Statistical Papers **44**, 47–65 (2003)

Statistical Papers

Maximum likelihood estimators in regression models with infinite variance innovations

Vygantas Paulaauskas[1] and Svetlozar T. Rachev[2]

[1] Department of Mathematics, Vilnius University, Vilnius, Lithuania
[2] Institute of Statistics and Mathematical Economics, University of Karlsruhe, Karlsruhe, Germany, and Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, USA

Received: September 6, 1999; revised version: November 22, 2000

In this paper we consider the problem of maximum likelihood (ML) estimation in the classical AR(1) model with i.i.d. symmetric stable innovations with known characteristic exponent and unknown scale parameter. We present an approach that allows us to investigate the properties of ML estimators without making use of numerical procedures. Finally, we introduce a generalization to the multivariate case.

AMS 1980 classifications: 60E07, 60F17, 62F12, 62J05.
Key words and phrases: Autoregression, stable distributions, Lévy processes, maximum likelihood estimators.

1 Introduction and Formulation of Results

We study the classical autoregressive model:

$$X_i = bX_{i-1} + \varepsilon_i, \qquad i = 1, 2, ..., n\,, \tag{1}$$

where $X_1, ..., X_n$ are observed variables, b is the unknown parameter to be estimated, and $(\varepsilon_i)_{i\geq 1}$ is the innovation process. We assume the initial value X_0 to be known. The asymptotic properties (such as consistency, rate of convergence and limiting distribution) of the ordinary least squares (OLS) estimator of b, given by

$$\tilde{b}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2}\,, \tag{2}$$

will depend on the probabilistic structure of the innovation process. From now on, we assume that the innovations ε_i are i.i.d. random variables (independence is an essential assumption, while the hypothesis that all ε_i are identically distributed, can be relaxed). Next, it is usually assumed that ε_1 has a completely specified distribution, for example $N(0,\sigma^2)$, (that is, normal with mean zero and variance σ^2), or a stable law with known parameters, or just a distribution belonging to the domain of attraction of a given stable law. From this, there exists an appropriate normalization factor a_n and a limiting law for $a_n(\tilde{b}_n - b)$, where b is the true value of the parameter in (1). Generally speaking, this limiting law depends on the asymptotic behavior of the appropriately normalized sums $\sum_{i=1}^n \varepsilon_i$, and is different in the cases $|b| < 1$, $|b| > 1$, and $|b| = 1$. There exists vast literature devoted to OLS estimators in this classical model: we mention here only the seminal papers by Mann and Wald (1943), White (1958) and Anderson (1959). For more recent results encompassing more general settings, we refer to Phillips (1987), Rachev, Kim and Mittnik (1997) and Mijnheer (1997).

The assumption that the distribution of the innovations $(\varepsilon_i)_{i\geq 1}$ is known is rather difficult to justify. It is by far a more realistic situation when only information on the form of the distribution of ε_1 is available. Suppose that ε_1 is distributed as $N(0,\sigma^2)$, where σ^2 is unknown. Since

$$\varepsilon_i = X_i - bX_{i-1}, \qquad i = 1,2,...,n,$$

we can write the likelihood function (LF) explicitly and obtain the maximum likelihood (ML) estimators of both unknown parameters b and σ as follows:

$$\tilde{b}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2}, \qquad \tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \tilde{b}_n X_{i-1}\right)^2. \tag{3}$$

It is then not difficult to find limiting distributions for properly normalized errors of the estimators, i.e. for

$$\sqrt{n}(\tilde{b}_n - b) \qquad \text{and} \qquad \sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2).$$

For the first quantity, we need to consider separately the cases $|b| < 1$ ("stationary" case), and $|b| = 1$ ("unit root" case).

If the assumptions of our model are compatible with the true underlying generating process, then the sample residuals

$$\widehat{\varepsilon}_i = X_i - \tilde{b}_n X_{i-1} \qquad i = 1,2,...,n \tag{4}$$

should comply with the hypothesis that the ε_i's are normally distributed. However, if they exhibit "heavy tails" or skewness, (such features are frequently observed in financial data, for instance), we can infer that our assumption on the probabilistic structure of the innovations is incorrect. The natural candidate for the distribution of ε_i, which allows heavy tails and skewness, is the family of stable laws. We recall that stable (non-gaussian) distributions are usually defined by means of their characteristic function

$$\varphi_\alpha(t;\sigma,\beta) = \begin{cases} \exp\Big\{ -\sigma^\alpha|t|^\alpha\Big[1 - i\beta(\operatorname{sign} t)\operatorname{tg}\frac{\pi\alpha}{2}\Big]\Big\} & \text{for } \alpha \neq 1 \\ \exp\Big\{ -\sigma|t|\big[1 + i\beta\frac{2}{\pi}(\operatorname{sign} t)\ln|t|\big]\Big\} & \text{for } \alpha = 1, \end{cases}$$

where $0 < \alpha < 2$, $\sigma > 0$, and $-1 \le \beta \le 1$. The parameter α is called the stable (or tail) exponent; σ is the scale parameter; and β is the skewness parameter, (if $\beta = 0$, then the corresponding stable distribution is called strictly stable). Note that in this representation it is assumed that the shift (or location) parameter is 0.

Denote by $G_\alpha(x;\sigma,\beta)$ and $g_\alpha(x;\sigma,\beta)$ the distribution and the density functions, respectively, of a stable distribution with characteristic function $\varphi_\alpha(t;\sigma,\beta)$. Now suppose that the distribution of ε_1 is $G_\alpha(x;\sigma,\beta)$ with known α, but unknown parameters σ and β. This assumption is analogous to the Gaussian case mentioned above (known $\alpha = 2$, but σ^2 is unknown). However, if we proceed as in the Gaussian case to obtain the likelihood function, we face a serious problem: stable densities, except for a few cases, do not admit an analytical expression, so we cannot write the likelihood function explicitly. When dealing with a similar problem of estimating the parameters of a stable law, one can try to perform a numerical maximization of the log likelihood function with respect to the unknown parameters (see, for example, Mittnik, Rachev and Paolella (1998), Nolan (1997), and references therein). However, this approach is not attractive. It does not lead to analytical expressions of the underlying estimators, while the asymptotic analysis of numerically obtained estimators is just intractable.

Even in the few cases for which an analytic expression of the stable density is available, the ML estimators cannot be easily analyzed.

For example, suppose that $(\varepsilon_i)_{i\ge1}$ follows the Cauchy distribution with density

$$g_1(x,\sigma,0) = \frac{\sigma^2}{\pi(\sigma^2 + x^2)},$$

where σ is an unknown parameter. The ML estimators $\widehat{b}_n$ and $\widehat{\sigma}_n$ can be obtained as solutions of the system:

$$\begin{cases} \displaystyle\sum_{i=1}^n \frac{(X_i - bX_{i-1})X_{i-1}}{\sigma^2 + (X_i - bX_{i-1})^2} = 0 \\ \displaystyle\sigma^2\sum_{i=1}^n \frac{1}{\sigma^2 + (X_i - bX_{i-1})^2} = n. \end{cases}$$

The equations for the unknown σ and b are highly nonlinear hence, it is hard to find simple explicit expressions for the estimators.

A similar situation arises in the case of the Lévy law with exponent $\alpha = \frac{1}{2}$ and density

$$g_{1/2}(x,\sigma) = \Big(\frac{\sigma}{2\pi}\Big)^{1/2} x^{-3/2}\exp\Big\{-\frac{\sigma}{2x}\Big\}, \qquad x > 0.$$

Again, the equations obtained by maximizing the likelihood function do not lead to explicit expressions for the estimators of the unknown parameters b and σ.

So far these difficulties do not allow us to study the general case of stable innovations. However, we can propose a solution in the particular case of symmetric α–stable innovations with unknown scale parameter σ. To this end, we assume that $(\varepsilon_i)_{i\geq 1}$ are i.i.d. symmetric α–stable ($S\alpha S$) random variables with characteristic function

$$\varphi_\alpha(t) = \exp\big(- \sigma^\alpha |t|^\alpha \big), \tag{5}$$

where α is known and σ is unknown. The random variable ε_1 with characteristic function (5) is called subgaussian (see for example, Feller (1971) or Samorodnitsky and Taqqu (1994)), and it admits a representation as a product of two independent random variables: $\varepsilon_1 = U_1 V_1$, where $U_1 \sim N(0, 2\sigma^2)$, $V_1 = A_1^{1/2}$ and A_1 is an $\alpha/2$–stable subordinator, that is, A_1 is a positive random variable with characteristic function $\varphi_{\frac{\alpha}{2}}\Big(x, \big(\cos \frac{\pi\alpha}{4}\big)^{2/\alpha}, 1\Big)$.

Taking independent sequences of i.i.d. random variables $(V_i)_{i\geq 1}$ and $(U_i)_{i\geq 1}$ with U_1 and V_1 as defined above, we can write

$$\varepsilon_i = V_i U_i. \tag{6}$$

From (1) we have

$$U_i = \frac{X_i - bX_{i-1}}{V_i}.$$

Now, since the U_i's are normally distributed, we can write the likelihood function and obtain the following ML estimators:

$$\widehat{b}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1} V_i^{-2}}{\sum_{i=1}^{n} X_{i-1}^2 V_i^{-2}}, \tag{7}$$

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \big(X_i - \widehat{b}_n X_{i-1}\big)^2 V_i^{-2}. \tag{8}$$

The estimators (7) and (8) are similar to those in (3), with the only difference that the variables X_i and X_{i-1} are now replaced by $X_i V_i^{-1}$ and $X_{i-1} V_i^{-1}$, respectively. There is however, the following problem: while in (6) V_i and U_i are independent, in the relation $U_i = \varepsilon_i V_i^{-1}$, we cannot assume ε_i and V_i to be independent. If ε_i and V_i were independent, then U_i would be a heavy-tailed random variable in the domain of normal attraction of α–stable law, denoted shortly as DNA(α). The problem of generating values of V_i's seems rather difficult and we shall address this issue later on. The selection of α is a separate problem. One possible way to estimate α is to first consider the estimator (2), then to evaluate the sample residuals

(4). We consider these residuals as a sample from stable distribution and then we can estimate the exponent α of a stable distribution. There is vast literature on the estimation of parameters of stable distributions, both in univariate and in multivariate cases, see for example survey paper McCulloch (1996), or a recent paper Davydov, Paulauskas and Račkauskas (2000), where it is proposed that an asymptotically unbiased and consistent estimator of the exponent of a multivariate stable law, is asymptotically normal with standard $\sqrt{n}$ rate.

Next, we study the asymptotic properties of the (appropriately normalized) error terms $\widehat{b}_n - b$ and $\widehat{\sigma}_n^2 - \sigma^2$, where b and σ are the true values of the parameters under consideration. We separately treat the cases $|b| < 1$ and $b = 1$. Our main result is the following theorem whose proof is given in section 2.
In what follows, "$\Rightarrow$" stands for convergence in distribution.

Theorem 1 *Suppose that in model (1) the innovations $(\varepsilon_i)_{i\geq 1}$ are i.i.d. symmetric α–stable random variables with unknown scale parameter σ. Then, as $n \to \infty$,*

$$\left(\frac{1}{2}n\right)^{1/2}\left(\frac{\widehat{\sigma}_n^2}{\sigma^2} - 1\right) \Rightarrow N(0,1). \tag{9}$$

If $|b| < 1$ (the "stationary case"), then

$$n^{1/\alpha}\left(\widehat{b}_n - b\right) \Rightarrow \frac{S_1}{S_2}. \tag{10}$$

If $b = 1$ (the "unit root case"), then

$$n^{1/2+1/\alpha}\left(a(\alpha)\right)^{1/2}\left(\widehat{b}_n - 1\right) \Rightarrow \frac{\int_0^1 Y_\alpha^-(t)\,dW(t)}{\int_0^1 \left(Y_\alpha(t)\right)^2 dt}. \tag{11}$$

In (10), S_1 and S_2 are stable random variables defined in (30) and (31), respectively. In (11), $a(\alpha) := EV_1^{-2}$, $Y_\alpha(t)$ is a standard α–stable Lévy motion, (see Lemma 3 below), $Y_\alpha^-(t) = \lim_{s\uparrow t} Y_\alpha(s)$, and $W(t)$ is a standard Brownian motion, independent of Y_α.

The main advantage of the ML estimator $\widehat{b}_n$ in the case of unit root is its better rate of convergence compared with that of the OLS estimator. Namely, Chan and Tran (1989) showed that the rate of convergence for the OLS estimator is independent of α and is of order n, while in (11), we achieve a rate of convergence of order $n^{1/2+1/\alpha}$, which is better for all $\alpha < 2$. We recall that in (10), the limiting distribution depends on the unknown parameters b and σ, and this is a drawback of our theorem. However, the limiting distributions in (9) and (11) are independent of b and σ, and therefore they can be used to construct confidence intervals. Moreover, the limiting distribution in (11) can be obtained by a simulation method (see Mittnik, Paulauskas and Rachev (1999)).

Another drawback of Theorem 1 is the difficulty of generating values of V_i. Indeed, we can write the joint two–dimensional density function of the pair (ε, V), and then obtain the following expression of the conditional density of V, given that $\varepsilon = x_0$

$$f_V(y|\varepsilon = x_0) = \frac{y^{-1} p_V(y) \exp\left\{ - \frac{x_0^2}{4\sigma^2 y^2} \right\}}{\int_0^\infty z^{-1} p_V(z) \exp\left\{ - \frac{x_0^2}{4\sigma^2 z^2} \right\} dz}, \tag{12}$$

where $p_V(x) = 2x g_{\frac{\alpha}{2}}\left(x^2; \left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 1\right)$. Since the density $g_{\frac{\alpha}{2}}(x; \sigma, 1)$ has no explicit expression (except for the case $\alpha = 1$), it seems that the only possible way to generate values of V_i is to evaluate the density p_V numerically. Furthermore, the question of choosing x_0 is also nontrivial. We suggest the following procedure: evaluate OLS estimate $\tilde{b}_n$ given by (2), then compute the sample residuals $\widehat{\varepsilon}_i = X_i - \tilde{b}_n X_{i-1}$ for all $i = 1, 2, ..., n$. To generate the value V_i, take $x_0 = \widehat{\varepsilon}_i$ in (12). The unknown scale parameter σ is present in (12), therefore, from the obtained residuals $\widehat{\varepsilon}_i$, we need to estimate σ. At this point, we can propose the following estimator for the scale parameter σ of a stable random variable. The estimator is based on the following formula (see Samorodnitsky and Taqqu (1994)): if ε_1 is an $S\alpha S$ random variable with characteristic function (5), then for any $0 < p < \alpha$, it holds

$$E|\varepsilon_1|^p = C(\alpha, p)\sigma^p,$$

where $C(\alpha, p)$ stands for a constant depending on α and p. Therefore, we can take

$$\widehat{\sigma}_n = \left((C(\alpha, p))^{-1} \frac{1}{n} \sum_{i=1}^{n} |X_i - \tilde{b}_n X_{i-1}|^p \right)^{1/p}. \tag{13}$$

For all values of p in the interval $0 < p < \alpha$, the consistency of this estimator will follow from the law of large numbers and the consistency of the estimator $\tilde{b}_n$. Instead of $\tilde{b}_n$, we can take any consistent estimator of b, but the rate of convergence of $\widehat{\sigma}_n$ to σ may depend on the choice of the estimator for b. In fact, the choice of p is a separate problem of considerable interest and we intend to investigate the properties of the estimator (13) elsewhere. Since the main goal of the paper is the discussion of theoretical issues of ML estimation in the case of stable innovations, we intend to discuss simulation results and related problems in a separate paper, as it was done in the case of OLS estimators (see Paulauskas and Rachev (1998) and Mittnik, Paulauskas and Rachev (1999)).

Here we give only a table with preliminary simulation results for several values of α, σ, and b. The last column of the table gives the value of a parameter A, which has the following meaning. In order to generate values V_i, we need to evaluate the conditional density (12). We approximate the integral in the denominator with an integral over the interval $[0, A]$ with step $h = 0.01$. In most cases we took $A = 500$ and $n = 10000$, but it seems that for the smallest values of α, even $A = 500$ is too small. Therefore, for $\alpha = 1.1$, we tried $A = 1000$, then to make the computations in a reasonable

Table 1 Preliminary simulation results

α	σ	b	$\widehat{b}_n$ (MLE)	$\tilde{b}_n$ (OLS)	n	A
0.5	0.1	0.5	0.499962	0.4999579	10000	500
0.5	0.1	0.8	0.7999551	0.7999551	10000	500
0.5	0.1	0.9	0.8988169	0.8988137	10000	500
0.5	0.1	0.95	0.9422095	0.9420699	10000	500
0.5	0.1	0.99	0.9901045	0.9901055	10000	500
0.5	0.1	1	1.0000384	1.0000562	10000	500
1.1	0.1	0.5	0.5114295	0.5112379	1000	1000
1.1	0.1	0.8	0.7952818	0.7954026	1000	1000
1.1	0.1	0.9	0.9013114	0.9052705	1000	1000
1.1	0.1	1	1.0000114	0.9996152	1000	1000
1.4	0.1	0.5	0.500875	0.502669	10000	500
1.4	0.1	0.8	0.7947087	0.7910217	10000	500
1.4	0.1	0.9	0.90010468	0.90183797	10000	500
1.4	0.1	1	0.99931134	0.99855534	10000	500

time frame, we were forced to lower n. (Evaluating the residuals $\widehat{\varepsilon}_i$ for $\alpha = 0.5$, we observed the values of the order 10^7, and this shows that even $A = 1000$ is too small).

For $\alpha = 1.1$ and $\alpha = 1.4$ and $b = 1$, we can see the effect of the better rate of convergence of ML versus OLS estimators, and the simulation results fit well to theoretical comparison given after formulation of Theorem 1. For example, the ratio of errors for OSL and ML estimators in the case of $\alpha = 1.1$ is 33.75, and it is of the same order as predicted by theory $n^{\frac{2-\alpha}{2\alpha}} = 10^{1.227}$. Only the case $\alpha = 0.5$, $b = 1$ does not fit well into the picture and at present, the only explanation for this is that the value of A is too small for this case.

We now consider the multivariate generalization of the model (1):

$$X_i = BX_{i-1} + \varepsilon_i, \tag{14}$$

where $X_i = \left(X_{i1}, ..., X_{ik}\right)$, $B = \left\{b_{i,j}\right\}_{i,j=1,...,k}$ is an unknown matrix, and $\varepsilon_i = \left(\varepsilon_{i1}, ..., \varepsilon_{ik}\right)$. We shall assume that $(\varepsilon_i)_{i\geq 1}$ are i.i.d. random vectors. The case $\varepsilon_1 \sim N(0, \Sigma)$ with known or unknown covariance matrix Σ is well investigated: see, for example, Park and Phillips (1988), Johansen (1988), (1996) and references therein. (Here, as usual, $N(0, \Sigma)$ stands for a normal distribution with mean zero and covariance matrix Σ.) Furthermore, the OLS method can even be extended to the case of innovations ε_i belonging to the domain of normal attraction of the known operator–stable law, (see Paulauskas and Rachev (1998)). However, difficulties arise when we assume ε_i to have a multivariate stable law with unknown spectral measure and at present, we are not able to analyze this general case.

Next, we recall some facts about multivariate stable distributions. A random vector $X = (X_1, ..., X_k)$ is called multivariate stable with exponent $0 < \alpha < 2$ if its characteristic function has the following form

$$Ee^{i(t,X)} = \begin{cases} \exp\{ - \int_{\mathbb{S}^k} |(t,s)|^\alpha (1 - i\, \mathrm{sign}((t,s)) \mathrm{tg} \frac{\pi\alpha}{2}) \Gamma(ds) \}, & \alpha \neq 1 \\ \exp\{ - \int_{\mathbb{S}^k} |(t,s)| (1 + i\, \frac{\pi}{2} \mathrm{sign}((t,s)) \ln |(t,s)|) \Gamma(ds) \}, & \alpha = 1, \end{cases}$$

where $\mathbb{S}^k = \{x \in \mathbb{R}^k : ||x|| = 1\}$ is the unit sphere in $\mathbb{R}^k$ and Γ is a finite measure on $\mathbb{S}^k$. Γ is called the spectral measure of the stable random vector X, and the pair (α, Γ) completely characterizes the stable law. (Again, as in the one–dimensional case, we have assumed that the shift vector is zero). For a detailed survey on multivariate stable vectors, we refer to Samorodnitsky and Taqqu (1994), and to Jurek and Mason (1993) for facts on operator–stable vectors.

The straightforward generalization of the approach used in Theorem 1 is obtained assuming that $\varepsilon_1 = U_1 V_1$, where U_1 is a k–dimensional normal law with mean zero and covariance matrix Σ (which we assume to be unknown), and $V_1 = A_1^{1/2}$, with A_1 one–dimensional $\alpha/2$–stable subordinator, α being a known parameter, $0 < \alpha < 2$. In other words, ε_1 is an $S\alpha S$ random vector with a subgaussian characteristic function

$$Ee^{i(t,\varepsilon_1)} = \exp\Big\{ - \Big(\frac{1}{2}\Sigma t, t\Big)^{\alpha/2}\Big\}. \tag{15}$$

Then, arguing in a similar way as in the one–dimensional case of Theorem 1, we obtain the following ML estimators:

$$\widehat{B}_n = C_n^{-1} \sum_{i=1}^{n} \frac{1}{V_i^2} (X_{i-1} X_i') \tag{16}$$

and

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{V_i^2} (X_i - \widehat{B}_n X_{i-1})(X_i - \widehat{B}_n X_{i-1})', \tag{17}$$

where $X_i = (X_{i1}, ..., X_{ik})$ is a $k \times 1$ column vector, X_i' is a $1 \times k$ row vector, and $C_n = \sum_{i=1}^{n} \frac{1}{V_i^2} (X_{i-1} \cdot X_{i-1}')$.

We formulate the generalization of Theorem 1 in the stationary case only; the proof follows the same type of arguments as in Theorem 1, but is more cumbersome.

Theorem 2 *Suppose that in model (14), the innovations* $(\varepsilon_i)_{i \geq 1}$ *are i.i.d. symmetric* α*–stable random vectors with sub-gaussian characteristic function (15) and unknown matrix* Σ*. Suppose that the unknown matrix* B *satisfies* $||B|| < 1$*. Then, as* $n \to \infty$*,*

$$n^{1/\alpha}(\widehat{B}_n - B) \Rightarrow Z_2^{-1} Z_1, \tag{18}$$

where Z_1 is $k \times k$ matrix with α–stable entries depending on Σ and B. Furthermore, Z_2 is a diagonal $k \times k$ matrix with $\alpha/2$–stable positive entries also depending on Σ and B. As for the estimator $\widehat{\Sigma}_n$, the following relation holds:

$$\sqrt{n}\big(\widehat{\Sigma}_n - \Sigma\big) \Rightarrow Z_3. \tag{19}$$

Here, Z_3 is a $k \times k$ random normal matrix with mean–zero entries Z_3^{ij} and covariances

$$\operatorname{Cov}\big(Z_3^{ij}, Z_3^{ml}\big) = E\big(U_{1i}U_{1j} - \sigma_{ij}\big)\big(U_{1m}U_{1l} - \sigma_{ml}\big), \qquad i, j, m, l = 1, \dots, k.$$

Remark. It is easy to see that (19) is the multivariate analog of the "traditional" difference $\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2)$ in the one–dimensional case. Therefore, the limiting distribution depends on the unknown matrix Σ. Most likely, in order to obtain a limiting distribution that does not depend on the unknown parameters, one needs to consider the difference

$$\sqrt{n}\big(\Sigma^{-1}\widehat{\Sigma}_n - I\big), \tag{20}$$

where I is the $k \times k$ identity matrix. (However, (20) does not provide a method for easily constructing the confidence regions for Σ.)

In Paulauskas and Rachev (1998), we considered the more general model with innovations being multivariate with coordinates having different exponents. However, the assumption that the distribution of ε_1 is specified. It seems that such an extension is also possible here. To see the difficulties in this case, let us introduce the coordinate-wise multiplication and division of vectors as follows: for $x, y \in \mathbb{R}^k$, $x \odot y := (x_1y_1, ..., x_ky_k)$ and, if $y_i \neq 0$ for all $i = 1, ..., k$, then

$$\frac{x}{y} = \Big(\frac{x_1}{y_1}, \dots, \frac{x_k}{y_k}\Big).$$

We can assume then that

$$\varepsilon_1 = U_1 \odot V_1 := \big(U_{11}V_{11}, ..., U_{1k}V_{1k}\big). \tag{21}$$

In (21) we assume that U_1 is $N(0, \Sigma)$, and $V_1 = \big(V_{11}, ..., V_{1k}\big)$, $V_{1j} = A_j^{1/2}$, with A_j being $\alpha_j/2$–stable subordinator, $0 < \alpha_j < 2$, $j = 1, ..., k$. Next, let V_{1j}, $j = 1, ..., k$, be independent random variables and independent of the vector U. Now ε_1 has a more complex structure: it is no longer $S\alpha S$ (furthermore, it does not seem possible to get an explicit expression of the characteristic function of ε_1), but its coordinates ε_{1j} are $S\alpha_jS$ random variables with unknown scale parameter σ_j, where $\sigma_1^2 = EU_{1j}^2$. Assume that the multi-index $\bar{\alpha} = (\alpha_1, ..., \alpha_k)$ is known. (The problem of estimating $\bar{\alpha}$ should be treated separately as in the one–dimensional case).

In model (14), (21) we have

$$U_i = \frac{X_i - BX_{i-1}}{V_i}, \qquad i = 1, 2, ..., n,$$

and then we can write the likelihood function as a function of two unknown matrices B and Σ. However, the equations for the ML estimators $\widehat{B}_n$ and $\widehat{\Sigma}_n$ are too complicated and we were not able to obtain explicit expressions of the estimators. (Of course, one can try to solve these equations numerically, but we do not consider this approach here.)

One reason for these difficulties is the fact that the usual matrix and vector multiplication and coordinate-wise division do not commute, that is, if B is a $k \times k$ matrix and $x, y \in \mathbb{R}^k$, $(y_i \neq 0,\ i = 1, ..., k)$, then

$$\frac{Bx}{y} \neq B \cdot \frac{x}{y}. \tag{22}$$

Even if one assumes that B is diagonal, (then there is equality in (22)), and thus the model (14) becomes

$$X_{ij} = B_j X_{i-1,j} + U_{i,j} V_{i,j}, \qquad j = 1, ..., k, \quad i = 1, ..., n,$$

the equations for the variates B and Σ do not separate due to the dependence between U_{1j}, $j = 1, ..., k$. As a consequence, one cannot find explicit expressions for the ML estimators of the unknown parameters.

2 Proofs

Proof of Theorem 1. We consider the model defined by (1) and (6), and the ML estimators (7) and (8). Assume that $|b| < 1$. In this case it is known that under an appropriate choice of the value X_0, there exists a stationary solution of (1):

$$X_n = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{n-i}, \qquad n = 1, 2, ... \tag{23}$$

with exponentially decreasing coefficients Ψ_i. (Here we have assumed that the sequence of i.i.d. ε_n is defined both for positive and negative values of n). In fact, it is enough to set $X_0 = \sum_{i=0}^{\infty} b^i \varepsilon_{-i}$ to obtain (23) with $\Psi_i = b^i$.

From (9) and (10), using (1) we obtain

$$\widehat{b}_n - b = \Big(\sum_{i=1}^{n} X_{i-1}^2 A_i^{-1}\Big)^{-1} \sum_{i=1}^{n} \varepsilon_i X_{i-1} A_i^{-1}, \tag{24}$$

and

$$\widehat{\sigma}_n^2 - \sigma^2 = n^{-1} \sum_{i=1}^{n} \left(\varepsilon_i^2 A_i^{-1} - \sigma^2\right) - 2n^{-1}\left(\widehat{b}_n - b\right) \sum \varepsilon_i X_{i-1} A_i^{-1}$$
$$+ (\widehat{b}_n - b)^2 n^{-1} \sum_{i=1}^{n} X_{i-1}^2 A_i^{-1}, \tag{25}$$

where $A_i = V_i^2$. Denote $\tilde{\varepsilon}_i = \varepsilon_i A_i^{-1} = U_i V_i^{-1}$. From the independence of U_i and V_i,

$$E\tilde{\varepsilon}_1 = EU_1 EV_1^{-1} = 0,$$
$$E\tilde{\varepsilon}_1^2 = EU_1^2 EV_1^{-2} = \sigma^2 \cdot a,$$

where $a = a(\alpha) := EV_1^{-2}$ is finite as the stable subordinator has exponentially decreasing density at zero. The expectations EV_1^{-1} and EV_1^{-2} are finite due to the following lemma. Let p_1 and p_2 denote the densities of the random variables V_1^{-1} and V_1^{-2}, respectively.

Lemma 1 *The following asymptotic relations hold: for* $0 < \alpha < 2$,

$$\begin{aligned}
p_1(x) &\sim c(\alpha)x^{\alpha-1}, \qquad \text{as} \quad x \to 0,\\
p_1(x) &\sim c(\alpha)x^{\frac{2(\alpha-1)}{2-\alpha}} \exp\Big\{ -\Big(1-\frac{\alpha}{2}\Big)x^{\frac{2\alpha}{\alpha-2}}\Big\}, \quad \text{as} \quad x \to \infty,\\
p_2(x) &\sim c(\alpha)x^{\frac{\alpha}{2}-1}, \qquad \text{as} \quad x \to 0,\\
p_2(x) &\sim c(\alpha)x^{-\frac{4-3\alpha}{2(2-\alpha)}} \exp\Big\{ -\Big(1-\frac{\alpha}{2}\Big)x^{\frac{\alpha}{2-\alpha}}\Big\}, \quad \text{as} \quad x \to \infty,
\end{aligned}$$

where $a(x) \sim b(x)$, *as* $x \to a$, *means* $\lim_{x\to a} \dfrac{a(x)}{b(x)} = 1$, *and the generic constant* $c(\alpha)$ *can be different in the above relations.*

Proof We use the following well–know fact (see Zolotarev (1986)): for $0 < \alpha < 1$, it holds

$$\begin{aligned}
g_\alpha(x,1,1) &\sim c_1(\alpha)x^{2-\alpha/2(\alpha-1)} \exp\big\{ -(1-\alpha)x^{\alpha/\alpha-1}\big\}, \qquad \text{as} \quad x \to 0\\
g_\alpha(x,1,1) &\sim c_3(\alpha)x^{-(\alpha+1)}, \qquad \text{as} \quad x \to \infty.
\end{aligned}$$

The positively skewed random variable A_1 has density $g_{\alpha/2}(x, c(\alpha), 1)$, and therefore $p_2(x) = \frac{1}{x^2} g_{\alpha/2}\big(\frac{1}{x}, c(\alpha), 1\big)$, $p_1(x) = 2xp_2(x^2)$. These relations prove the lemma.

Next, we need to find a proper normalization and joint limiting laws for the sums

$$n^{-1}\sum_{i=1}^{n}\big(\varepsilon_i^2 A_i^{-1} - \sigma^2\big), \quad \sum_{i=1}^{n}\varepsilon_i X_{i-1}A_i^{-1}, \quad \sum_{i=1}^{n} X_{i-1}^2 A_i^{-1}, \tag{26}$$

(see (24), (25)). Because $\varepsilon_i^2 A_i^{-1} - \sigma^2 = U_i^2 - \sigma^2$, $i = 1, ..., n$ are i.i.d. with mean zero and finite variance, the first sum in (26) is easy to analyze: by the Strong Law of Large Numbers and the Central Limit Theorem, as $n \to \infty$,

$$n^{-1}\sum_{i=1}^{n}\Big(\varepsilon_i^2 A_i^{-1} - \sigma^2\Big) \to 0 \qquad \text{a.s.,} \tag{27}$$

$$\frac{1}{\sqrt{n}c\sigma^2}\sum_{i=1}^{n}\Big(\frac{\varepsilon_i^2}{A_i} - \sigma^2\Big) \Rightarrow N(0,1). \tag{28}$$

Consider then $W_{n1} = \sum_{i=1}^{n} \varepsilon_i X_{i-1} A_i^{-1} = \sum_{i=1}^{n} \tilde{\varepsilon}_i X_{i-1}$. Using (23) we write

$$W_{n1} = \sum_{i=1}^{n} \tilde{\varepsilon}_i \sum_{j=0}^{\infty} \Psi_j \varepsilon_{i-1-j} = \sum_{j=0}^{\infty} \Psi_j \sum_{i=1}^{n} \tilde{\varepsilon}_i \varepsilon_{i-1-j}.$$

For a fixed j, consider the random variables $\eta_i^{(j)} = \tilde{\varepsilon}_i \varepsilon_{i-1-j}$, $i = 1, 2, ..., n$, and the sum $\sum_{i=1}^{n} \eta_i^{(j)}$. For each $j \geq 0$, $\tilde{\varepsilon}_i$ and ε_{i-1-j} are independent and $\tilde{\varepsilon}_i$ has finite variance, so the marginal distribution of $\eta_i^{(j)}$ belongs to the DNA of a $S\alpha S$ random variable. For small values of n ($n \leq j+1$), the random variables $\eta_i^{(j)}$, $i = 1, 2, ..., n$ are independent. Nevertheless, to study the limiting relations as $n \to \infty$, we need to consider the case $j < n$. In the latter case, it is possible to rearrange $\eta_i^{(j)}$, $i = 1, 2, ..., n$ in such a way that they will be 1-dependent random variables. (Recall that a two-sided sequence $(X_k, k \in \mathbb{Z})$ of random vectors is said to be m–dependent if, for every $n \in \mathbb{N}$, the σ–algebras $\sigma(..., X_{n-1}, X_n)$ and $\sigma(X_{n+m+1}, X_{n+m+r}, ...)$ are independent.) To prove the relation

$$n^{-1/\alpha} \sum_{i=1}^{n} \eta_i^{(j)} \Rightarrow \zeta_j, \tag{29}$$

where ζ_j is an $S\alpha S$ random variable with scale parameter a_j depending on unknown σ, we apply a result of Davis (1983). (In our case the verification of conditions D and D' from that paper is standard, so we omit it.)

Using (29), for any fixed m we obtain the limiting relation

$$n^{-1/\alpha} \sum_{j=1}^{m} \Psi_j \sum_{i=1}^{n} \tilde{\varepsilon}_i \varepsilon_{i-1-j} \Rightarrow \sum_{j=1}^{m} \Psi_j \zeta_j \qquad \text{as} \quad n \to \infty,$$

for any fixed m. Then, applying Theorem 4.2 from Billingsley (1968), we conclude that

$$\lim_{n \to \infty} n^{-1/\alpha} W_{n1} = \sum_{j=1}^{\infty} \Psi_j \zeta_j := S_1. \tag{30}$$

S_1 is an $S\alpha S$ random variable with scale parameter $\left(\sum_1^{\infty} \Psi_j^{\alpha} a_j\right)^{1/\alpha}$ which depends on both unknown parameters b and σ. For similar calculations we refer to Davis and Resnick (1986).

Now we consider the third sum in (26). Let

$$W_{n2} := \sum_{i=1}^{n} X_{i-1}^2 A_i^{-1} = \sum_{i=1}^{n} A_i^{-1} \Big(\sum_{j=0}^{\infty} \Psi_j \varepsilon_{i-1-j} \Big)^2 = W_{n2}^{(1)} + W_{n2}^{(2)},$$

where

$$W_{n2}^{(1)} = \sum_{i=1}^{n} A_i^{-1} \sum_{j=0}^{\infty} \Psi_j^2 \varepsilon_{i-1-j}^2 = \sum_{j=0}^{\infty} \Psi_j^2 \sum_{i=1}^{n} A_i^{-1} \varepsilon_{i-1-j}^2,$$

$$W_{n2}^{(2)} = \sum_{i=1}^{n} A_i^{-1} \sum_{j \neq k} \Psi_j \Psi_k \varepsilon_{i-1-j} \varepsilon_{i-1-k}.$$

The analysis of $W_{n2}^{(1)}$ goes along the same lines as for the sum W_{n1}. Let $Z_i^{(j)} = A_i^{-1}\varepsilon_{i-1-j}^2$. Note that A_i^{-1} has finite moments of all orders and is independent of ε_{i-1-j}, therefore the marginal distribution of $Z_i^{(j)}$ (for all j) is in the DNA of an $\alpha/2$–stable random variable with $\beta = 1$, as $Z_i^{(j)} \geq 0$, for all i and j. The random variables $Z_i^{(j)}$ are independent if $j \geq n-1$ and (after rearrangement) at most 1–dependent, if $j < n-1$. Therefore, we obtain the following limiting relation

$$n^{-2/\alpha}\sum_{i=1}^{n} Z_i^{(j)} \Rightarrow \kappa_j,$$

where κ_j is a stable random variable with density $g_{\alpha/2}(x; d_j, 1)$ and d_j is a scale parameter depending on σ. Now, arguing in a similar way as in the derivation of (30), we show that

$$n^{-2/\alpha}W_{n2}^{(1)} \Rightarrow \sum_{j=0}^{\infty}\Psi_1^2\kappa_j := S_2. \tag{31}$$

Random variable S_2 has density $g_{\alpha/2}\big(x; \big(\sum\Psi_j^{2/\alpha}d_j\big)^{\alpha/2}, 1\big)$. It remains to show that

$$n^{-2/\alpha}W_{n2}^{(2)} = o_p(1). \tag{32}$$

For $\alpha > 1$, we use the bound:

$$\begin{aligned}
P\{|n^{-2/\alpha}W_{n2}^{(2)}| > \varepsilon\} &\leq \frac{1}{\varepsilon}n^{-2/\alpha}E|W_{n2}^{(2)}| \\
&\leq \frac{n^{-2/\alpha}}{\varepsilon}\sum_{j\neq k}\Psi_j\Psi_k\sum_{i=1}^{n}E\left|\frac{\varepsilon_{i-1-j}\varepsilon_{i-1-k}}{A_i}\right| \\
&\leq \frac{n^{-2/\alpha+1}}{\varepsilon}\big(E|\varepsilon_1|\big)^2E|A_i^{-1}|\sum_{j\neq k}\Psi_j\Psi_k \to 0, \quad \text{as } n\to\infty.
\end{aligned}$$

If $\alpha < 1$, we apply a similar moment inequality with some $\delta < \alpha$, and (32) is proved. Formulae (31) and (32) imply

$$n^{-2/\alpha}W_{n2} \Rightarrow S_2. \tag{33}$$

Although (30) and (33) only provide convergence results of the marginal distributions, it is not difficult to show that joint convergence holds: for example, one could use the Cramer–Wold device (see Billingsley (1968)). This leads to (12).

From (25), using relations (27), (28), (30), (33) and (12), we obtain

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_i - \sigma^2\right) + O_p\big(n^{-1/2}\big)\ ,$$

and this relation proves (11).

Note that the assumption of stationarity of the solution of model (1), based on the specific choice of X_0, is not essential. Namely, instead of (23), we can use the equality

$$X_n = b^n X_0 + \sum_{i=0}^{n-1} b^i \varepsilon_{n-i}.$$

Indeed, X_0 does not affect the limiting relations implying that we can get the same limiting result, making non-essential changes in the proof.

Now we consider the "unit root" case $b = 1$. Here $X_n = X_0 + \sum_{i=1}^{n} \varepsilon_i$ is indeed a nonstationary sequence. Formulae (24) and (25) remain valid setting $b = 1$. Again, we need to find the limiting distributions of the normalized sums (26). Consider the three-dimensional random vectors $\zeta_i = (\varepsilon_i, \tilde{\varepsilon}_i, \tilde{\varepsilon}_i^*)$, $i \geq 1$, where ε_i and $\tilde{\varepsilon}_i$ were introduced earlier and $\tilde{\varepsilon}_i^* = V_i^{-2}$. We have

$$\begin{aligned}
E\tilde{\varepsilon}_1 &= 0, \quad E\tilde{\varepsilon}_1^* = EV_1^{-2} = a = a(\alpha), \\
E\tilde{\varepsilon}_1^2 &= EU_1^2 EV_1^{-2} = \sigma^2 a, \\
E(\tilde{\varepsilon}_1^* - a)^2 &= EV_1^{-4} - a^2 := b_1^2, \\
E\tilde{\varepsilon}_1(\tilde{\varepsilon}_1^* - a) &= E\tilde{\varepsilon}_1\tilde{\varepsilon}_1^* = EU_1V_1^{-3} = 0, \\
E\varepsilon_1\tilde{\varepsilon}_1^* &= EU_1V_1^{-1} = 0, \\
E\varepsilon_1\tilde{\varepsilon}_1 &= EU_1^2 = \sigma^2.
\end{aligned}$$

These relations show that despite the fact that all three coordinates of ζ_1 are dependent, the pairs $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_1^*)$ and $(\varepsilon_1, \tilde{\varepsilon}_1^*)$ are uncorrelated, and only ε_1 and $\tilde{\varepsilon}_1$ are linearly correlated. Furthermore, the third coordinate of ζ_1 has positive mean, therefore, in order to construct the partial sum process, we need to use centering. Let $\bar{Z}_n(t) = \left(Z_n^{(1)}(t), Z_n^{(2)}(t), Z_n^{(3)}(t)\right)$, $0 \leq t \leq 1$, where $Z_n^{(j)}(t) = S_n^{(j)}([nt])$, $S_n^{(1)}(k) = \sigma^{-1} n^{-1/\alpha} \sum_{i=1}^{k} \varepsilon_i$, $S_n^{(2)}(k) = (an)^{-1/2}\sigma^{-1} \sum_{i=1}^{k} \tilde{\varepsilon}_i$, $S_n^{(3)}(k) = (nb_1)^{-1/2} \sum_{i=1}^{k} (\tilde{\varepsilon}_i^* - a)$. Let $D_3 \equiv D([0,1], \mathbb{R}^3)$ be the usual Skorohod space of cadlag functions on $[0,1]$ with values in $\mathbb{R}^3$. We recall that by $Y_\alpha(t)$, we denote a standard α–stable Lévy motion, i.e., $Y_\alpha(0) = 0$ a.s., Y_α has independent increments and $Y_\alpha(t) - Y_\alpha(s)$ is an $S\alpha S$ random variable having characteristic function (5) with $\sigma = (t-s)^{1/\alpha}$ for $s < t$. The following lemma is important for the rest of the proof.

Lemma 2 *For the partial sum process $\bar{Z}_n$, we have*

$$\bar{Z}_n \Rightarrow W \qquad \text{in} \quad D_3, \tag{34}$$

where $W(t) = \left(W_1(t), W_2(t), W_3(t)\right)$, $W_1(t) = Y_\alpha(t)$, W_2 and W_3 are standard Brownian motions. All three components of W are independent.

Proof of Lemma 2 is standard (see, for example, Mittnik, Paulauskas, Rachev (1999)). We shall only show the independence of the coordinates of

W. Independence of W_1 from W_2 and W_3 follows from Sharpe (1969), and the independence of W_2 and W_3 follows from their uncorrelatedness:

$$EW_2(t)W_3(t) = tE\frac{\tilde{\varepsilon}_1\tilde{\varepsilon}_1^*}{\sqrt{a}\sigma b_1} = 0.$$

Again, for simplicity of notation, we consider the marginal distributions only. We have

$$\begin{aligned}
\frac{1}{\sqrt{a\sigma^2}} n^{-\frac{1}{2}-\frac{1}{\alpha}} \sum_{i=1}^{n} \varepsilon_i X_{i-1} A_i^{-1} &= n^{-\frac{1}{2}-\frac{1}{\alpha}} a^{-\frac{1}{2}} \sigma^{-2} \sum_{i=1}^{n} \tilde{\varepsilon}_i \Big(X_0 + \sum_{j=1}^{i-1} \varepsilon_j \Big) \\
&= \sum_{i=1}^{n} \frac{\tilde{\varepsilon}_i}{\sqrt{na}\sigma} \Big(\frac{X_0}{n^{1/\alpha}\sigma} + \frac{1}{n^{1/\alpha}\sigma} \sum_{j=1}^{i-1} \varepsilon_j \Big) \\
&= \sum_{i=1}^{n} S_n^{(1)}(i-1)\big(S_n^{(2)}(i) - S_n^{(2)}(i-1)\big) + X_0 n^{-1/\alpha}\sigma^{-1} S_n^{(2)}(n) \\
&= \sum_{i=1}^{n} Z_n^{(1)}\Big(\frac{i-1}{n}\Big)\Big(Z_n^{(2)}\big(\frac{i}{n}\big) - Z_n^{(2)}\big(\frac{i-1}{n}\big)\Big) + O_p(n^{-1/2}) \\
&= \int_0^1 Z_n^{(1)}(t) dZ_n^{(2)}(t) + O_p\big(n^{-1/2}\big), \qquad (35)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n^{1+2/\alpha}\sigma^2 b_1} \sum_{i=1}^{n} A_i^{-1} X_{i-1}^2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(\tilde{\varepsilon}_i^* - a + a)}{\sqrt{n} b_1} \Big(\frac{X_0}{\sigma n^{1/\alpha}} + \frac{1}{\sigma n^{1/\alpha}} \sum_{j=1}^{i-1} \varepsilon_j \Big)^2 \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(S_n^{(3)}(i) - S_n^{(3)}(i-1)\Big)\Big(\frac{X_0}{n^{1/\alpha}\sigma} + S_n^{(1)}(b-1)\Big)^2 \\
&\quad + a \sum_{i=1}^{n} \frac{1}{n} \Big(\frac{X_0}{\sigma\sqrt{n}^{1/\alpha}} + S_n^{(1)}(i-1)\Big)^2 \\
&= a \int_0^1 \big(Z_n^{(1)}(u)\big)^2 du + O_p\big(\max n^{-1/\alpha}, n^{-1/2}\big). \qquad (36)
\end{aligned}$$

To get the limiting distribution in (35), we use results concerning convergence of stochastic integrals. This topic was discussed in detail in a similar setting in our previous work Paulauskas, Rachev (1998), so we do not verify here the so-called UT condition for the sequence $Z_n^{(2)}(t)$. The relations (34), (35) and (36) lead to

$$\frac{1}{\sigma^2\sqrt{an}n^{1/\alpha}} \sum_{i=1}^{n} \varepsilon_i X_{i-1} A_i^{-1} \Rightarrow \int_0^1 W_1(t) dW_2(t), \qquad (37)$$

$$\frac{1}{n^{1+2/\alpha}\sigma^2 b_1} \sum_{i=1}^{n} A_i^{-1} X_{i-1}^2 \Rightarrow a \int_0^1 W_1^2(u) du. \qquad (38)$$

From (25), using (37), (38) and the relation $\widehat{b}_n - 1 = O_p(n^{-1/2-1/\alpha})$, we obtain

$$\sqrt{n}\Big(\frac{\widehat{\sigma}_n^2}{\sigma^2} - 1\Big) = \frac{1}{\sqrt{n}\sigma^2}\sum_{i=1}^{n}(U_i^2 - \sigma^2) + O_p(n^{-1/2}). \tag{39}$$

Thus, we have (11) in the case $b = 1$. Formulae (37)–(39) prove (13) and the proof of Theorem 1 is completed.

Proof of Theorem 2. Theorem 2 is a straightforward generalization of the one–dimensional result, and we shall give only a sketch of the proof.

From (16) and (17) we have

$$\widehat{B}_n - B = C_n^{-1}\sum_{n=1}^{n}\frac{1}{V_i^2}(X_{i-1}\cdot\varepsilon_i'),$$

$$\widehat{\Sigma}_n - \Sigma = \frac{1}{n}\sum_{i=1}^{n}\big(U_iU_i' - \Sigma\big) + J_1 + J_2 + J_3,$$

where

$$J_1 = \frac{1}{n}\Big(\sum_{i=1}^{n}U_i\cdot\Big(\frac{X_{i-1}'}{V_i}\Big)\Big)(B - \widehat{B}_n),$$

$$J_2 = (B - \widehat{B}_n)\frac{1}{n}\sum_{i=1}^{n}\frac{X_{i-1}}{V_i}U_i',$$

$$J_3 = \frac{1}{n}(B - \widehat{B}_n)\sum_{i=1}^{n}\frac{X_{i-1}}{V_i}\Big(\frac{X_{i-1}}{V_i}\Big)'(B - \widehat{B}_n)',$$

$$C_n = \sum_{i=1}^{n}\frac{1}{V_i^2}(X_{i-1}\cdot X_{i-1}').$$

In order to prove (18), we need to consider the joint distribution of (C_n, W_n), where $W_n = \sum_{i=1}^{n}V_i^{-2}(X_{i-1}\cdot\varepsilon_i')$. Again, for simplicity of notation, we study the marginal distributions only. Denoting $\tilde{\varepsilon}_i = U_iV_i^{-1}$, we have

$$W_n = \sum_{i=1}^{n}X_{i-1}\tilde{\varepsilon}_i'.$$

Under the assumption $\|B\| < 1$, without loss of generality, we assume that

$$X_n = \sum_{j=0}^{\infty}B^j\varepsilon_{n-j}.$$

Then

$$n^{-1/\alpha}W_n = \sum_{j=0}^{\infty}B^jn^{-1/\alpha}\sum_{i=1}^{n}\varepsilon_{i-1-j}\cdot\tilde{\varepsilon}_i'.$$

Similarly to the one-dimensional case, we show that

$$n^{-1/\alpha} W_n \Rightarrow Z_1, \tag{40}$$

where $Z_1 = \left(Z_{ij}^{(1)}\right)_{i,j=1\dots k}$ is jointly α-stable vector with scale parameters depending on the unknown B and Σ, (here we use results of Jakubowski and Kobus (1984)), A complete description of this vector is rather complicated and we do not provide it here (see Theorem 5.3 in Jakubowski and Kobus (1984)).

Similarly,

$$C_n = \sum_{i=1}^{n} \frac{1}{V_i^2} (X_{i-1} \cdot X_{i-1}') = \sum_{j,\ell} B^j \Big(\sum_{i=1}^{n} \frac{1}{V_i^2} \varepsilon_{i-1-j} \varepsilon_{i-1-\ell}' \Big) B'^{\ell} = C_{n1} + C_{n2},$$

where

$$C_{n1} = \sum_{j=0}^{\infty} B^j \Big(\sum_{i=1}^{n} \frac{1}{V_i^2} \varepsilon_{i-1-j} \varepsilon_{i-1-j}' \Big) B'^{j},$$

$$C_{n2} = \sum_{j \neq \ell} B^j \Big(\sum_{i=1}^{n} \frac{1}{V_i^2} \varepsilon_{i-1-j} \varepsilon_{i-1-\ell}' \Big) B'^{\ell}.$$

Since V_i^{-2} has all moments and the distributional tail behavior of the entries of the matrix $\varepsilon_{i-1-j}\varepsilon_{i-1-j}'$ is different on the diagonal and outside the diagonal, it is not difficult to check that there exist diagonal random matrices T_j with positive $\alpha/2$–stable random variables $T_j^{(i)}$, $i = 1, \dots, k$, on the diagonal such that

$$n^{-2/\alpha} \sum_{i=1}^{n} \frac{1}{V_i^2} (\varepsilon_{i-1-j} \varepsilon_{i-1-j}') \Rightarrow T_j$$

and

$$n^{-2/\alpha} C_{n1} \Rightarrow \sum_{j=0}^{\infty} B^j T_j B'^j = Z_2.$$

As in the one-dimensional case, we show that

$$n^{-2/\alpha} C_{n2} = o_p(1).$$

Thus, we obtain

$$n^{-2/\alpha} C_n \Rightarrow Z_2. \tag{41}$$

From (40), (41) we get (18).

To prove (19), note that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i U_i' - \Sigma) \Rightarrow Z_3$$

and $J_i = o_p(1)$ for $i = 1, 2, 3$. Theorem 2 is now proved.

Acknowledgment. The authors are most grateful to Carlo Marinelli and the referee for a careful reading of the manuscript and for suggesting many improvements and clarifications.

References

1. Anderson T. W. (1959), On asymptotic distributions of estimates of parameters of stochastic difference equations, *Annals of Mathematical Statistics*, **30**, 676–687.
2. Billingsley P. (1968), *Convergence of Probability Measures*, Wiley, New York.
3. Chan N. H. and Tran L. T. (1989), On the first order autoregressive process with infinite variance, *Econometric Theory*, **5**, 354–362.
4. Davis R. (1983), Stable limits for partial sums of dependent random variables, *Annals of Probability*, **11**, 262–269.
5. Davis R. and Resnick S. (1986), Limit theory for the sample covariance and correlation functions of moving averages, *Annals of Statistics*, **14**, 543–558.
6. Davydov Yu., Paulauskas V., and Račkauskas A. (2000), More on P-Stable Convex Sets in Banach Spaces, *J. Theoretical Probab.*, **13**, 39-64.
7. Feller W. (1971), *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York, 2nd edition.
8. Jakubowski A. and Kobus M. (1984), α–stable limit theorems for sums of dependent random vectors, *Journal of Multivariate Analysis*, **29**, 219–251.
9. Johansen S. (1996), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Advanced Texts in Econometric, Oxford UP, Oxford UK.
10. Johansen S. (1988), Statistical analysis of cointegration vectors, *Journal of Economics and Dynamics and Control*, **12**, 231–254.
11. Jurek Z. J. and Mason J. D. (1993), *Operator-Limit Distributions in Probability Theory*, John Wiley & Sons, New York.
12. Mann H. B. and Wald A. (1943), On the statistical treatment of linear stochastic difference equations, *Econometrica*, **11**, 173–220.
13. McCulloch J.H. (1996), Financial Applications of Stable Distributions, in Maddala G.S. and Rao C.R. (eds), *Handbook of Statistics*, **14**,393-425.
14. Mijnheer J. (1997), Asymptotic inference for AR(1) processes with (non–normal) stable errors, *Journal of Mathematical Sciences*, **83**, 401–406.
15. Mittnik S., Paulauskas V. and Rachev S. T. (1999), Statistical Inference in Regression With Heavy–Tailed Integrated Variables, in Rachev S.T. and Mittnik S. (eds.), *Stable Models in Finance*, Pergamon Press, to appear.
16. Mittnik S., Rachev S. T. and Paolella M. (1998), Stable Paretian modeling in finance: some empirical and theoretical aspects, in R. Adler, R. E. Feldman and M. S. Taqqu (eds.), *A Practical Guide to Heavy Tails*, pp. 79–110, Birkhäuser, Boston.
17. Nolan J. P. (1997), Maximum likelihood estimation and diagnostics for stable distributions, Preprint.
18. Paulauskas V. and Rachev S. T. (1998), Cointegrated processes with infinite variance innovations, *Annals of Applied Probability*, **8**, 775–792.
19. Park J. Y. and Phillips P. C. B. (1988), Statistical inference in regressions with integrated processes: part I, *Econometric Theory*, **4**, 468–497.
20. Phillips P. C. B. (1987), Towards a unified asymptotic theory for autoregression, *Biometrika*, **74**, 535–547.

21. Rachev S. T., Kim J-R. and Mittnik S. (1997), Econometric modeling in the presence of heavy-tailed innovations: a survey of some recent advances, *Stochastic Models*, **13**, 841–866.
22. Samorodnitsky G. and Taqqu M. S. (1994), *Stable non-Gaussian Random Processes: Models with Infinite Variance*, Chapman and Hall, New York.
23. Sharpe M. (1969), Operator-stable probability distributions on vector groups, *Transactions of the American Mathematical Society*, **136**, 51–65.
24. Zolotarev V. M. (1986), *One–dimensional Stable Distributions*, American Mathematical Society, Providence, RI.
25. White J. S. (1958), The limiting distribution of the serial correlation coefficient in the explosive case, *Annals of Mathematical Statistics*, **29**, 1188–1197.

V. Paulauskas
Department of Mathematics
Vilnius University
Naugarduko 24
Vilnius 2006
Lithuania
E-mail: vpaul@ieva.maf.vu.lt

S.T. Rachev
Institut of Statistics
and Mathematical Economics
Department of Economics
University of Karlsrhue
Kollegium am Schloss
D-76128 Karlsruhe
Germany
E-mail: rachev@lsoe.uni-karlsruhe.de