherits the tail index of the minimum of α_φ and α_ε.

3.2.4. *Asymptotic normality.* Remark 3.11 suggests to consider the large sample properties of $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$. This will be done in this section. It turns out that, by virtue of the central limit theorem, the heavy tails of the ε_t's and φ_t's have no influence on the limit distribution. This is the content of the following result. In what follows, B is standard Brownian motion on $[0,\infty)$ and $\stackrel{d}{\rightarrow}$ refers to weak convergence in $\mathbb{D}[0,\infty)$ equipped with the J_1-metric. For more information on weak convergence in metric spaces, we refer to Billingsley (1968) or Jacod and Shiryaev (1987). For any random variable A we write $\sigma_A^2 = \operatorname{var}(A)$.

Proposition 3.15. *In addition to Assumptions* (1) – (4) *in Section 3.2.1 assume the following conditions:*

(1) (X_t) *is stationary and ergodic.*
(2) $EX = E\varepsilon = E\varphi = 0$.
(3) $EX^4 < \infty$.
(a) *If* $\alpha_\varepsilon > 2$, *then*

$$\widetilde{\rho}_{n,\varepsilon} := n^{1/2}\,(\rho_{[nt],\varepsilon})_{t\geq 0} \stackrel{d}{\rightarrow} \frac{(\sigma_\varepsilon^2 EX^4)^{1/2}}{\sigma_X^2}\,B\,,\quad n\to\infty\,. \tag{3.15}$$

(b) *If* $\alpha_\varphi > 2$ *then*

$$\widetilde{\rho}_{n,\varphi} := n^{1/2}\,(\rho_{[nt],\varphi})_{t\geq 0} \stackrel{d}{\rightarrow} \frac{\sigma_\varphi}{\sigma_X}\,B\,,\quad n\to\infty\,. \tag{3.16}$$

(c) *If* $\min(\alpha_\varepsilon,\alpha_\varphi) > 2$ *and the sequences* (ε_t), (φ_t) *are independent, then*

$$(\widetilde{\rho}_{[nt],\varepsilon} + \widetilde{\rho}_{[nt],\varphi})_{t\geq 0} \stackrel{d}{\rightarrow} \frac{(\sigma_\varepsilon^2 EX^4 + \sigma_\varphi^2\sigma_X^2)^{1/2}}{\sigma_X^2}\,B\,,\quad n\to\infty. \tag{3.17}$$

Proof. Since $\sigma_X^2 < \infty$, the ergodic theorem gives $n^{-1}\sum_{t=1}^n X_t^2 \to \sigma_X^2$ a.s. Hence

$$\widetilde{\rho}_{[n\cdot],\varepsilon} = (\sigma_X^{-2} + o_P(1))\frac{1}{\sqrt{n}}\sum_{i=1}^{[n\cdot]}\varepsilon_i X_i^2\,.$$

Since (X_t) and (ε_t) are independent, the sequence $(\varepsilon_t X_t^2)$ is a stationary ergodic martingale difference sequence with mean zero and finite variance $\sigma_\varepsilon^2 EX^4$. The functional central limit theorem for such sequences (see Billingsley (1968)) and a Slutsky argument yield (3.15). A similar argument gives (3.16). The same argument also applies to the sequence $(\varepsilon_t X_t^2 + \varphi_t X_t)$ which is a mean zero, variance $\sigma_\varepsilon^2 EX^4 + \sigma_\varphi^2\sigma_X^2$ stationary ergodic martingale difference sequence. □

Remark 3.16. Similar results can be obtained if $\alpha_\varepsilon < 2$ or $\alpha_\varphi < 2$. The limits are infinite variance stable processes with dependent increments. The limit theory then becomes much more delicate because point process convergence results enter. We refer to Davis and Hsing (1995) for more details.

A particular consequence of Proposition 3.15 and the continuous mapping theorem is the following:

Corollary 3.17. *Assume the conditions of part (c) of Proposition 3.15 hold. Then*

$$n^{1/2} \sup_{0\le t\le 1} |\rho_{[nt],\varepsilon} + \rho_{[nt],\varphi}| \stackrel{d}{\to} \frac{(\sigma_\varepsilon^2 EX^4 + \sigma_\varphi^2 \sigma_X^2)^{1/2}}{\sigma_X^2} \sup_{t\le 1} |B_t|\,. \tag{3.18}$$

Remark 3.18. The quantiles of the limit distribution are well-known from Kolmogorov-Smirnov type tests; see for example Shorack and Wellner (1986). One-sided versions of (3.18) can be derived in the same way.

4. A simulation study

We conduct a small simulation study to gain further insight into the theoretical results. We assume the model (2.1) with $\beta = 1$ and estimate β by the OLS estimator $\widehat{\beta}$ defined in (2.4).

4.1. **Pure multiplicative noise.** Consider the model

$$Y_t = (1+\varepsilon_t)X_t \tag{4.1}$$

satisfying the conditions:

(1) (X_t) is iid $N(1, 0.04)$.
(2) (ε_t) is iid $N(0, 0.01)$ or (ε_t) is iid Student-t distributed with ν degrees of freedom, $\nu = 2, 3, 4, 5, 10$, and rescaled such that the sample variance of the ε_t's is 0.01.
(3) (X_t) and (ε_t) are independent.

Under these conditions Corollary 3.12 applies if ε is Student$t$-distributed. Indeed, then ε is regularly varying with index ν, and we may conclude that $\widehat{\beta} = \beta + \rho_{n,\varepsilon}$ is regularly varying with index ν. Moreover, the tail asymptotics of $\rho_{n,\varepsilon}$ are described in part (1) of Proposition 3.8. Since the X_t's are iid and ε is symmetric, this means that

$$P(\rho_{n,\varepsilon} > x) = P(\rho_{n,\varepsilon} \le -x) \sim n\, E\left[\frac{X_1^2}{\sum_{s=1}^n X_s^2}\right]^\nu P(\varepsilon > x)\,.$$

We repeated 20,000 simulations of a time series $Y_1, \ldots, Y_n$ for $n = 25, 50, 100$. The results are reported as boxplots in Figures 4.1 and in the corresponding Table 4.2 with the sample mean, median, standard deviation, quartiles, minimum and maximum of the values $\widehat{\beta}$ coming from the 20,000 independent experiments for each parameter set.

The boxplots in Figure 4.1 and the sample characteristics reported in Table 4.2 show only slight differences between the cases of normal and heavy tailed multiplicative noise. There is a tendency for the distribution of the OLS estimator $\widehat{\beta}$ to be more spread for Student-t noise, in particular for noise with low degrees of freedom. This is indicated by the range of the data. The remaining characteristics such as median, mean, standard deviation and quartiles are very much the same in the heavy and light tailed cases.

For comparison, we included asymptotic 95% confidence bands which are based on the central limit theorem for $\widehat{\beta}$. For the most left boxplot, the central limit theorem with $\sqrt{n}$-rates does not apply to the distribution of $\widehat{\beta}$ since it has infinite variance. Therefore we indicate the asymptotic confidence bands only in the cases when $\widehat{\beta}$ has finite variance. They are obviously much smaller than the range prescribed by the whiskers and do not properly cover the set where $\widehat{\beta}$ assumes its values. For larger n, the whiskers and the inter-quartile range become smaller, but even for $n = 100$ the deviation of $\widehat{\beta}$ from its true value can be significant due to the wild sample behavior of $\widehat{\beta}$.

The simulation study shows that there is good and bad news. The good news is that the distributions of $\widehat{\beta}$ for heavy tailed and light tailed noise (ε_t) are not as different as one might expect from the theory in Section 3.2.1. The bad news is that the central limit theorem is not a

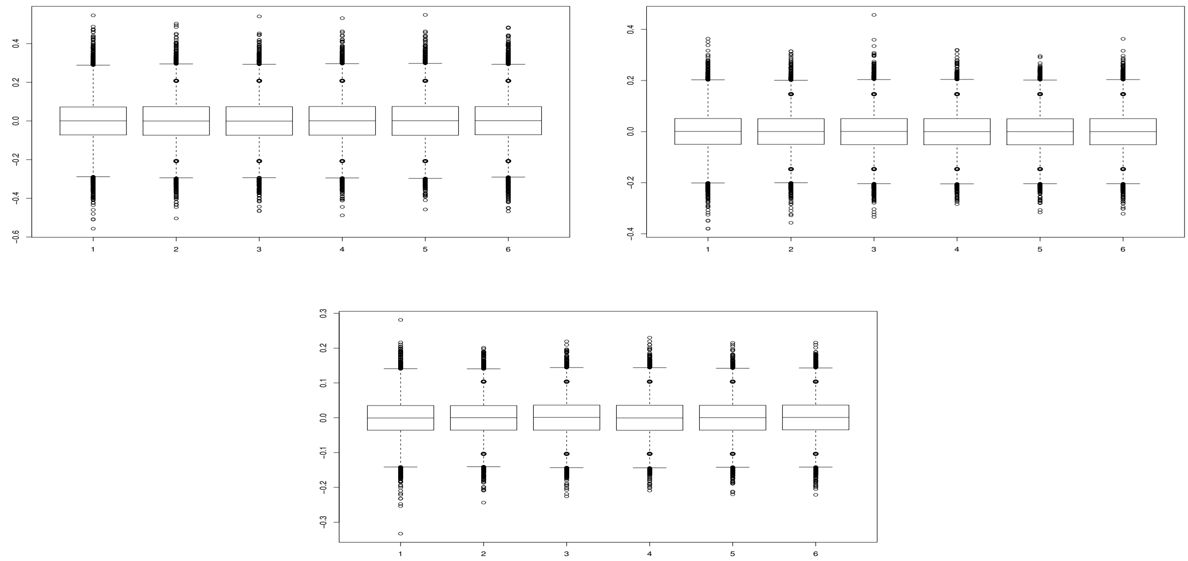

Figure 4.1. *Boxplots for* $20,000$ *repetitions of the OLS estimator* $\widehat{\beta}-1$ *estimated from the model* $Y_t = (1+\varepsilon_t)X_t$ *with pure multiplicative noise. In each graph, boxplots* $1-5$ *correspond to* ε_t *Student-t distributed with* $\nu = 2,3,4,5,10$ *degrees of freedom; plot No* 6 *corresponds to* ε_t *iid* $N(0,0.01)$. *The random variables* X_t *are iid* $N(1,0.04)$. Top left: $n=25$. Top right: $n=50$. Bottom: $n=100$. *The black dots in the* 5 *right boxplots indicate* 95% *asymptotic confidence bands based on the central limit theorem.*

	IID normal X_t's								
	Student-2			Student-3			Student-4		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.11	0.08	0.05	0.11	0.08	0.05	0.11	0.08	0.05
1st quartile	0.93	0.95	0.96	0.93	0.95	0.96	0.93	0.95	0.96
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.07	1.05	1.03	1.07	1.05	1.03	1.07	1.05	1.03
minimum	0.44	0.62	0.66	0.50	0.64	0.75	0.53	0.66	0.77
maximum	1.54	1.36	1.28	1.50	1.31	1.20	1.54	1.45	1.21
	Student-5			Student-10			Normal		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.11	0.08	0.05	0.11	0.08	0.05	0.11	0.08	0.05
1st quartile	0.93	0.95	0.96	0.93	0.95	0.96	0.92	0.95	0.96
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.08	1.05	1.04	1.08	1.05	1.04	1.07	1.05	1.04
minimum	0.51	0.71	0.80	0.54	0.68	0.78	0.53	0.68	0.78
maximum	1.53	1.32	1.23	1.54	1.29	1.21	1.48	1.36	1.21

Table 4.2. *Sample characteristics of* $\widehat{\beta}$ *for iid normal* $N(1,0.04)$ (X_t), *based on* $20,000$ *repetitions.*

good guideline as regards the distributional tails of $\widehat{\beta}$ even for sample sizes $n=50$ or 100, including the case where the noise is normally distributed.

We experimented with other models and distributions, including (X_t) iid uniform, (X_t) Gaussian AR(1) and (ε_t) with distributions with different variances. These simulation studies showed the same qualitative behavior: there is hardly any difference between the distributions of $\rho_{n,\varepsilon}$ for t-distributed and for normal noise (ε_t).

In order to mimic the behavior of regression coefficient estimators $\widehat{\beta}$ based on real-life data, the discussions in Sections 1 and 5 show that these estimators vary wildly across samples, we drop the assumption (2) regarding the independence of the ε_t's in model (4.1). We retain the assumptions (1) and (3).

	IID Normal X_t's								
	Student-2			Student-3			Student-4		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.36	0.24	0.19	0.25	0.19	0.17	0.24	0.19	0.17
1st quartile	0.85	0.87	0.88	0.85	0.87	0.88	0.85	0.87	0.88
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.15	1.13	1.12	1.15	1.13	1.12	1.15	1.13	1.12
minimum	-5.33	-9.88	-3.97	-1.40	-1.62	-1.67	-3.28	-0.42	-4.91
maximum	20.02	10.68	4.65	4.63	3.52	2.21	4.10	2.15	2.13
	Student-5			Student-10			Normal		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.23	0.19	0.17	0.23	0.19	0.17	0.23	0.19	0.17
1st quartile	0.85	0.87	0.88	0.85	0.87	0.88	0.84	0.87	0.88
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.15	1.13	1.12	1.15	1.13	1.12	1.15	1.13	1.12
minimum	-0.71	-0.04	0.19	-0.44	0.07	0.32	-0.27	0.14	0.34
maximum	2.31	2.35	2.59	2.29	2.16	2.11	2.29	1.85	1.64

Table 4.3. *Sample characteristics of $\widehat{\beta}$ for iid normal $N(0,0.04)$ (X_t), autoregressive $\varepsilon_t = 0.7X_{t-1} + Z_t$ with iid student noise (Z_t) with $\nu = 2,3,4,5,10$ degrees of freedom, scaled to variance 0.01. In Column 6 one finds the results for Gaussian autoregressive (ε_t) with the same autoregressive coefficients and variance as in the Student case. The simulations are based on 20,000 repetitions.*

We model the dependence of the ε_t's by an autoregressive process of order 1. This means that

$$\varepsilon_t = \gamma\,\varepsilon_{t-1} + Z_t\,, \quad t \in \mathbb{Z}\,,$$

where (Z_t) is an iid regularly varying sequence, as described in Section 3.2.2. We choose Z_t to be Student-t distributed with ν degrees of freedom, resulting in regular variation of Z_t, ε_t and $\rho_{n,\varepsilon}$ with index ν, and $\gamma = 0.7$. From the first part of Proposition 3.13 we have that

$$P(\rho_{n,\varepsilon} > x) = P(\rho_{n,\varepsilon} \le -x) \quad \sim \quad P(Z > x) \sum_{j=-\infty}^{n} E\left[\frac{\sum_{t=\max(1,j)}^{n} \gamma^{t-j} X_t^2}{\sum_{t=1}^{n} X_t^2}\right]^{\nu}.$$

Notice that the factor after $P(Z > x)$ is larger than in the case of iid ε_t. Therefore the heavy tails of ε have a stronger influence on the tails of $\rho_{n,\varepsilon}$. This is illustrated in Table 4.3 and Figure 4.4. For comparison we also include Figure 4.5 where both (X_t) and (ε_t) constitute dependent stationary sequences. The combination of heavy tails and dependence in the innovations ε make the distribution of $\rho_{n,\varepsilon}$ more spread out than their Gaussian counterparts; and the distribution is the more spread out the smaller the index of regular variation. This observation applies for samples of small and moderate size. From both figures, but also from Table 4.3 we see that the characteristics of the centers of the distributions of $\rho_{n,\varepsilon}$ exhibit very similar behavior for t-distributed and Gaussian dependent noise (ε_t).

Finally, we mention that other models for (ε_t) which imply both dependence and regularly varying finite-dimensional distributions (such as GARCH processes, see Section 3.2.3) yield the same qualitative behavior for the tails of $\rho_{n,\varepsilon}$. However, if the dependence in (ε_t) does not range over a sufficiently large period of time (such as in an MA(q) process with small q and small coefficients) one observes that the tails of $\rho_{n,\varepsilon}$ do not differ very much in the cases of heavy and light tails. To conclude, it is not the heavy tails 'an sich', but the combination with persistence in the noise by which outliers affect the OLS estimator in samples of moderate size. Larger sample estimators in sizes countervail the persistence effect.

4.2. **Pure additive noise.** Next, consider the model

$$Y_t = X_t + \varphi_t \,,$$

satisfying the conditions

(1) (X_t) is iid $N(1, 0.04)$.
(2) (φ_t) is iid $N(0, 0.09)$ or (φ_t) is iid Student-t distributed with ν degrees of freedom, $\nu = 2, 3, 4, 5, 10$, and rescaled such that the sample variance of the φ_t's is 0.09.
(3) (X_t) and (φ_t) are independent.

If φ is Student-t distributed, Corollary 3.14 applies. In particular, we conclude from part (2) of Proposition 3.8 that

$$P(\rho_{n,\varphi} > x) = P(\rho_{n,\varphi} \le -x) \sim P(\varphi > x)\, n\, E\left[\frac{|X_1|}{\sum_{s=1}^n X_s^2}\right]^\nu .$$

We repeated 20,000 simulations of a time series $Y_1, \dots, Y_n$ for $n = 25, 50, 100$. The results are reported as boxplots in Figure 4.6 and in the corresponding Table 4.7 with the sample mean, median, standard deviation, quartiles, minimum and maximum of the values $\widehat{\beta}$ coming from the 20,000 independent experiments for each parameter set.

	IID normal X_t's and iid φ_t's								
	Student-2			Student-3			Student-4		
sample size	25	50	100	25	50	100	25	50	100
mean	0.96	0.98	0.99	0.96	0.98	0.99	0.96	0.98	0.99
s.d.	0.30	0.21	0.15	0.30	0.21	0.15	0.30	0.21	0.15
1st quartile	0.75	0.83	0.89	0.75	0.84	0.89	0.75	0.84	0.89
median	0.96	0.98	0.99	0.96	0.98	0.99	0.96	0.98	0.99
3rd quartile	1.16	1.12	1.09	1.16	1.12	1.09	1.17	1.13	1.09
minimum	-0.15	0.03	0.40	-0.17	0.14	0.39	-0.32	0.15	0.38
maximum	2.41	1.87	1.60	2.38	1.92	1.61	2.19	1.90	1.58
	Student-5			Student-10			Normal		
sample size	25	50	100	25	50	100	25	50	100
mean	0.96	0.98	0.99	0.96	0.98	0.99	0.96	0.98	0.99
s.d.	0.30	0.21	0.15	0.30	0.21	0.15	0.30	0.21	0.15
1st quartile	0.75	0.83	0.89	0.75	0.84	0.89	0.76	0.84	0.89
median	0.96	0.98	0.99	0.96	0.98	0.99	0.96	0.98	0.99
3rd quartile	1.17	1.12	1.09	1.17	1.12	1.09	1.16	1.12	1.09
minimum	-0.75	0.13	0.38	-0.40	0.06	0.45	-0.31	0.06	0.41
maximum	2.11	2.05	1.60	2.14	1.89	1.51	2.29	2.18	1.65

Table 4.7. *Sample characteristics of $\widehat{\beta}$ for iid normal $N(1, 0.04)$ (X_t), based on 20,000 repetitions.*

Figure 4.6 and Table 4.7 indicate that the OLS estimator $\widehat{\beta}$ oscillates wildly but it is difficult to see clear qualitative differences between the cases of heavy tailed and normally distributed noise. In all cases, $\widehat{\beta}$ is downward biased and the 95% asymptotic confidence bands prescribed by the central limit theorem are too optimistic. The picture does not change significantly as long as the φ_t's are independent. We experimented with (X_t) iid uniform or Gaussian AR(1) and (φ_t) iid normal or t-distributed; the outcome was again very much as reported above.

In order to study the influence of dependence in the sequence (φ_t) on the OLS estimator, we consider a GARCH(1,1) process

$$\varphi_t = \sigma_t\, Z_t \,, \quad \sigma_t^2 = 10^{-5} + \alpha_1\, X_{t-1}^2 + \beta_1\, \sigma_{t-1}^2 \,,$$

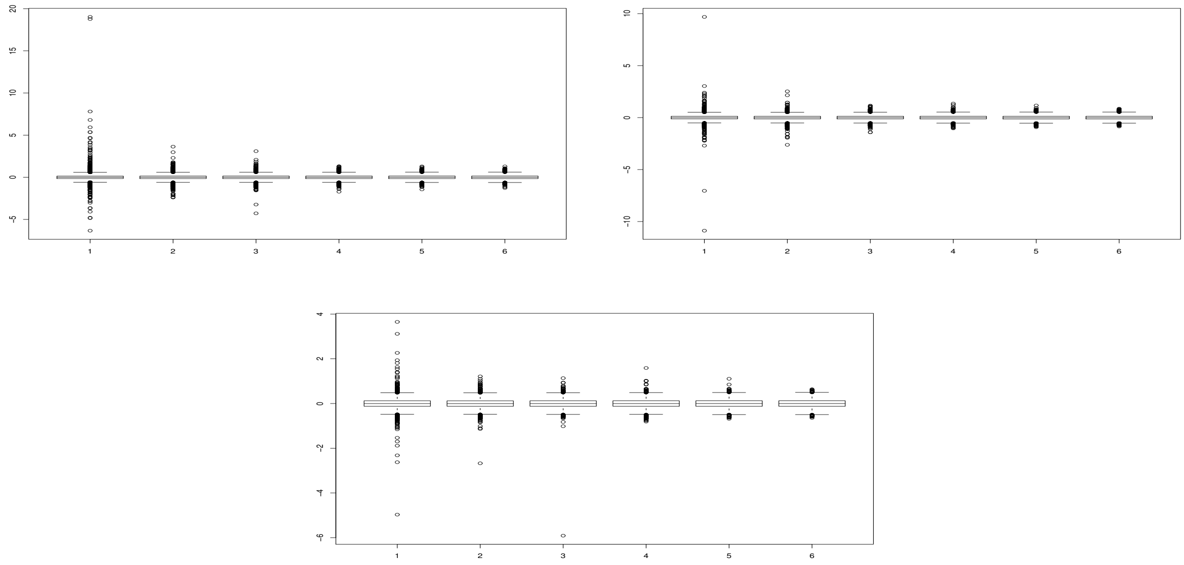

Figure 4.4. *Boxplots for* $20,000$ *repetitions of the OLS estimator* $\widehat{\beta}-1$ *estimated from the model* $Y_t = (1+\varepsilon_t)X_t$ *with pure multiplicative noise. In each graph, boxplots* $1-5$ *correspond to the autoregressive process* $\varepsilon_t = 0.7\varepsilon_{t-1} + Z_t$*, where* (Z_t) *is iid Student-t distributed with* $\nu = 2,3,4,5,10$ *degrees of freedom, and* ε_t *is rescaled to empirical variance* 0.01*; plot No* 6 *corresponds to Gaussian autoregressive* ε_t *with the same autoregressive coefficient* 0.7 *and variance* 0.01*. The random variables* X_t *are iid* $N(1,0.04)$. Top left: $n = 25$. Top right: $n = 50$. Bottom: $n = 100$.

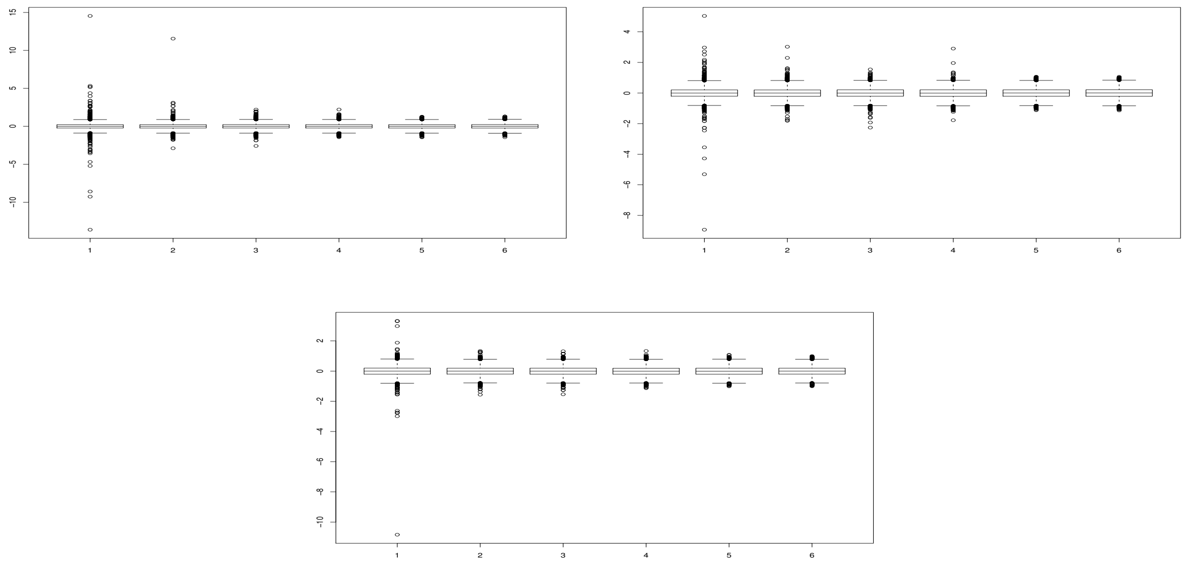

Figure 4.5. *Boxplots for* $20,000$ *repetitions of the OLS estimator* $\widehat{\beta}-1$ *estimated from the model* $Y_t = (1+\varepsilon_t)X_t$ *with pure multiplicative noise. In each graph, boxplots* $1-5$ *correspond to the autoregressive process* $\varepsilon_t = 0.7\varepsilon_{t-1} + Z_t$*, where* (Z_t) *is iid Student-t distributed with* $\nu = 2,3,4,5,10$ *degrees of freedom, and* ε_t *is rescaled to empirical variance* 0.01*; plot No* 6 *corresponds to Gaussian autoregressive* ε_t *with the same coefficient* 0.7 *and variance* 0.01*. The random variables* X_t *are autoregressive* $X_t = 0.5X_{t-1} + \gamma_t$*, where* (γ_t) *is iid centered Gaussian, and* X_t *is rescaled to variance* 0.04. Top left: $n = 25$. Top right: $n = 50$. Bottom: $n = 100$.

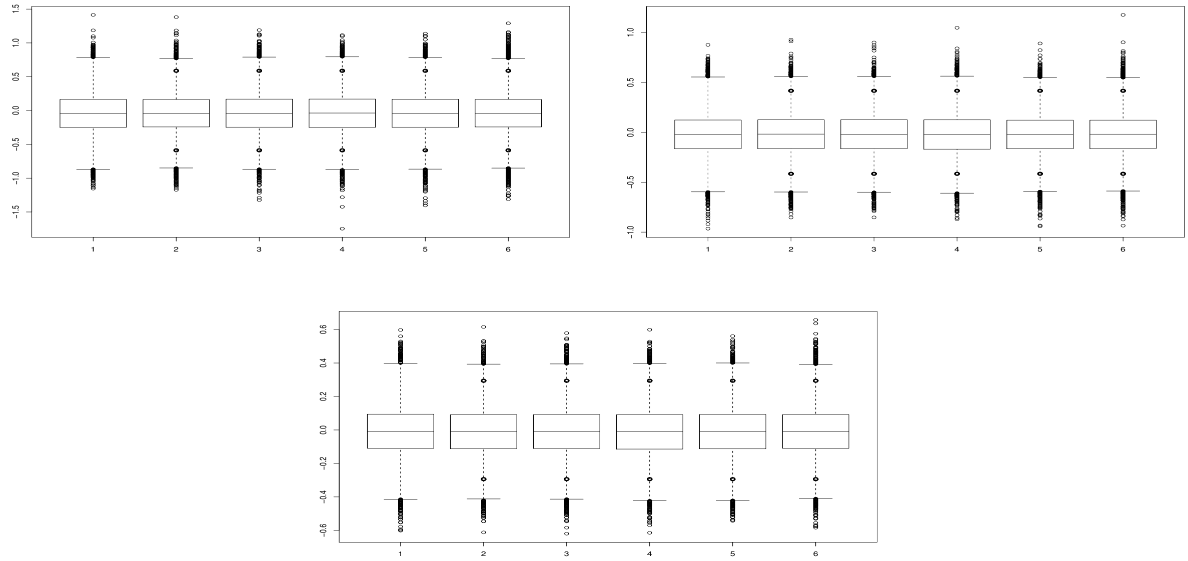

Figure 4.6. *Boxplots for* $20,000$ *repetitions of the OLS estimator* $\widehat{\beta}-1$ *estimated from the model* $Y_t = X_t+\varphi_t$ *with pure additive noise. In each graph, boxplots* $1-5$ *correspond to* φ_t *Student-t distributed with* $\nu = 2,3,4,5,10$ *degrees of freedom rescaled to empirical variance* 0.09*; plot No* 6 *corresponds to* φ_t *iid* $N(0,0.09)$*. The random variables* X_t *are iid* $N(1,0.04)$*.* Top left: $n = 25$. Top right: $n = 50$. Bottom $n = 100$. *The black dots in the* 5 *right boxplots indicate* 95% *asymptotic confidence bands based on the central limit theorem.*

for iid $N(0,1)$ noise (Z_t). It follows for example from Basrak et al. (2002) that φ is regularly varying with index ν which is given as the unique solution to the equation

$$E|\alpha_1 Z^2+\beta_1|^{\nu/2} = 1\,.$$

This equation can be solved numerically. We consider the following five sets of parameter values:

$$\begin{aligned}
\alpha_1 &= 0.1\,, & \beta_1 &= 0.900\,, & \nu &= 2\,,\\
\alpha_1 &= 0.1\,, & \beta_1 &= 0.895\,, & \nu &= 3\,,\\
\alpha_1 &= 0.1\,, & \beta_1 &= 0.890\,, & \nu &= 4\,,\\
\alpha_1 &= 0.1\,, & \beta_1 &= 0.884\,, & \nu &= 5\,,\\
\alpha_1 &= 0.1\,, & \beta_1 &= 0.840\,, & \nu &= 10\,.
\end{aligned}$$

We simulated 20,000 realizations of GARCH samples with these parameters. Then we scaled the samples to empirical variance 0.09. For comparison we also drew an iid $N(0,0.09)$ sample of φ_t's. Independently of (φ_t) we drew iid $N(0,0.04)$ random variables X_t. Subsequently we calculated the 20,000 values of the OLS coefficient estimates. The boxplots of this experiment are reported in Figure 4.9 with the corresponding Table 4.8. There is a clear difference between the heavy tailed cases with GARCH additive noise and the case of iid normal additive noise. The distributions are much more spread in the case of GARCH noise. Surprisingly, there are hardly any differences between the distributional characteristics of the OLS estimators for the different GARCH models. But the difference with the iid normal innovations is considerable. To conclude, both in the model with multiplicative and the model with additive noise, the OLS coefficient estimates vary wildly in medium sized samples if the noise process is a cocktail of heavy tailed distributed innovations which exhibit some persistence.

	IID normal X_t's and GARCH φ_t's								
	$\nu=2$			$\nu=3$			$\nu=4$		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.32	0.22	0.15	0.32	0.22	0.15	0.32	0.22	0.15
1st quartile	0.78	0.85	0.90	0.78	0.85	0.90	0.78	0.85	0.90
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.21	1.15	1.10	1.21	1.15	1.10	1.21	1.15	1.10
minimum	-0.91	0.22	0.42	-0.90	0.22	0.40	-0.90	0.22	0.40
maximum	2.19	2.00	1.61	2.19	2.00	1.62	2.19	2.00	1.63
	$\nu=5$			$\nu=10$			Normal		
sample size	25	50	100	25	50	100	25	50	100
mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
s.d.	0.32	0.22	0.15	0.32	0.22	0.15	0.16	0.21	0.08
1st quartile	0.78	0.85	0.90	0.78	0.85	0.90	0.90	0.93	0.95
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3rd quartile	1.21	1.15	1.10	1.21	1.15	1.10	1.11	1.07	1.05
minimum	-0.90	0.22	0.39	-0.87	0.22	0.38	0.24	0.50	0.68
maximum	2.19	2.00	1.64	2.20	1.99	1.65	1.63	1.41	1.34

Table 4.8. *Sample characteristics of* $\widehat{\beta}$ *for iid normal* $N(1,0.04)$ (X_t) *and GARCH* (φ_t) *with tail parameter* ν*, based on* 20,000 *repetitions.*

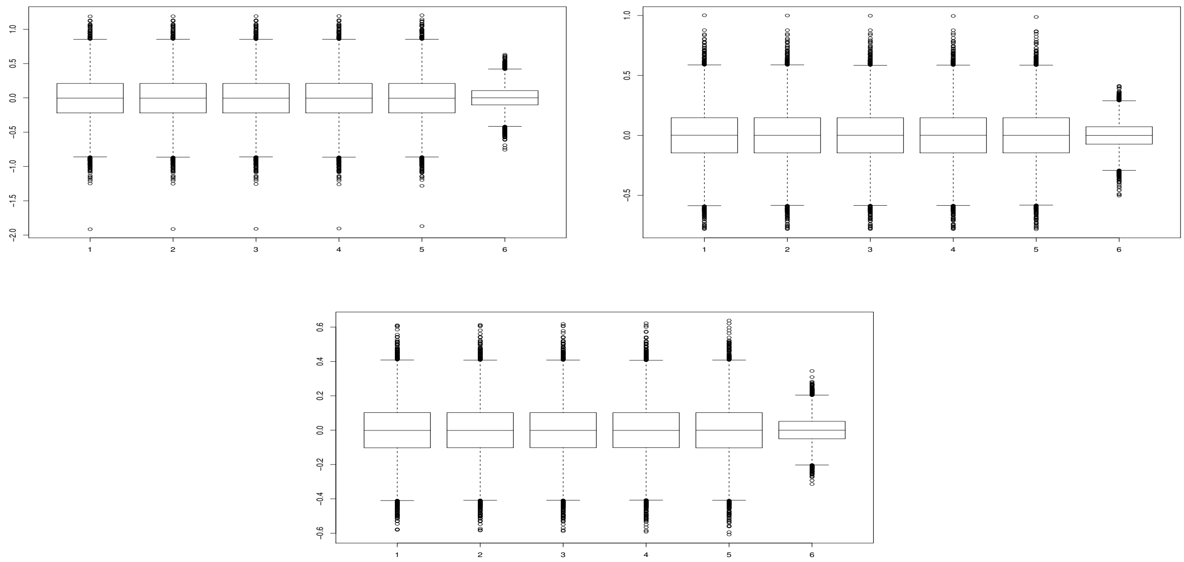

Figure 4.9. *Boxplots for* 20,000 *repetitions of the OLS estimator* $\widehat{\beta}-1$ *estimated from the model* $Y_t = X_t+\varepsilon_t$ *with pure additive noise. In each graph, boxplots* $1-5$ *correspond to* (ε_t) *from a GARCH process with tail parameter* $\nu=2,3,4,5,10$ *scaled to empirical variance* 0.09*; plot No* 6 *corresponds to* ε_t *iid* $N(0,0.09)$*. The random variables* X_t *are iid* $N(1,0.04)$. Top left: $n=25$. Top right: $n=50$. Bottom: $n=100$.

5. Application

The motivation for this study comes from the empirical fact that in various applications involving small and medium size samples of financial data there appears a wide variation in reported coefficient estimates across different samples. One possible explanation for this variation is the regular variation of the distributions of the innovations in combination with some dependence. As a consequence, the abatement in coefficient variability may require much longer time series than if the data came from a normal distribution. Both from an estimation point of view and for policy making it is important to capture the uncertainty in the estimates, as standard central limit type of error bands can be quite misleading in smaller samples with dependent heavy tailed noise. The previous simulation study demonstrated this claim on artificial data. We now turn to an economic application to demonstrate that this theory also has a bite in explaining the observed variability of empirical coefficient estimates.

The application focuses on the yield curve. Mankiw and Miron (1986) report typical slope estimates of the expectations coefficient in yield curve regressions for quite different samples. There are two important data features. First, the point estimates come in a wide range. Only in one of their five samples, with about 80 observations each, the point estimates come close to the benchmark theoretical slope of two (zero term premium). Second, all reported point estimates are less than two, and two slope estimates are even negative (one is significantly negative). Fama and Bliss (1987) emphasize the variability of the coefficient estimates. A sizeable literature has focused on the apparent downward bias. While the latter issue is not the focus of this paper, we need to address it before we can zoom in on the variability question. The first subsection introduces the topic from a simple three period investment decision problem. The model includes a risk premium and a liquidity premium, which potentially could account for the variability and downward difference from two. Subsequently, the second subsection discusses the data and the results from standard time series and panel regressions, basically replicating the results from previous studies. The last subsection provides evidence for the coefficient variability stemming from dependent heavy tailed noise by means of cross section regressions.

5.1. **Economic theory.** To introduce the issue of yield curve estimation, consider a three period investment problem under uncertainty:

$$
\begin{aligned}
\max EU(x, X_i) &= \pi V(x) + (1-\pi)\Sigma_i \pi_i \rho V(X_i), \;\; \Sigma_i \pi_i = 1 \\
\text{subject to } w &= d + b, \\
x &= (1+r)d + (1+q)b, \\
X_i &= (1+s)b + (1+R_i)(1+r)d\,, \quad i = 1, ..., n.
\end{aligned}
\tag{5.1}
$$

Here $V(.)$ and $\rho V(.)$ are the strictly concave first and second period utility functions; the expected utility function $EU(x, X_i)$ is time separable. The pure rate of time preference is $\rho \leq 1$. There are two types of uncertainty. The first type of uncertainty stems from the uncertain liquidity needs of agents as in Diamond and Dybvig (1983): with probability π an agent wants to consume early, and with probability $1-\pi$ the agent finds out he desires to consume late during the third period. The other uncertainty, π_i, pertains to the return on the second period short term bond. At the beginning of the first period, wealth w can be invested in a one-period bond d yielding $1+r$ in the second period and a two-period zero-coupon bond b yielding s in the third period. The one-period bond investment can be rolled over into a new one-period bond investment if one turns out to be a late consumer. All bonds are default-free government issues. The interest rate on the second period short term bond is uncertain at the time when the first investment decision has to be taken; R_i materializes with probability π_i, and there are n states of the world $i = 1, ..., n$. If the consumer has early liquidity needs, the long term bond investment has to be liquidated early. This comes at a cost, and we assume that $q < r$, where $r - q$ is the liquidation premium. The costs of early liquidation are due to irreversibilities of longer term capital investments and transactions costs, see Diamond and Dybvig (1983). If $(1+s) > (1+r)E[1+R_i]$, a risk averse agent will want to hedge against liquidity needs by partly investing in the long term bond and partly investing in the short term bond (no short sale condition).

The first order conditions for a maximum to the problem (5.1) imply the following pricing kernel relationship

$$E[M_i(1+R_i)] = \frac{q-r}{1+r} + E\left[M_i\frac{1+s}{1+r}\right], \tag{5.2}$$

where

$$M_i = \rho\frac{1-\pi}{\pi}\frac{\partial V(X_i)/\partial X_i}{\partial V(x)/\partial x}$$

is the intertemporal marginal rate of substitution or discount factor, for short. Under risk neutrality the discount factor reduces to $\rho(1-\pi)/\pi$. Assuming that the second moment of R_i is bounded, we can use the covariance $\mathrm{cov}(M_i, R_i - r)$ between the discount factor and the short term interest rate innovations to rewrite (5.2). First expand (5.2) as follows:

$$E\left[M_i\left((R_i - r) + (r+1)\right)\right] = \frac{q-r}{1+r} + \frac{1+s}{1+r}EM_i\,,$$

and subsequently use the covariance to obtain

$$E[R_i - r] = -(1+r) + \frac{1+s}{1+r} + \frac{1}{EM_i}\frac{q-r}{1+r} - \frac{\mathrm{cov}(M_i, R_i - r)}{EM_i}\,. \tag{5.3}$$

Let y be the 1-period yield on the two period bond, i.e., $(1+y)^2 = (1+s)$. Then (5.3) can be restated as

$$E[R_i - r] = 2(y-r) + T(y,r) + P(M_i, R_i, q), \tag{5.4}$$

where

$$P(M_i, R_i, q) = \frac{1}{EM_i}\frac{q-r}{1+r} + \frac{\mathrm{cov}(-M_i, R_i - r)}{EM_i} \quad \text{and} \quad T(y,r) = \frac{(y-r)^2}{1+r}\,.$$

Here P is the term premium and T is a convexity term. The term premium consists of two parts, a liquidity premium $(q-r)/[(1+r)EM_i]$ and a risk premium $\mathrm{cov}(-M_i, R_i - r)/EM_i$. Confusingly, in the applied yield curve literature the risk premium is sometimes referred to as the liquidity premium.

The liquidity premium is negative, since $q < r$, which reflects the costs of liquidating the long term investment early. While this liquidity premium is often studied in the banking literature, it is mostly ignored in the literature on the yield curve. It is sometimes discussed under the heading of segmented market hypothesis or the less absolute preferred habitat theory; but these do not consider the liquidity needs as uncertain, rather in this literature different agents have a priori different investment horizons. This segmentation results in possible mismatches between supply and demand for bonds within a certain maturity range. The yield curve literature traditionally focuses on the risk premium $\mathrm{cov}(-M_i, R_i - r)/EM_i$. If agents are risk neutral so that $M_i = \rho(1-\pi)/\pi$, then the risk premium drops out and the term premium (5.4) reduces to the liquidity premium

$$P = \pi(q-r)/\rho(1+r)(1-\pi) < 0. \tag{5.5}$$

If agents are risk averse, however, the risk premium part is positive, so that P can have either sign. To show this, assume e.g. that agents have a power utility function

$$V(x) = \frac{1}{1-\gamma}x^{1-\gamma},\ \gamma > 0.$$

Some manipulation yields

$$\begin{aligned}\frac{\pi}{1-\pi}\frac{1}{\rho}\left(\frac{(1+r)d}{x}\right)^{\gamma}\operatorname{cov}(-M_i, R_i - r) &= E\left[\frac{R_i}{-(k+R_i)^{\gamma}}\right] - E\left[\frac{1}{-(k+R_i)^{\gamma}}\right] ER_i \\ &= E\left[\frac{k+R_i}{-(k+R_i)^{\gamma}}\right] - E\left[\frac{1}{-(k+R_i)^{\gamma}}\right] E[R_i + k]\,,\end{aligned}$$

where

$$k = 1 + \frac{1+s}{1+r}\frac{b}{d}\,.$$

Note that $k > 1$ if there is no short selling (which we will assume). For any random variable $X \geq 0$ and a monotone non-decreasing function $g(x)$ it follows from association that the relation $E[X\,g(X)] \geq EX\,Eg(X)$ holds; see Resnick (1987), Lemma 5.32(iv). Taking $X = k + R_i \geq 0$, and $g(x) = -x^{-\gamma}$ so that $g'(x) = \gamma x^{-\gamma-1} \geq 0$ whenever $x \geq 0$, it follows that $\operatorname{cov}(-M_i, R_i - r) \geq 0$. Thus the term premium P can be of either sign, since the liquidity premium is negative and the risk premium is positive.

A linear yield curve results if $P = 0$ and if the convexity term is absent as well. Note that since typically both r and y are close to zero, so that $1 + r$ can be ignored in the convexity term, $T(y, r) = o(y - r)$. Hence little is lost by setting $T(y, r) = 0$ in empirical work (except for hyperinflation economies). But the benchmark case of $P = 0$ which is referred to in most empirical studies is more contentious, since even if agents are approximately risk neutral the liquidity premium turns the term premium negative.

The validity of model (5.4) cum $P = 0$ for the term structure of interest rates is often tested by whether $H_0 : \beta = 2$ can be rejected in a linear regression of $R_i - r$ on $y - r$:

$$R_i - r = \theta + \beta(y - r) + \varepsilon, \quad E\varepsilon = 0. \tag{5.6}$$

The hypothesis $\beta = 2$ is known as the *Expectations Hypothesis.* It is typical for the research in this area that the null hypothesis $H_0 : \beta = 2$ in (5.6) is soundly rejected, yielding evidence for a non-zero term premium P. Time series analyses such as presented in Mankiw and Miron (1986) almost always yield coefficients below two, and sometimes even negative coefficients are reported. Specifically, this indicates that the term premium $P(M_i, R_i, q)$ is negatively correlated with the spread $y - r$. Possibly, this stems from the negative liquidity premium, see below.

In the application, we first re-run the specification (5.6) for a number of different countries and basically find the same results as reported in the literature. To rule out any contamination from the convexity term, we also estimate the specification

$$R_i - r = \theta + \beta(y - r) + \tau\frac{(y-r)^2}{1+r} + \varepsilon. \tag{5.7}$$

Since the model (5.7) applies to each country individually, one can also investigate the model relative to a benchmark country. Due to the negative correlation between the unobserved term premium and the spread, the relative specification may diminish the omitted variable bias. Thus we also estimate

$$\widetilde{R}_i - \widetilde{r} = \theta + \beta(\widetilde{y} - \widetilde{r}) + \tau\widetilde{T}(y, r) + \widetilde{\varepsilon}, \tag{5.8}$$

where $\tilde{X} = X - X^*$, and X^* denotes the benchmark country variable.

To show that this specification may help reduce the omitted variable bias, note that

$$\operatorname{cov}(X - X^*, P - P^*) = \operatorname{cov}(X, P) + \operatorname{cov}(X^*, P^*) - \operatorname{cov}(X, P^*) - \operatorname{cov}(X^*, P), \tag{5.9}$$

and where $X = y - r$. If $\operatorname{cov}(X, P) + \operatorname{cov}(X^*, P^*)$ is negative, the bias is reduced if $-\operatorname{cov}(X, P^*) - \operatorname{cov}(X^*, P)$ is positive. This happens for example if countries experience simultaneously similar movements in their yield curves, so that P (and P^*) also co-vary negatively with $y^* - r^*$ (and $y - r$ respectively).

In the second part of the empirical application we investigate the cross section properties of the model. Assuming that the β's are more or less identical across countries at any point in time, i.e., they have a common component, but vary over time, there is value in estimating the expectations hypothesis model at any point in time. Subsequently, from the country cross section regressions we try to infer the properties of the distribution of the innovations and the variability of the error terms $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ in particular.

5.2. **Time series regressions and downward bias.** We obtained data on the one month and two month interest rates from February 1995 to December 1999 for 14 countries, yielding 59 observations per country. The countries with the abbreviations used in the tables are respectively Belgium (Bel), Canada (Can), Germany (Ger), Denmark (Den), France (Fr), United Kingdom (UK), Italy (It), Japan (Jap), Austria (Aus), Portugal (Por), Sweden (Swe), United States of America (USA), Switzerland (Swit), and The Netherlands (Neth). The data are beginning of the month figures on short term treasury paper available from datastream. Since for the purpose of the paper we are interested in running both time series, panel and cross section regressions, we are squeezed by data availability. Only since the middle of the 1990's there are more than 10 countries which report such rates. By the end of this decade, though, we are confronted by the data squeeze due to the European monetary unification process, which implies that at the short end of the yield curve rates became about equal. In Table 5.1 we report the per country regression coefficient estimates on the

count	OLS	relative	relative & convex	pool & relative	pool & relative & convex
Bel	1.01	1.86	0.58	0.99	0.54
Can	1.20	1.91	2.14	1.98	2.09
Ger	1.28	0.97	0.22	1.00	0.47
Den	2.02	2.21	2.27	1.44	1.93
Fr	1.99	2.89	2.96	2.68	3.35
UK	0.87	1.37	1.35	1.15	1.28
It	0.98	1.30	1.27	1.22	0.97
Jap	0.95	1.02	1.26	1.15	1.28
Aus	0.51	1.39	1.40	1.25	1.46
Por	0.54	0.70	0.42	0.42	0.49
Swe	0.84	1.29	1.16	1.05	1.10
USA	0.36	0.91	1.10	0.66	1.09
Swit	1.15	1.01	0.99	1.00	0.91
Neth	1.22	/	/	/	/
mean	1.06	1.44	1.31	1.23	1.30
range	1. 66	2. 19	2. 74	2. 26	2. 88
s.d.	0.48	0.61	0.76	0.57	0.79

Table 5.1. *Slope estimates for expectations hypothesis model. The table records coefficient estimates for the slope coefficient of the expectations model. The first column contains the per country OLS results for* (5.6)*; column* 2 *for* (5.8) *without the convexity term and column* 3 *for* (5.8)*. The base country in the regressions is The Netherlands.*

slope coefficient for the yield differential $y - r$ using different regression techniques.

The first column gives the results for specification (5.6). The next column labeled relative is based on (5.8) but without the convexity term, and where The Netherlands is taken as the base country. The third column labeled relative & convex is based on the full specification of (5.8) including the convexity term $T(y, r)$. The last two columns repeat the same exercise but use panel regressions in which the data are pooled. The table conveys two main results. (i) The β-coefficients are almost all significantly positive, but hover more closely around 1, rather than around the Expectations Hypothesis value 2. (ii) There appears to be quite a considerable variation in the coefficient estimates as can be seen from the range statistic, which is the difference between the highest and the lowest estimated value. These results thus corroborate the results reported previously in the literature, although in our sample we do not find any negative coefficients. Not reported are the intercept estimates which were invariably very small and never significantly different from zero. We discuss the average value of the point estimates and their variability more extensively.

Since most of the β-coefficients and their averages are well below the Expectations Hypothesis value of two, the economic interpretation is that the term premium $P(M_i, R_i, q)$ is negatively correlated with $(y-r)$, causing a downward omitted variable bias in the β-estimates. To understand what may happen, consider the relationship between the observed spread $(y-r)$ and the term premium $P(M_i, R_i, q)$ in some detail in the benchmark case when agents are risk neutral. Thus only the liquidity premium (5.5) is present. Assume that (5.4) applies and that $T(y, r)$ is negligible small. Then, since the intercept is zero, approximately

$$\widehat{\beta} = 2 + \Sigma(y-r)\left(\frac{\pi/\rho}{1-\pi}\frac{q-r}{1+r}\right)/\Sigma(y-r)^2 \tag{5.10}$$

The second term in (5.9) is likely to be negative for the following economic reasons. Most of the time the yield curve is upward sloping, so that $y - r > 0$. The liquidity premium $q - r < 0$, though, is negative. This induces a negative covariance in (5.9). Hence, even if agents are risk neutral, the coefficient estimate is downward biased vis a vis the expectations value of two.

Comparing the first column of Table 5.1 to the other columns, it appears that using the relative country spreads is helpful in reducing the downward bias. Adding the convexity term or using pooled estimation does not seem to change much. Apparently, the use of the relative country spreads increases the $\widehat{\beta}$'s, which by recalling (5.9) would suggest that the covariances $-\mathrm{cov}(X, P^*)$ and $-\mathrm{cov}(X^*, P)$ are positive and partially offset the negative intercorrelation of $\mathrm{cov}(X, P) + \mathrm{cov}(X^*, P^*)$ with the spread.

In trying to diminish the influence of the individual country disturbances on the regression coefficients further, we also pooled the data into a panel with a fixed time effect, and equality of slope coefficient β across countries as suggested in Mayfield and Murphy (1996). Panel estimation is a way to artificially lengthen the sample. The panel regression analog of the relative yield curve model (5.8) is reported in Table 5.2. The panel estimates with equal country coefficients are lower than the panel estimates with country individual slopes and the convexity term has the wrong sign, albeit it is not significant. Apparently not much of a large sample effect is obtained by pooling the country information on β. This may be due to the fact that all data come from the same relatively short time period and only involve OECD countries which pursued quite similar economic policies during the time. To conclude, the relative country approach appears to be the best we can do to lessen the omitted variable bias; at least if countries have similar economic positions it is likely that the cross correlations $-\mathrm{cov}(X, P^*)$ and $-\mathrm{cov}(X^*, P)$ are positive.

statistics	regression 1	regression 2
$\widehat{\theta}$	-0.03	-0.02
s.d.	0.01	0.01
$\widehat{\beta}$	1.16	1.21
s.d.	0.10	0.10
$\widehat{\tau}$	-	-0.50
s.d.	-	0.28
R^2	0.13	0.13
F-test	124.14	63.77
Prob(F-test)	0	0

Table 5.2. *Panel slope estimates. The table records coefficient estimates for the slope coefficient of the expectations model specification as in* (5.8). *The first column gives the outcome omitting the convexity term and the second column comprises this term. The base country in the regressions* (5.8) *is The Netherlands.*

5.3. **Cross section regressions and coefficient variability.** The empirical results from the previous section contained two major results. One is the well known downward bias vis a vis the expectations value of two. The other major result is the considerable variation in the individual cross country estimates, recall Table 5.1. It is a well known empirical fact that almost all variation is in the short term interest innovations $R_i - r$ rather than in the variation of the expected changes. In other words $\sigma_{R_i - r} \gg \sigma_{y-r}$, see e.g. Fama and Bliss (1987) and Cochrane (2001, Chapter 20). Thus the yield spread is not very informative about the future interest rates. This news dominance feature may contain the explanation for the considerable variation in the coefficient estimates across the countries. The interesting question is therefore as to whether the distribution of news is heavy tailed, in such a way that this could explain the wide dispersion of the coefficient estimates. In this subsection we try to elicit some information about this coefficient variation and relate this to the concept of regular variation dealt with the econometric theory part.

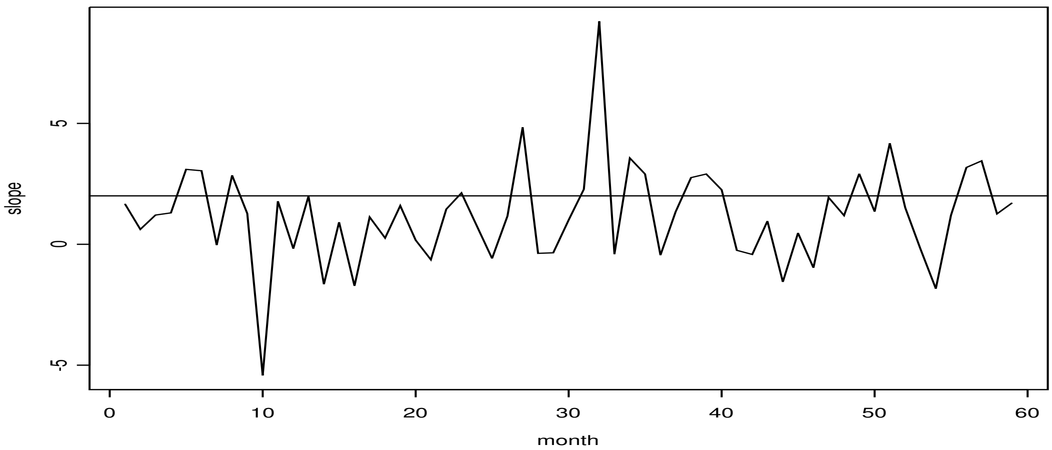

Figure 5.3. *Cross section slope estimates.*

From the panel estimates we concluded that restricting the slopes across countries over time did not yield extra information over and above the average of the individual country estimates. Since several countries pursued quite similar economic (monetary) policies during the second half of the nineties, the variation is perhaps not so much across countries, but rather over time. Therefore, to investigate the amount of variation in the slopes, we now assume identical slopes per country but allow for temporary disturbances. This model can be estimated by running per period cross section regressions. A graph of the 59 cross country $\widehat{\beta}_t$ OLS estimates for the specification (5.8) is provided in Figure 5.3. There appears to be quite a bit of variation in the slope estimates. Since the yield curve is widely known to be rotating and shifting, this was to be expected. Summary statistics are given in Table 5.4. The standard error of the slope estimates, the range and the kurtosis confirm the sizeable variation. But this is not due to a bad fit, since the R-squared statistic is mostly acceptable for the small cross section regressions. The variation appears to be genuine. The average of the slope estimates re-confirms the downward bias, and is somewhat more severe in these small samples.

The summary statistics indicate that the slope estimates are highly variable and, moreover, the statistics point to a non-normal distribution feature of the slopes, given the high kurtosis and strongly negative skewness. In Section 3 we showed that the heavy tail feature of the innovations carries over to the distribution of the coefficient estimates. Therefore we continue and investigate

mean	1.18
standard error (of mean)	0.26
median	1.21
standard deviation	2.03
sample variance	4.13
kurtosis	4.52
skewness	0.50
range	14.65
minimum	-5.43
maximum	9.22
mean R^2	0.31
standard error (of mean R^2)	0.03
standard deviation R^2	0.22

Table 5.4. *Summary statistics of cross country slope estimates.*

Statistic	Slopes	Slopes Squared
intercept	1.16	5.198
s.d.	0.31	1.411
t-value	3.70	3.682
AR(1)	0.01	-0.073
s.d.	0.13	0.133
t-value	0.07	-0.554
R^2	0.01	0.005

Table 5.5. $AR(1)$ *estimation results for the level and the squares.*

the evidence for regular variation in $\widehat{\beta}$. To this end we construct a number of QQ-plots, and we estimate the index of regular variation.

We start with a QQ-plot of the slope estimates against the Gumbel distribution. We choose the Gumbel distribution since the limit theory for the sample maxima of iid random variables X_i (cf. Embrechts et al. (1997), Chapter 3) says that, if the scaled and centered sample maxima have a limit distribution, the distribution of the X_i's is in the domain of attraction of either one of three limit laws. If there is an upper endpoint, it must be the Weibull law or the Gumbel. Without an endpoint, the limit law is either Fréchet, if the underlying distribution has heavy tails in the sense of regular variation at infinity, or it is Gumbel in the case of light tails (exponential) and medium type heavy tails (subexponential), as in the case of the log-normal distribution. Thus the limiting distribution of the extremes is tightly connected with the shape of the tail of the distribution of the X_i's. The Gumbel distribution moreover, has no parameters, while the other two limit laws require a parametrization of the tail. This can be exploited as follows. Suppose we use a QQ-plot of the data against the Gumbel quantiles, where $G(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$, and hence the QQ-plot is a plot of the double log transformed empirical probabilities against the rank ordered slopes. If the distribution is light tailed, i.e., the data are generated by a distribution in the domain of attraction of a Gumbel, then the QQ-plot should be approximately linear in the extreme North-East and South-West corners of the plot. On the contrary, if the generating distribution is in the domain of attraction of a Weibull, then the plot curves upward in the North-East corner of the plane, while in the case of a Fréchet law, when $F(x) = \exp(-x^{-\alpha})$, $x > 0$, it curves downward.

In the left graph of Figure 5.6 we plot the double log transformed empirical Gumbel probabilities against the rank ordered absolute values of the slope estimates. We take absolute values to increase the number of observations in the tail area and to save space.[1] One sees a clear deviation from

[1]Separate plots per tail are available upon request; these tell basically the same story.

the diagonal, and the curvature in the North-East corner is indicative of the Fréchet case and, therefore, hints at regular variation.

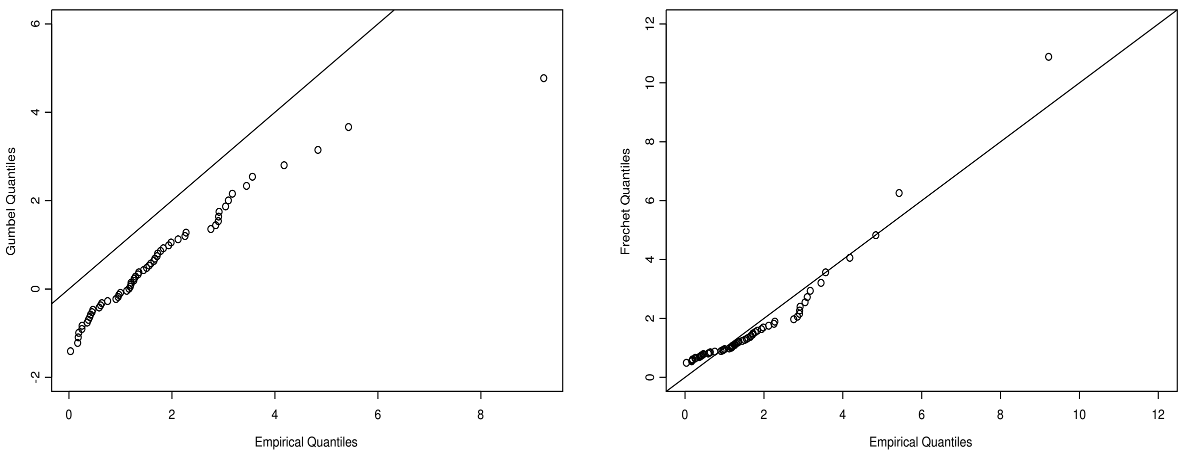

Figure 5.6. Left: *Gumbel QQ-plot for the absolute slope values.* Right: *QQ-plot of the absolute slope estimates against the Fréchet*(2) *law.*

The standard approach to estimating the index of regular variation is by means of computing the Hill statistic, which coincides with the maximum likelihood estimator of the tail index in case the data are exactly Pareto distributed. If the Pareto approximation is only good in the tail area, one conditions the estimator on a high threshold s, say, to obtain

$$\widehat{\frac{1}{\alpha}} = \frac{1}{M} \sum_{i=1, X_i < -s} \log \frac{-X_i}{s}\,, \tag{5.11}$$

where M is the random number of extreme observations X_i that fall below the threshold $-s$. To investigate the nature of the randomness in β, we give the Hill plot of the absolute values of the β estimates in the left graph of Figure 5.7. We considered all slope estimates below two as belonging to the left tail, since two is the theoretical benchmark from the Expectations Hypothesis. Hence, before taking absolute values we deducted two from the original estimates. The left plot in Figure 5.7 thus combines the slope estimates from both tails (all slopes minus two). Judging from the graph of tail index estimates, one would say that the coefficient of regular variation α hovers between 1 and 2, suggesting very heavy tails indeed. The way to read these plots is further explained in Embrechts et al. (1997, p. 341). Since the Hill estimator is biased, stemming from the fact that the distribution is not exactly Pareto, there is a region where if one uses too many observations the bias dominates, while the variance exerts a dominating influence if one uses too few observations. There exists an intermediate region in which the two vices are optimally balanced.

The Hill estimator is constructed on the assumption that the data are generated by a distribution which is regularly varying. It is also of interest not to make this assumption from the outset, so that one can test whether the data support this maintained hypothesis. From the limit law for the distribution of the maximum we know the only possible limit laws are those of Fréchet, Gumbel and Weibull. One can parameterize the limit laws in such a way that this corresponds to a power law with respectively a positive coefficient (the case treated by the Hill estimator), a value of zero (as in case of the normal distribution for example) and a negative coefficient (as in case of a uniform distribution). The Dekkers-Einmahl-de Haan (DEdH) estimator, see e.g. Embrechts et al. (1997),

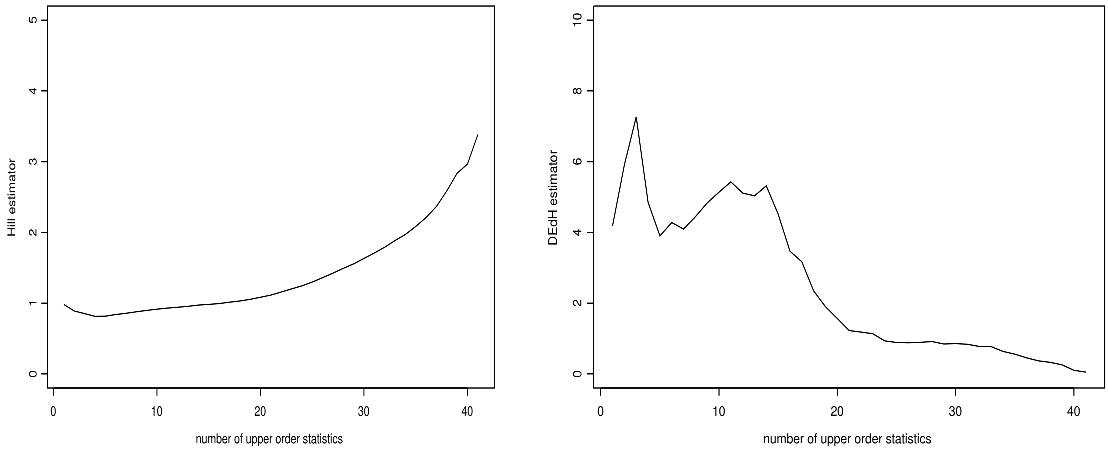

Figure 5.7. Left: *Hill plot. Combined tails* (-2). Right: *DEdH plots. Combined tails* (-2).

estimates this power without precommitting to a particular limit law. If one knows that the data are in the domain of attraction of the Fréchet case, then the Hill estimator is to be preferred since it is more efficient, but the DEdH should produce similar estimates as the Hill procedure in this case. The DEdH plot corresponding to the Hill plot of Figure 5.7 for the combined absolute values of the slope estimates is given in the right graph of Figure 5.7. Note that while this estimator allows for negative and zero values, the DEdH estimates are nevertheless always positive, therefore confirming the heavy tail features. In contrast to the Hill plots, the bias goes in the opposite direction, but this has to do with the fact that the DEdH estimator has different bias and variance properties. But unlike the Hill plot, one finds two plateaus, i.e., around 4-5 and 2, and it is less clear which value to choose. Nevertheless, one should bear in mind that the 59 slope estimates constitute a very limited amount of data for judging the tail behavior, so that some uncertainty about the tail index value is only to be expected.

Suppose we fix the tail index at two. Given this particular value for the tail index, we can come back to the QQ-plot technique. If the true distribution is indeed regularly varying with index two, then a QQ-plot against the Fréchet law with index two should give a straight line in the tail area. We use again the absolute values of the slopes, plotted against the quantiles of the Fréchet distribution $F(x) = \exp(-x^{-2})$. As can be seen in the right graph of Figure 5.6, the data now nicely align along the diagonal, corroborating the regular variation hypothesis and tail index value of two. Compare this with the left graph in Figure 5.6.

For the theory to have a bite, the simulation study showed that the heavy tailed noise must be dependent, since otherwise the effect of a large shock is rapidly diminished as the sample size increases. Therefore we investigate the dependency structure of the data, both over time and cross sectionally.

We first investigated the time series properties of the cross sectionally estimated slopes, to see whether there was evidence for (time) dependency in the multiplicative innovations. From the correlograms for the slope estimates and the squares thereof, there was no indication for the level or the squares to be time dependent. We further investigated the autocorrelation structure by estimating an AR(1) model for the level and the squares. Table 5.5 gives the AR(1) estimation results for the level and the squares. These estimates yield the same information, there is no indication that these slope estimates are time dependent.

Secondly, we looked at the time structure of the per country residuals from the cross sectional slope estimates, to see whether there was evidence for (time) dependency in the additive noise. On basis of correlograms, ARMA and ARCH estimates, again little or no evidence was found for a time structure in the residuals. To conclude, it appears that there is little or no evidence for dependency over time in the data.

Thirdly, we investigated the per period cross-sectional dependency. Since several countries pursued similar monetary policies (relative to the base country), cross-sectional dependency would be consistent with the economic facts. From the cross-sectional slope estimates we constructed the matrix (across months and countries) of residuals. From this matrix the covariance and correlation matrices were formed. The average of the absolute value of the 91 off-diagonal elements in the correlation matrix is 0.21. If we use the trace of squared correlation matrix as a measure of dependence, one finds that this number is comparable to a correlation matrix in which all off-diagonal elements are about $\rho = 0.25$. Thus this points to considerable cross country dependence. Since the data are non-normally distributed, we also computed the rank correlation matrix. Rank correlations reflect the copula and are therefore a more robust measure of the dependency if the innovations are non-normal. The average absolute value of the rank correlation matrix elements is 0.16. Since each element is constructed from 58 monthly observations, the asymptotically normal based critical value for a significant rank correlation coefficient is $1.96/\sqrt{58}$. A 31 elements of the 91 lower triangular elements exceed this critical value. If we take the covariance matrix and use the asymptotic χ^2 based test for equality with the identity matrix, i.e., using the 2-log criterion as in Anderson (1984, p. 434-7), we obtain a test value of 2432, which overwhelmingly rejects the null (critical value of 129). To conclude, while there is little or no dependency over time, there is clear evidence for cross-sectional dependency. This cross sectional dependence makes that large shocks, which occur due to the non-normal error structure, appear simultaneously in several countries. As was explained in the theory section and the simulation study, this causes the high variability in the small sample cross sectional slope estimates as depicted in Figure 5.3.

In sum, the slope estimates corroborate the presumption that the error terms $\rho_{n,\varepsilon}$ and $\rho_{n,\varphi}$ in the estimator $\widehat{\beta}$ are heavy tail distributed. This is due to a combination of heavy tailed distributed innovations and strong cross sectional dependence and a small cross section.

6. Conclusion

The paper provides a theory for tail probabilities for the linear regression estimator distribution in medium sized samples if the multiplicative and additive error terms follow a heavy tailed distribution. We show that even if standard moment conditions such as the existence of the variance are satisfied, the usual benchmarks based on central limit theory can be misleading. The results hinge on relatively weak assumptions regarding the stochastic nature of the explanatory variable. With additive uncertainty we require that the joint density of the explanatory variables is bounded in some neighborhood of the origin. A restriction is the condition that the regressor be exogenous. On the other hand, we allow for the possibility that the random multiplicative noise component be correlated with the additive noise term, and in this sense there can be correlation between the economic explanatory part and the additive noise structure. Moreover, both the noise and the regressor are allowed to be time dependent. It is shown that for a fixed sample size and if the perturbations are regularly varying, the OLS regression coefficient estimator has a tail probability which is the product of the tail probability of the perturbations and the expected 'kernel weight'. The tail influence of the perturbation term is lost in large samples by virtue of the central limit theorem. To derive these results we started with a novel result on scaling properties of products and ratios of regularly varying random variables.

A small Monte Carlo study showed the importance of the alternative assumptions of normally distributed versus heavy tail distributed innovations. Regardless of whether the noise in the regression is additive or multiplicative, there exists a clearly discernible effect of wider spread of the OLS estimator in medium sized samples, in contrast to the normal approximation and in contrast to normally distributed noise. This effect is particularly striking when the noise is dependent.

The application to yield curve estimation demonstrates the relevance of the theoretical results. Traditional slope estimates are clearly downward biased in contrast to the theoretical value suggested by the Expectations Hypothesis, but are not necessarily in conflict with economic theory (liquidity premium). Relative country estimation attenuates the bias. The estimates showed considerable variation across countries and over time. A thorough analysis of the set of cross country slope estimates revealed that the estimated random components come from a distribution with a regular varying tail. The heavy tail feature is significant in the cross sectional estimates due to the small cross section sample size and the cross sectional dependence.

We conclude that the theory is applicable to economic data and potentially explains the wide variability of observed regression estimates.

7. Appendix

We give the proofs to Lemmas 3.2, 3.4 and 3.7, and the expectations inequality used in Section 5.

7.1. **Proofs to Section 3.** We start with Lemma 3.2.

Proof. Assume first that the Z_i's are independent. If the Z_i's are non-negative the result is standard; see Feller (1971), p. 278, or Embrechts et al. (1997), Lemma 1.3.1. and Appendix A3.3. For general Z_i and Z_j, $i \neq j$, using the tail balance conditions and independence,

$$\lim_{x\to\infty} \frac{P(Z_i > x, Z_j > x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{p_i\, P(|Z_i| > x)\, p_j\, P(|Z_j| > x)}{\overline{F}(x)} = 0\,.$$

Similarly,

$$\lim_{x\to\infty} \frac{P(Z_i > x, Z_j \leq -x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{P(Z_i \leq -x, Z_j \leq -x)}{\overline{F}(x)} = 0\,.$$

Taking these calculations and assumption (3.2) into account, we see that the conditions of Lemma 3.1 are satisfied and so it follows that

$$\lim_{x\to\infty} \frac{P(Z_1 + \cdots + Z_n > x)}{\overline{F}(x)} = c_1^+ + \cdots + c_n^+\,,$$

implying that

$$P(Z_1 + \cdots + Z_n > x) = (1 + o(1))\,(P(Z_1 > x) + \cdots + P(Z_n > x))\,.$$

The case of the left tail $P(Z_1 + \cdots + Z_n \leq -x)$ is analogous.
For dependent Z_1, Z_2 with $\alpha_1 < \alpha_2$, $P(|Z_2| > x) = o(P(|Z_1| > x))$. Then similar calculations as above yield that the assumptions of Lemma 3.1 are satisfied. □

We continue with Lemma 3.4.

Proof. The fact that Z is regularly varying with index α follows by a straightforward application of Proposition A.1 in Basrak et al. (2002). The techniques used there also allow one to derive (3.4), which relation we show in detail. Let $M > 0$. Then

$$\begin{aligned} P(Z > x) &= P(Z > x, A_1) + P(Z > x, A_2) + P(Z > x, A_3) \\ &=: p_1 + p_2 + p_3, \end{aligned}$$

where

$$A_1 = \left\{|\mathbf{Y}| \leq M^{-1}\right\},\ A_2 = \left\{M^{-1} < |\mathbf{Y}| \leq M\right\},\ A_3 = \{|\mathbf{Y}| > M\}.$$

Then we have

$$p_1 \le P(|\mathbf{X}|\,|\mathbf{Y}| > x, A_1) \le P(|\mathbf{X}| > xM).$$

Clearly, $|\mathbf{X}|$ is regularly varying with index α, and so

$$\lim_{M\to\infty} \limsup_{x\to\infty} \frac{p_1}{P(|\mathbf{X}| > x)} = 0. \tag{7.1}$$

Recall from Breiman (1965) that for independent non-negative random variables ξ, η such that $E\eta^{\alpha+\delta} < \infty$ for some $\delta > 0$ and $P(\xi > x)$ regularly varying with index $\alpha > 0$.

$$P(\xi\eta > x) \sim E\eta^\alpha P(\xi > x),\ x \to \infty.$$

For p_3 we have, using Breiman's result, regular variation of $|\mathbf{X}|$ and Lebesgue dominated convergence,

$$\begin{aligned}\lim_{M\to\infty} \limsup_{x\to\infty} \frac{p_3}{P(|\mathbf{X}| > x)} &\le \lim_{M\to\infty} \limsup_{x\to\infty} \frac{P(|\mathbf{X}|\,|\mathbf{Y}|\, I_{(M,\infty)}(|\mathbf{Y}|) > x)}{P(|\mathbf{X}| > x)} \\ &= \lim_{M\to\infty} E[|\mathbf{Y}|^\alpha\, I_{(M,\infty)}(|\mathbf{Y}|)] = 0. \end{aligned} \tag{7.2}$$

By virtue of (7.1) and (7.2) the result must follow by a consideration of p_2. Indeed,

$$\begin{aligned}\lim_{x\to\infty} \frac{p_2}{P(|\mathbf{X}| > x)} &= \lim_{x\to\infty} \int_{M^{-1}<|Y|\le M} \frac{P(Z > x\,|\mathbf{Y})}{P(|\mathbf{X}| > x)} P(d\mathbf{Y}) \\ &= E\left[I_{(M^{-1},M)}(|\mathbf{Y}|) \sum_{i=1}^d (c_i^+ E[Y_i^\alpha I_{\{Y_i>0\}}] + c_i^- E\left[|Y_i|^\alpha\, I_{\{Y_i<0\}}\right]) \right]. \end{aligned} \tag{7.3}$$

In the last step of the proof we made use of Pratt's lemma (see Pratt (1960)) and Lemma 3.2. Now let $M \to \infty$ in (7.3) to conclude that the statement of the lemma is correct for $P(Z > x)$. The case $P(Z \le -x)$ is completely analogous. $\square$

The proof of Lemma 3.7 goes as follows.

Proof. Calculation shows that $EY_t^{-\alpha/2} < \infty$ if and only if for some $x_0 > 0$,

$$\int_0^{x_0} P(Y_t \le x^{2/\alpha}) x^{-2} dx < \infty. \tag{7.4}$$

If the X_t's are iid we have

$$\begin{aligned} P(Y_t \le x^{2/\alpha}) &\le P\left(\max_{i=1,\dots,n} X_i^2 \le x^{2/\alpha}\right) \\ &= P^n(|X| \le x^{1/\alpha}) \le \text{ const } x^{n\gamma/\alpha}. \end{aligned} \tag{7.5}$$

The function $x^{n\gamma/\alpha-2}$ is integrable on $[0, x_0]$ if $n\gamma/\alpha > 1$, hence (7.4) holds. Now assume that $(X_1, ..., X_n)$ has a bounded density f_n in some neighborhood of the origin. We conclude from (7.5) that

$$P(Y_t \le x^{2/\alpha}) \le \int_{\max_{i=1,\dots,n} |y_i| \le x^{1/\alpha}} f_n(\mathbf{y})\, d\mathbf{y} \le \text{const}\, x^{n/\alpha},$$

for sufficiently small x, and so we may conclude that (7.4) holds for $n > \alpha$. This concludes the proof. $\square$

Next we give the proof of Proposition 3.13.

Proof. (1) We observe that

$$\sum_{t=1}^{n} \varepsilon_t X_t^2 = \sum_{j=-\infty}^{t} Z_j \sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2 .$$

Write

$$\widetilde{\psi}_j = \frac{\sum_{t=\max(1,j)}^{n} \psi_{t-j} X_t^2}{\sum_{s=1}^{n} X_s^2} .$$

For fixed $k \le n$ we may then conclude from Lemma 3.4 that

$$P\left(\sum_{j=k}^{n} Z_j \widetilde{\psi}_j > x\right) \quad \sim \quad P(Z > x) \sum_{j=k}^{n} E[\widetilde{\psi}_j^{+}]^{\alpha} + P(Z \le -x) \sum_{j=k}^{n} E[\widetilde{\psi}_j^{-}]^{\alpha} .$$

Without loss of generality assume that $k \le 1$. Observe that

$$|\widetilde{\psi}_j| \le \sum_{t=1}^{n} |\psi_{t-j}| ,$$

Then similar calculations as in Mikosch and Samorodnitsky (2000) show that

$$\lim_{k \to -\infty} \limsup_{x \to \infty} \frac{P\left(\left|\sum_{j=-\infty}^{k-1} Z_j \widetilde{\psi}_j\right| > x\right)}{P(|Z| > x)} = 0 .$$

This proves the asymptotics for the right tail of $\rho_{n,\varepsilon}$. The asymptotics for the left tail of $\rho_{n,\varepsilon}$ are analogous.

(2) The proof of the second part is analogous, making use of the representation

$$\sum_{t=1}^{n} \varphi_t X_t = \sum_{j=-\infty}^{n} \gamma_j \sum_{t=\max(1,j)}^{n} c_{t-j} X_t ,$$

the moment condition (3.8) and, again, Lemma 3.4. □

References

[1] Anderson, T.W. (1984) *Introduction to Multivariate Statistical Analysis.* Wiley, New York

[2] Basrak, B., Mikosch, T. and Davis, R.A. (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* **99**, 95–116.

[3] Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[4] Bilson, J.F.O. (1981) The speculative efficiency hypothesis. *Journal of Business* **54**, 435–451.

[5] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge (UK).

[6] Brainard, W. (1967) Uncertainty and the effectiveness of policy. *American Economic Review, Papers and Proceedings* **57**, 411–425.

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

[8] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

[9] Campbell, J.Y., Lo, A.W. and McKinley, A.C. (1997) *The Econometrics of Financial Markets.* Princeton University Press, Princeton.

[10] Campbell, J.Y. and Shiller, R.J. (1991) Yield spreads and interest rate movements: a bird's eye view. *Review of Economic Studies* **58**, 495–514.

[11] Chow, G.C. (1984) Random and changing coefficient models. In: Griliches, Z. and Intriligator, M.D. (Eds.) *Handbook of Econometrics* Vol. **II**, North-Holland, Amsterdam.

[12] Cochrane, J.H. (2001) *Asset Pricing.* Princeton University Press, Princeton.

[13] Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability* **23**, 879–917.

[14] Davis, R.A. and Resnick, S.I. (1996) Limit theory for bilinear processes with heavy tailed noise. *Annals of Applied Probability* **6**, 1191–1210.
[15] Diamond, D.W. and Dybvig, P.H. (1983) Bank runs, deposit insurance, and liquidity. *Journal of Political Economy* **91**, 401–419.
[16] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.
[17] Fama, E. (1976) Forward rates as predictors of future spot rates . *Journal of Financial Economics* **3**, 361–377.
[18] Fama, E.F. and Bliss, R.B. (1987) The information in long maturity forward rates. *The American Economic Review* **77,** 680–692.
[19] Feige, E.L. and Swamy, P.A.V.B. (1974) A random coefficient model of the demand for liquid assets. *Journal of Money Credit and Banking* **6**, 241–252.
[20] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications II.* Wiley, New York.
[21] Fisher, I. (1930) *The Theory of Interest.* MacMillan, New York.
[22] Frenkel, J.A. (1993) *On Exchange Rates.* The MIT Press, Cambridge (US).
[23] Haan, L. de and Resnick, S.I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**, 317–337.
[24] Hodrick, R.J. (1987) *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets,* Harwood Publishers, Chur.
[25] Jacod, J. and Shiryaev, A.N. (1987) *Limit Theorems for Stochastic Processes.* Springer, Berlin.
[26] Lettau, M. and Ludvigson, S. (2001) Resurrecting the (C)CAPM: a cross sectional test when risk premia are time-varying. *Journal of Political Economy* **109**, 1238–1288.
[27] Lewis, K. (1995) Puzzles in international financial markets. In: Grossman, G.M. and Rogoff, K. (Eds.) *Handbook of International Economics Vol. III.* North-Holland, Amsterdam, 1913–1971.
[28] Maddala, G.S. (1977) *Econometrics.* McGraw-Hill, New York.
[29] Mankiw, N.G. and Miron, J.A. (1986) The changing behavior of the term structure of interest rates. *The Quarterly Journal of Economics* **51**, 211–228.
[30] Mayfield, E.S. and Murphy, R.G. (1996) Explaining the term structure of interest rates: a panel data approach. *Journal of Economics and Business* **48**, 11–22.
[31] Mikosch, T. (2003) Modeling Dependence and Tails of Financial Time Series. In: Finkenstädt, B. and Rootzén, H. (Eds.) *Extreme Value Theory and Applications.* Chapman and Hall, Chichester, pp. 185–286.
[32] Mikosch, T. and Samorodnitsky, G. (2000) The supremum of a negative drift random walk with dependent heavy-tailed steps. *Annals of Appied Probabability* **10**, 1025–1064.
[33] Pratt, J. (1960) On interchanging limits and integrals. *Annals of Mathematical Statistics* **31**, 74–77.
[34] Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Advances in Applied Probability* **18**, 66–138.
[35] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.
[36] Sack, B. (2000) Does the fed act gradually? A VAR analysis. *Journal of Monetary Economics* **46**, 229–256.
[37] Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, and MaPhySto, and EURANDOM, Eindhoven
E-mail address: mikosch@math.ku.dk , www.math.ku.dk/∼mikosch

Department of Economics, H8-33, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands, and EURANDOM, Eindhoven
E-mail address: cdevries@few.eur.nl