ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES **6-2015**

A Note on Mixed Stable Limit Theory for the OLSE and the Robustness of Subsampling Wald Tests

Stelios Arvanitis

76 Patission Str., Athens 104 34, Greece
Tel. (++30) 210-8203911 - Fax: (++30) 210-8203301
www.econ.aueb.gr

A Note on Mixed Stable Limit Theory for the OLSE and the Robustness of Subsampling Wald Tests

Stelios Arvanitis

Department of Economics, Athens University of Economics and Business.

Patision 76, P.O. Box 10434, Athens, Greece.

Abstract

We are occupied with a simple example concerning the limit theory of the OLSE when the innovation process of the regression has the form of a martingale transform the i.i.d. part of which lies in the domain of attraction of an α-stable distribution, the scalling sequence has a potentially diverging truncated α-moment and the regressor process is asymptotically stationary. We obtain rates that reflect the stability parameter as well as the slow variations present in the aforementioned sequences and mixtures of stable limits. We also derive asymptotic exactness, consistency as well as local asymptotic unbiaseness under appropriate local alternatives for relevant Wald tests derived by subsampling.

Keywords: OLSE, Innovation Process, Martingale Transform, Martingale Limit Theorem, Stable Distribution, Mixtures, Domain of Attraction, Slow Variation, Asymptotic Stationarity, Non Standard Rate, Wald Test, Subsampling, Asymptotic Exactness, Consistency, Local Alternatives, Asymptotic Locally Unbiased Test.

Acknowledgement. The author acknowledges financial support from the Greek General Secretariat for Research & Technology under the grand ARISTEIA II-08-5413.

1 Introduction

In this note we are occupied with a simple example concerning the limit theory of the OLSE when the innovation process of the regression has the form of a martingale transform the i.i.d. part of which lies in the domain of attraction of an α-stable distribution, the scalling sequence has a potentially diverging truncated α-moment and the regressor process is asymptotically stationary.

Our motivation is twofold. First we are interested in the development of limit theorems to stable limits when the rate contains also information about the aforementioned divergence whilst the scalling sequence itself has a multiplicative structure. Second we are interested in the question of asymptotic robustness of Wald tests based on subsampling in such a framework.

The issue of the limit theory for parameter estimators in linear models, autoregressions, moving average processes e.t.c. with heavy tailed innovations is vast. We incompletely mention only Knight [9] as well as Kokoszka and Taqqu [10] and direct to Hall and Yao [6] (p. 291) for a comprehensive list of references. We note that in the relevant literature the innovation process is usually assumed i.i.d. following a stable distribution. In the present paper we allow for stationary martingale transforms with marginals that belong to the general domain of attraction of a stable law. This implies among others that the rates that we obtain are generally more complex. Furthermore, even though our assumption framework does not allow for ARMA type of regressors[1] our framework allows for the extraction of limits that are mixtures of stable distributions with common stability parameters.

The limit theory for regularly varying martingale transforms with applications to the limit theory of the QMLE in GARCH-type modes, has been the subject of a literature consisting of papers such as the ones of Hall and Yao [6], Mikosch and Straumann [11] or Surgailis [15]. There the martingale transform is consisted by an i.i.d. process supposed to lie in the domain of (normal in the latter case) attraction of an α-stable distribution while the scalling process is a stationary ergodic sequence with either high enough moments and/or obeying some mixing condition. The limits are α stable distributions in some cases partially characterized. Arvanitis and Louka [1] generalize the aforementioned results by retaining only stationarity for the scalling process while requiring only the α moment to exist, albeit with some restrictions to the marginal regular variation of the i.i.d. process when higher moments do not exist for the scalling process. They obtain mixed α-stable distributions while they are able to obtain analogous results in some trivial cases of non-stationarity.

In the present note we rely on a limit theorem partially involving an analogous transform. We generalize the aforementioned result in the following ways. First we allow for the scalling process to have itself a multiplicative structure in order to avoid the i.i.d. assumption for the innovation process of the regression model. Second we allow for the latter process to be in a strong sense only asymptotically stationary. Third, we allow for the scalling process to have a slowly varying trunctated α-moment that is allowed to diverge. In this respect we obtain more complex rates than in the aforementioned literature. We derive our results by allowing a more restrictive framework for the regular variation of the i.i.d. sequence in comparison to the aforementioned paper.

The assumption framework, the aforementioned limit theorem and the limit theory of the relevant OLSE is presented in the following section. In the final one, based on the previous, we derive asymptotic exactness, consistency as well as local asymptotic unbiaseness under appropriate local alternatives for relevant Wald type testing procedures derived by subsampling.

[1] except for the case of the domain of attraction to the normal law.

2 Assumption Framework and OLSE Limit Theory

We employ the framework of an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In what follows, we are among others interested in the asymptotic behavior of the partial sums of a martingale transform type process of the form $(X_i v_i \xi_i)_{i\in\mathbb{N}}$. The assumption framework presented below, essentially specifies properties of the constituent factor processes.

Assumption 1. *$(\xi_i)_{i\in\mathbb{N}}$ is stationary. The $\log$-characteristic function of ξ_0 has the following representation locally around zero:*

$$-c|t|^{\alpha}h(1/|t|)\left(1-ib\,\mathrm{sgn}(t)\tan\left(\frac{1}{2}\pi\alpha\right)\right).$$

$\alpha\in(1,2]$, $b\in[-1,1]$, $c\in\mathbb{R}_{++}$, h is slowly varying at infinity. If h diverges then there exist m and g such that $\left|\frac{h(tx)}{h(x)}-1\right|\leq m(t)\,g(x)$ for all t in a neighborhood of infinity and large enough x, where m is increasing while $t^{-\alpha}m(t)$ is eventually bounded and $g(x)\to 0$ as $x\to\infty$. $\lim_{x\to\infty}h(x)>0$ if it exists.

The representation in Assumption 1 holds iff the distribution of ξ_0 lies in the domain of attraction of an α-stable law (Theorem 2.6.5 of Ibragimov and Linnik [8]). α is the index of stability, b the skewness and c the scale parameter. The aforementioned Tauberian theorem implies that α, b, c and the slowly varying function h represent the asymptotic behavior of the tails of the distribution of ξ_0. The existence of the pair m, g is ensured if h belongs to the *Zygmund* class of slowly varying functions (see Theorem 1.5.5 of Bingham et. al. [2]) and $xh'(x)$, is Lebesgue almost everywhere bounded. Examples are $C\frac{(\ln^{(m_1)}(x))^{\delta_1}}{(\ln^{(m_2)}(x))^{\delta_2}}$ for $\delta_1\in[0,1]$, $\delta_2\geq 0$, $C(\ln^{(m)}(\exp(1)+x))^{\delta}$ for $\delta\in[0,1]$. In any case the restrictions on the behavior of h imply that ξ_0 does not have moments of order greater or equal to α.

Assumption 2. *$(v_i)_{i\in\mathbb{N}}$ is a stationary, ergodic $\mathbb{R}_{++}$-valued process, for which*

$$\mathbb{P}\left(v_0^{\alpha}>x\right)=o\left(\mathbb{P}\left(|\xi_0|^{\alpha}>x\right)\right),\ \textit{as}\ x\to+\infty.$$

Define $(\kappa_n)_{n\in\mathbb{N}}$ by

$$(n\kappa_n)^{-1/\alpha}\doteqdot\sup\left\{x>0: x^{\alpha}h^{\star}\left(x^{-1}\right)\leq 1/n\right\},$$

for $h^{\star}(x)\doteqdot\mathbb{E}\left[v_0^{\alpha}1_{\{v_i\leq x\}}\right]$. If κ_n diverges then

$$\sum_{i=1}^{n}\mathbb{C}ov\left(v_0^{\star},v_i^{\star}\right)=o\left(n\kappa_n^2\right),$$

for $v_i^{\star}\doteqdot v_i^{\alpha}1_{\left\{v_i\leq(n\kappa_n)^{1/\alpha}\right\}}$.

$(v_i)_{i\in\mathbb{N}}$ is (part of) the scaling process of the transform. v_0 has a regularly varying tail of index at least $-\alpha$ and its α moment may not exist. In this case, i.e. when κ_n diverges, the covariance summability condition facilitates the derivation of a Generalized LLN that extends Theorem 2 in Section VII.7 of Feller [4] (see Lemmata 2 and 4 in the Appendix). It is implied by the previous and uniform mixing for the process with summable coefficients, by Theorem 1.4.(b) of Rio [13]. The existence of $(\kappa_i)_{i\in\mathbb{N}}$ is ensured by the right-continuity of $h^\star$ as in Theorem 2 in Section VII.7 of Feller [4].

Assumption 3. *$X_i = K_i + A_i$. $(K_i)_{i\in\mathbb{N}}$ is a stationary $\mathbb{R}^q$-valued process and for some $\Delta > 0$, $\mathbb{E}\left(\|K_0\|^{2+\Delta} < +\infty\right)$. For $\mathcal{I}$ the invariant σ-field of the process and $\mathbb{V} = \mathbb{E}\left(K_0 K_0^{Tr}/\mathcal{I}\right)$*

$$\textit{rank}\,[\mathbb{V}] = q,\ \mathbb{P}\ \textit{a.s.}$$

$(A_i)_{i\in\mathbb{N}}$ *is a $\mathbb{R}^q$-valued process*, such that $A_i A_i^{Tr}$ is uniformly integrable and converges to 0 $\mathbb{P}$ a.s.

Assumption 3 specifies the asymptotically stationary regressor process as the pointwise sum of a stationary process and an asymptotically $\mathbb{L}_2$-degenerate at zero process. The LLN of Doob implies convergence of the mean of the squared stationary part to the relevant non degenerate conditional expectation. The existence of at least the second moment of X_i cannot be easily dispensed, except for the case where $\alpha = 2$, due to the restriction that shoud be imposed on the distribution of $K_i K_i^{Tr}$ by the aforementioned Generalized LLN and our interest for the asymptotic behavior of $\frac{1}{n}\sum_{i=1}^n X_i X_i^{Tr}$. The condition on the boundness of the sequence of second moments facilitates among others results on the variation of the "tails" of the (vector) martingale transform $(X_i v_i \xi_i)_{i\in\mathbb{N}}$.

Assumption 4. *ξ_i is independent of $\mathcal{F}_i \doteqdot \sigma\left(\xi_{i-j} v_{i-j} X_{i-j}, j > 0\right)$, and conditionally on $\mathcal{F}_i$, ξ_i, v_i, X_i are mutually independent $\forall i \in \mathbb{N}$.*

Assumption 4 establishes the dependence structure between the processes involved. It enables the use of the Principle of Conditioning of Jakubowski [7]. It does not preclude non-contemporaneous dependence between the $(v_i)_{i\in\mathbb{N}}$ and the $(X_i)_{i\in\mathbb{N}}$ processes.

Consider $(y_i)_{i\in\mathbb{N}}$ specified by

$$y_i = X_i^{Tr}\beta_0 + \xi_i v_i,$$

where $\beta_0 \in \mathbb{R}^q$. Given the random element $(y_i, X_i)_{i=1}^n$ we are interested in the asymptotic behavior of the OLSE for β_0, i.e.

$$\beta_n = \left(\sum_{i=1}^n X_i X_i^{Tr}\right)^{-1} \sum_{i=1}^n X_i y_i = \beta_0 + \left(\sum_{i=1}^n X_i X_i^{Tr}\right)^{-1} \sum_{i=1}^n X_i v_i \xi_i.$$

This is established in the following proposition the basic part of which is essentially a joint limit theorem for the partial sums of a martingale transform and the quadratic variation of the regressors. In what follows $S_\alpha(\beta, c, \gamma)$ denotes the α-stable distribution on $\mathbb{R}$ with b, c

as above and location parameter γ.[2] When $\alpha = 2$, necessarily $\beta = 0$ and the resulting distribution is the $N(\gamma, 2c)$. In such a case and without loss of generality, we assume that $c = \frac{1}{2}$.

Furthermore the notation $\mathbb{E}\left[S_{\alpha}\left(b, c, \gamma\right)\right]$ denotes the mixture of the distributions of $S_{\alpha}\left(b, c, \gamma\right)$ w.r.t. $\mathbb{P}$ given that we allow (for some of) those parameters to be $\mathcal{F}$-measurable non-constant functions defined on Ω. Finally, the weak limits appearing below, are mixtures of multivariate α-stable distributions where the (random) spectral measures (for their definition see Paragraph 2 of Mikosch and Straumann [11]) are characterized by the action of the set of non-trivial linear transformations on the limiting random vectors due to Theorem 2.3 of Gupta et. al. [5].

Proposition 1. *Under Assumptions 1-4 for any* $\lambda \in \mathbb{R}^q - \left\{0_{q\times 1}\right\}$

$$\left(\frac{1}{\left[nr_n\kappa_n\right]^{1/\alpha}}\sum_{i=1}^{n} X_i v_i \xi_i, \frac{1}{n}\sum_{i=1}^{n} X_i X_i^{Tr}\right) \rightsquigarrow (\zeta, \mathbb{V})$$

and

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{\left[r_n\kappa_n\right]^{1/\alpha}}\left(\beta_n - \beta_0\right) \rightsquigarrow \mathbb{V}^{-1}\zeta$$

where

$$\lambda^{Tr}\zeta \backsim \mathbb{E}\left[S_{\alpha}\left(b_{\lambda,\mathcal{J}}, c_{\lambda,\mathcal{J}}, 0\right)\right],$$

$b_{\lambda,\mathcal{J}} = b\frac{\mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^{\alpha}\operatorname{sgn}(\lambda^{Tr}K_0)/\mathcal{J}\right]}{\mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^{\alpha}\right]}$, $c_{\lambda,\mathcal{J}} = c\mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^{\alpha}/\mathcal{J}\right]$ *and* $(r_n)_{n\in\mathbb{N}}$ *is defined by*

$$\left(n\kappa_n r_n\right)^{-1/\alpha} \doteqdot \sup\left\{x > 0 : x^{\alpha}h\left(x^{-1}\right) \leq 1/n\kappa_n\right\}.$$

Remark 1. a. The result on the joint weak convergence generalizes current results on stable weak limits of martingale transforms, in the sense that it allows for the non-existence of the α-moment of the scalling process of the transform, a fact that is to our knowledge novel in the relevant literature. In this respect the rate contains information on the tail properties of the distribution of ξ_0, as well as on the asymptotic behavior of $h^{\star}$. In the special case where $h^{\star}, h$ converge we obtain the standard rate $n^{1/\alpha}$. The rate of the OLSE also contains the aforementioned information and it is slower than $n^{1/\alpha}$ when $h^{\star}$ and/or h diverge.

b. The existence of $(r_n)_{n\in\mathbb{N}}$ can be established by analogous arguments to the ones in the proof of Theorem 2.6.5 in Ibragimov and Linnik .

c. The weak limit of the transform is a multivariate mixture of α-stable distributions. The mixture is trivial iff $\mathcal{J}$ is trivial, i.e. $(K_i)_{i\in\mathbb{N}}$ is ergodic, whence $\mathbb{V}$ is a constant positive definite matrix. When the latter is not trivial, the stability and the location parameters of the limit are independent of $\mathcal{J}$. α is also independent of K_0. In any case the spectral

[2] Which is equivalently defined by its characteristic function $\gamma it - c|t|^{\alpha}\left(1 - ib\operatorname{sgn}(t)\tan\left(\frac{1}{2}\pi\alpha\right)\right)$.

measures depends only on the relevant moments of K_0 and the analogous parameters of the distribution of ξ_0. Analogously the spectral measures of the mixing distributions characterizing the asymptotic distribution of the OLSE obviously depend also on $\mathbb{V}^{-1}$.

d. When $\alpha = 2$, $\zeta \backsim \mathbb{E}\left[N\left(0_q, \mathbb{V}\right)\right]$, whence $\sqrt{\frac{n}{r_n \kappa_n}}\left(\beta_n - \beta_0\right) \rightsquigarrow \mathbb{E}\left[N\left(0_q, \mathbb{V}^{-1}\right)\right]$. If $\kappa_n$and/or r_n diverge the rate is slower than the usual $\sqrt{n}$. For example suppose that $\xi_0 \sim t_2$, and $\left(v_i^2\right)_{i \in \mathbb{N}}$is a stationary ergodic conditional variance process in the context of some GARCH type model with finite second moment. Then some calculations imply that $\sqrt{\frac{n}{\ln n}}\left(\beta_n - \beta_0\right) \rightsquigarrow \mathbb{E}\left[N\left(0_q, 2\mathbb{E}\left(v_0^2\right)\mathbb{V}^{-1}\right)\right]$.

e. It can be easily proven that the results above are valid when $\alpha \in (0,1)$ since the local representation in Assumption 1 holds also in this case. Then a simple inspection implies that β_n is asymptotically non-tight and thereby *inconsistent*. The same would be true when $\alpha = 1$ which as a matter of fact would be a more preplexed case since it would involve a different local representation as well as a sequence of translating random vectors. Since those cases would not be relevant to the discussion below about the limit properties of Wald-type tests,[3] we do not pursue them to economize on space.

f. Suppose that for some $\beta^{\star} \in \mathbb{R}^q$ we have that $\beta_{0,n} = \beta_0 + \frac{\left[r_n \kappa_n\right]^{1/\alpha}}{n^{\frac{\alpha-1}{\alpha}}}\beta^{\star}$. The previous readily imply that under the law implied by $\beta_{0,n}$

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{\left[r_n \kappa_n\right]^{1/\alpha}}\left(\beta_n - \beta_0\right) \rightsquigarrow \beta^{\star} + \mathbb{V}^{-1}\zeta,\ \lambda^{Tr}\left(\beta^{\star} + \mathbb{V}^{-1}\zeta\right) \backsim \mathbb{E}\left[S_\alpha\left(b_{\lambda,\mathcal{J}}, c_{\lambda,\mathcal{J}}, \lambda^{Tr}\beta^{\star}\right)\right]$$

a result useful for the derivation of power properties under local alternatives for the testing procedures discussed below.

Proof. By the "Main Lemma for Sequences" of Jakubowski [7], the Cramer-Wold device, Doob's LLN, and Assumption 4 for the first result it suffices that

$$\frac{1}{\left[n r_n \kappa_n\right]^{1/\alpha}} \sum_{i=1}^{n} \lambda^{Tr} X_i v_i \xi_i \rightsquigarrow \lambda^{Tr}\zeta.$$

This follows if for all $t \in \mathbb{R}$

$$q_n(t) \doteqdot \prod_{i=1}^{n} \mathbb{E}\left(\exp\left(it \frac{1}{\left[n r_n \kappa_n\right]^{1/\alpha}} \lambda^{Tr} X_i v_i\right) / \mathcal{F}_i\right)$$

converges in probability to the cf of

$$S_\alpha\left(\beta \frac{\mathbb{E}\left[\left|\lambda^{Tr} K_0\right|^\alpha \operatorname{sgn}(\lambda^{Tr} K_0) / \mathcal{J}\right]}{\mathbb{E}\left[\left|\lambda^{Tr} K_0\right|^\alpha\right]}, c\mathbb{E}\left[\left|\lambda^{Tr} K_0\right|^\alpha / \mathcal{J}\right], 0\right).$$

[3] It is easy to see that in those cases Lemma 6 in the Appendix would seize to hold.

Suppose that the representation in Assumption 1 holds for all $t \in (-t_0, t_0)$, where $t_0 > 0$. Then for any $t \neq 0$, defining the event

$$C_{n,K} := \left\{ \omega \in \Omega : |\lambda^{Tr} X_i v_i| \leq M_t \left(n r_n\right)^{\frac{1}{\alpha}} \quad , \forall i = 1, \dots, n \right\}$$

where $M_t < \frac{t_0}{|t|}$, we have that $\mathbb{P}(C^c_{n,M}) \to 0$ by Lemma 5 as $n \to \infty$. Due to Assumption 1 if $\omega \in C_{n,M}$

$$\log q_n\left(t\right) = -\frac{c|t|^\alpha}{n r_n \kappa_n} \sum_{i=1}^{n} \left|t\lambda^{Tr} X_i v_i\right|^\alpha h\left(\left[n r_n \kappa_n\right]^{1/\alpha} \left|t\lambda^{Tr} X_i v_i\right|^{-1}\right) \left(1 - i\beta \operatorname{sgn}(t\lambda^{Tr} X_i) \tan\left(\frac{1}{2}\pi\alpha\right)\right).$$

Notice that for any $\delta, \epsilon > 0$, and $V_i = \left|t\lambda^{Tr} X_i v_i\right|^\alpha$ or $V_i = \left|t\lambda^{Tr} X_i v_i\right|^\alpha \operatorname{sgn}(t\lambda^{Tr} X_i)$

$$\begin{aligned}
&\mathbb{P}\left(\frac{h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)}{n \kappa_n r_n} \sum_{i=1}^{n} |V_i| \left|\frac{h\left(\left[n r_n \kappa_n\right]^{1/\alpha} \left|t\lambda^{Tr} X_i v_i\right|^{-1}\right)}{h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)} - 1\right| > \delta\right) \\
&\qquad \leq \mathbb{P}\left(B_n^\epsilon \frac{h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)}{r_n} \frac{1}{n\kappa_n} \sum_{i=1}^{n} |V_i| > \delta\right) \\
&\qquad + \mathbb{P}\left(C_n^\epsilon \frac{h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)}{r_n} q_n > \delta, \max_{i \leq n} \left|t\lambda^{Tr} X_i v_i\right| \leq \epsilon\right)
\end{aligned}$$

where

$$B_n^\epsilon = \epsilon^\alpha \max_{\frac{1}{\left[n r_n \kappa_n\right]^{1/\alpha}} \leq x \leq \frac{1}{\epsilon}} \left|\frac{\left[n r_n \kappa_n\right] x^\alpha h\left(\left[n r_n \kappa_n\right]^{1/\alpha} x\right)}{\left[n r_n \kappa_n\right] h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)} - x^\alpha\right|,$$

and due to the UCT for regularly varying functions (see Theorem 1.5.2 of Bingham et al. [2]) $B_n^\epsilon \to 0$. If h converges then

$$C_n^\epsilon = \max_{\frac{1}{\epsilon} \leq x < +\infty} \left|\frac{h\left(\left[n r_n \kappa_n\right]^{1/\alpha} x\right)}{h\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right)} - 1\right| \to 0, \; q_n = \tfrac{1}{n\kappa_n} \textstyle\sum_{i=1}^{n} |V_i|.$$

If h diverges since the pair m, g exists, there exists a $C > 0$,

$$C_n^\epsilon = g\left(\left[n r_n \kappa_n\right]^{1/\alpha}\right) \to 0, \; q_n = \tfrac{C}{\kappa_n}.$$

In every case the result follows from Assumptions 1, 3 and Lemma 4. The second follows from the previous and the CMT. □

3 Application: Robust Self-Normalized Subsampling Wald Tests

In the premise of the linear model

$$y_i = X_i\beta + \xi_i v_i,\ \beta \in \mathbb{R}^q$$

consider for some $\beta_\star \in \mathbb{R}^q$ the hypothesis structure

$$\mathbf{H}_0 : \beta_0 = \beta_\star,$$
$$\mathbf{H}_{\mathbf{alt}} : \beta_0 \neq \beta_\star.$$

Also, given the notation established in Remark 1.6 consider the local alternative hypothesis

$$\mathbf{H}_{\mathbf{alt},\mathbf{n}} : \beta_0 := \beta_{0,n} = \beta_\star + \frac{[r_n\kappa_n]^{1/\alpha}}{n^{\frac{\alpha-1}{\alpha}}}\beta^\star.$$

Notice that in our Assumption framework the asymptotic exactness of the usual Wald test for this structure, based on the asymptotic chi-squared distribution becomes generally invalidated. Proposition 1 provides with a way to robustify the procedure as the following Lemma implies.

Lemma 1. *Given Assumption 1, if $\alpha < 2$ suppose that $(K_i)_{i\in\mathbb{N}}$ is ergodic. Then,*

$$\sum_{i=1}^{n}\left(\frac{X_i v_i \xi_i}{[nr_n\kappa_n]^{1/\alpha}}, \frac{v_i^2\xi_i^2}{[nr_n\kappa_n]^{2/\alpha}}\right) \rightsquigarrow (\zeta, \varsigma),$$

where $\varsigma = 1$ when $\alpha = 2$. When $\alpha \neq 2$, $\varsigma \backsim S_{\alpha/2}$ which has non-negative support and the distribution of (ζ, ς) is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^2$. Also under the law implied by $\beta_{0,n}$

$$\eta_n \doteqdot n(\beta_n - \beta_\star)^{Tr}\frac{\sum_{i=1}^n X_i X_i^{Tr}}{\sum_{i=1}^n \epsilon_i^2}(\beta_n - \beta_\star) \rightsquigarrow \frac{(\beta^\star + \zeta)^{Tr}\mathbb{V}^{-1}(\beta^\star + \zeta)}{\zeta_2},$$

where $\epsilon_i \doteqdot y_i - X_i^{Tr}\beta_n$, $i = 1, \ldots, n$.

Proof. When $\alpha = 2$ we have that $\sum_{i=1}^n \frac{v_i^2\xi_i^2}{nr_n\kappa_n} \rightsquigarrow 1$ in a similar spirit to the proofs of lemmata 2 and 4. When $\alpha \neq 2$ the result follows as in the proofs of Theorems 2.1.c,e and 3.1 of Hall and Yao [6] by noting that $\mathcal{J}$ is trivial. □

The last result obviously provides with the asymptotic distribution of the self-normalized Wald test under $\mathbf{H}_0$. Notice that if $\alpha = 2$ the limit distribution is χ_q^2 even in the cases where the *second moments of both processes do not exist and/or we have mixed normality* due to the stochasticity of $\mathbb{V}$. Hence in this case the classical test remains asymptotically exact and consistent. Furthermore it is easy to see that it is locally asymptotically unbiased but w.r.t. the sequences described in $\mathbf{H}_{\mathbf{alt},\mathbf{n}}$.

This seizes to be true when $\alpha \neq 2$. Hence under our assumption framework in order for a feasible testing procedure to be established, an approximation of the relevant quantiles of the aforementioned distribution is needed. The following algorithm provides the well known modification based on subsampling.

Algorithm 1. *The testing procedure consists of the following steps:*

a. *Evaluate* η_n at the original sample value.

b. *For* $0 < c_n \leq n$ *generate subsamples from the original observations* $(\; Y_i, \quad X_i \;)_{i=t,\dots,t+c_n-1}$ *for all* $t = 1, 2, \dots, n - c_n + 1$.

c. *Evaluate the test statistic on each subsample thereby obtaining* $\eta_{n,c_n,t}$ *for the subsample indexed by* $t = 1, 2, \dots, n - c_n + 1$.

d. *Approximate the cdf of the asymptotic distribution under the null of* η_n by $s_{n,c_n}(y) = \frac{1}{n-c_n+1} \sum_{t=1}^{n-c_n+1} 1\left(\eta_{n,c_n,t} \leq y\right)$ and for $a \in (0,1)$ calculate

$$q_{n,c_n}(1-a) = \inf_y \left\{ s_{n,c_n}(y) \geq 1 - a \right\}.$$

e. *Reject* $\mathbf{H_0}$ *at* a *iff* $\eta_n > q_{n,c_n}(1-a)$.

In order to derive the asymptotic properties below we finally employ the following standard assumption that restricts the asymptotic behaviour of $(c_n)_{n\in\mathbb{N}}$.

Assumption 5. $(c_n)_{n\in\mathbb{N}}$, *possibly depending on* $(\; Y_i, \quad X_i \;)_{i=1,\dots,n}$, satisfies

$$\mathbb{P}\left(l_n \leq c_n \leq u_n\right) \to 1$$

where (l_n) and (u_n) are real sequences such that $1 \leq l_n \leq u_n$ for all n, $l_n \to \infty$ and $\frac{u_n}{n} \to 0$ as $n \to \infty$.

The main result is the following.

Proposition 2. *Suppose that Assumptions 1 and 5 as well as the the provisions of Lemma 1 hold. For the testing procedure described in Algorithm 1 we have that*

a. *If* $\mathbf{H_0}$ *is true then*

$$\lim_{n\to\infty} \mathbb{P}\left(\eta_n > q_{n,b_n}(1-a)\right) = a.$$

b. *If* $\mathbf{H_{alt}}$ *is true then*

$$\lim_{n\to\infty} \mathbb{P}\left(\eta_n > q_{n,b_n}(1-a)\right) = 1.$$

c. *If* $\mathbf{H_{alt,n}}$ *is true then*

$$\lim_{n\to\infty} \mathbb{P}\left(\eta_n > q_{n,b_n}(1-a)\right) = \mathbb{P}\left(\frac{(\beta^\star + \zeta)^{Tr}\,\mathbb{V}^{-1}(\beta^\star + \zeta)}{\zeta_2} > q(1-a)\right),$$

where $q(1-a)$ *is the* $(1-a)$ *quantile of the distribution of* $\frac{\zeta^{Tr}\mathbb{V}^{-1}\zeta}{\zeta_2}$.

Proof. The first and third result follow by a direct application of Theorem 3.5.1.i,iii respectively of Politis et al. [12], and this is enabled by the results of Lemma 1. For the second result notice that if $\mathbf{H}_{\mathbf{alt}}$ is true then due to Lemmata 1, 6, $\eta_n = k_{1,n} + k_{2,n}^{Tr} \frac{n^{\frac{\alpha-1}{\alpha}}}{\kappa_n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}} (\beta_0 - \beta_\star) + \left(\frac{n^{\frac{\alpha-1}{\alpha}}}{\kappa_n^{\frac{1}{\alpha}} r_n^{\frac{1}{\alpha}}} \right)^2 \|\beta_0 - \beta_\star\|^2$ where $k_{1,n}, \|k_{2,n}\| = O_p(1)$ and thereby it diverges to $+\infty$. □

The results in the previous proposition imply that the usual subsampling modification of the Wald test remains robust under our current assumption framework. More specifically asymptotic exactness and consistency remain invariant even when the limiting distributions are stable with $\alpha > 1$, while the rates also involve sequences that are asymptotically equivalent to diverging truncated moments of the scalling processes involved in the relevant martingale transforms. Furthermore the test also remains locally asymptotically unbiased when the sequences of local alternatives are appropriately modified.

A possibly interesting extension would be the derivation of suchlike results even in cases where the limiting distributions are mixtures and $\alpha < 2$. This would require a non-trivial extension of the results applied in the proof of Theorem 3.1 of Hall and Yao [6] and is thereby left for future research.

References

[1] Arvanitis and Louka. (2015). Martingale Transforms with Mixed Stable Limits and the QMLE for Conditionally Heteroskedastic Models. mimeo.

[2] Bingham N.H., Goldie C.M. and J.L. Teugels. (1992). *Regular Variation*. Encyclopedia of Mathematics and Its Applications, Vol. 27, C.U.P.

[3] Cline, D. B., & Samorodnitsky, G. (1994). Subexponentiality of the product of independent random variables. *Stochastic Processes and their Applications*, 49(1), 75-98.

[4] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, (Vol. 2). New York: Wiley.

[5] Gupta, A. K., Nguyen, T. T., & Zeng, W. B. (1994). Conditions for stability of laws with all projections stable. *Sankhyā: The Indian Journal of Statistics*, Series A, 438-443.

[6] Hall P. and Yao Q. (2003). Inference in ARCH and GARCH Models with Heavy Tailed Errors. *Econometrica*, 71, 285-317.

[7] Jakubowski A. (2012). Principle of Conditioning Revisited. *Demonstratio Mathematica*, Vol. XLV, No 2.

[8] Ibragimov I.A. and Yu. V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*. WOLTERS-NOORDHOFF PUBLISHING GRONINGEN.

[9] Knight, K. (1993). Estimation in dynamic linear regression models with infinite variance errors. *Econometric Theory*, 9(04), 570-588.

[10] Kokoszka, P. S., & Taqqu, M. S. (1996). Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistics*, 24(5), 1880-1913.

[11] Mikosch, T., & Straumann, D. (2006). Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics*, 493-522.

[12] Politis, D. N., J. P. Romano and M. Wolf (1999). *Subsampling*. Springer New York.

[13] Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences. MIMEO. url: https://hal.archives-ouvertes.fr/cel-00867106/document.

[14] Sung, S. H. (1999). Weak law of large numbers for arrays of random variables. *Statistics & probability letters*, 42(3), 293-298.

[15] Surgailis, D. (2008). A Quadratic ARCH(∞) Model with Long Memory and Levy Stable Behavior of Squares. *Advances in Applied Probability*, 40:1198-1222, 2008.

Appendix-Auxiliary Lemmata

Lemma 2. *Under Assumption 1.b-c for any $\delta > 0$*

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n} v_i^{\alpha} - 1\right| \geq \delta\right) \to 0.$$

Proof. If κ_n converges the result follows from Birkhoff's LLN. If not, then, letting $v_i^{\star} \doteqdot v_i^{\alpha} 1_{\left\{v_i \leq (n\kappa_n)^{1/\alpha}\right\}}$

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^{n} (v_i^{\alpha} - \kappa_n)\right| \geq \delta\right)$$

$$\leq \frac{\mathbb{E}\left[(v_0^{\star})^2\right]}{n\kappa_n^2\delta^2} + \frac{2\sum_{i=1}^{n}(n-t+1)\,\mathbb{E}\left[v_0^{\star}v_i^{\star}\right]}{n^2\kappa_n^2\delta^2} + \mathbb{P}\left(\max_{i=1,\ldots,n} v_i > (n\kappa_n)^{1/\alpha}\right) + o(1),$$

where the last display follows from the inequality of Chebychev. Due to the covariance summability condition in Assumption 1.b the second term of the last display converges to zero. For the first term we have that

$$\frac{\mathbb{E}\left[(v_0^{\star})^2\right]}{n\kappa_n^2} = -\frac{n}{\kappa_n}\mathbb{P}\left(v_0^{\alpha} > n\kappa_n\right) + \frac{2}{\kappa_n^2 n}\int_0^{n\kappa_n} x\mathbb{P}\left(v_0^{\alpha} > x\right)dx$$

which converges to zero again due to 1.b. Likewise

$$\mathbb{P}\left(\max_{i=1,\ldots,n} v_i > (n\kappa_n)^{1/\alpha}\right) \leq n\mathbb{P}\left(v_0 > (n\kappa_n)^{1/\alpha}\right) \to 0.$$

□

Lemma 3. *Let $\lambda \in \mathbb{R}^q - \left\{0_{q\times 1}\right\}$. Under Assumption 1.c, for any $\delta > 0$ and any $0 \leq \alpha \leq 2$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}X_i\right|^{\alpha} - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^{\alpha}/\mathcal{I}\right]\right| \geq \delta\right) \to 0,$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}X_i\right|^{\alpha}\operatorname{sgn}\left(\lambda^{Tr}X_i\right) - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^{\alpha}\operatorname{sgn}\left(\lambda^{Tr}K_0\right)/\mathcal{I}\right]\right| \geq \delta\right) \to 0,$$

and

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_iX_i^{Tr} - \mathbb{E}\left[K_0K_0^{Tr}/\mathcal{I}\right]\right| \geq \delta\right) \to 0.$$

Proof. For the first result we have that for any $\delta > 0$ due to the Triangle Inequality

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left|\lambda^{Tr}K_i + \lambda^{Tr}A_i\right|^{\alpha} - \sum_{i=1}^{n}\left|\lambda^{Tr}K_i\right|^{\alpha}\right| \geq \delta\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}A_i\right|^{\alpha} \geq \delta\right)$$

and the latter converges to zero due to the Cezaro Mean Theorem. Then the result follows from Doob's LLN for stationary sequences due to Assumption 1.c. Analogously for the second result we have that

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left|\lambda^{Tr}K_i + \lambda^{Tr}A_i\right|^{\alpha}\operatorname{sgn}\left(\lambda^{Tr}X_i\right) - \sum_{i=1}^{n}\left|\lambda^{Tr}K_i\right|^{\alpha}\operatorname{sgn}\left(\lambda^{Tr}X_i\right)\right| \geq \delta\right)$$
$$\leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}A_i\right|^{\alpha} \geq \delta\right).$$

Furthermore

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left|\lambda^{Tr}K_i\right|^{\alpha}\left[\operatorname{sgn}\left(\lambda^{Tr}X_i\right) - \operatorname{sgn}\left(\lambda^{Tr}K_i\right)\right]\right| \geq \delta\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}K_i\right|^{\alpha}\left|\lambda^{Tr}A_i\right| \geq \delta\right),$$

since the sgn function is Lipschitz. From Assumption 1.c and Holder's inequality, the probability in the rhs is less than or equal to

$$\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}K_i\right|^{\alpha+\Delta}\right)^{\frac{\alpha}{\alpha+\Delta}}\left(\frac{1}{n}\sum_{i=1}^{n}\left|\lambda^{Tr}A_i\right|^{\frac{\alpha+\Delta}{\Delta}}\right)^{\frac{\Delta}{\alpha+\Delta}} \geq \delta\right)$$

and the result follows from Doob's LLN for stationary sequences and the Cezaro Mean Theorem due to Assumption 1.b-c. For the last result we see that due to sub-multiplicativity and Cauchy-Schwarz inequality,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|K_iA_i^{Tr}\right\| \geq \delta\right) \leq \mathbb{P}\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|K_i\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|A_i\right\|^2} \geq \frac{\delta}{\Delta}\right)$$

and the result follows again from Doob's LLN for stationary sequences and the Cezaro Mean Theorem due to Assumption 1.b-c. □

Lemma 4. *Let* $\lambda \in \mathbb{R}^q - \left\{0_{q\times 1}\right\}$*. Under Assumption 1.a-d, for any* $\delta > 0$

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i v_i\right|^\alpha - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^\alpha / \mathcal{I}\right]\right| \geq \delta\right) \to 0,$$

and

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i v_i\right|^\alpha \operatorname{sgn}\left(\lambda^{Tr}X_i\right) - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^\alpha \operatorname{sgn}\left(\lambda^{Tr}K_0\right) / \mathcal{I}\right]\right| \geq \delta\right) \to 0.$$

Proof. For arbitrary $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i v_i\right|^\alpha - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^\alpha / \mathcal{I}\right]\right| \geq \delta\right)$$

$$\leq \mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i\right|^\alpha v_i^\alpha - \frac{1}{n}\sum_{i=1}^n \left|\lambda^{Tr}X_i\right|^\alpha\right| \geq \frac{\delta}{2}\right)$$

$$+\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \left|\lambda^{Tr}X_i\right|^\alpha - \mathbb{E}\left[\left|\lambda^{Tr}K_0\right|^\alpha / \mathcal{I}\right]\right| \geq \frac{\delta}{2}\right).$$

The second term in the rhs converges to zero due to Lemma 3. For the first one notice that

$$\mathbb{P}\left(\left|\frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i\right|^\alpha v_i^\alpha - \frac{1}{n\kappa_n}\sum_{i=1}^n \left|\lambda^{Tr}X_i\right|^\alpha v_i^\star\right| \geq \delta\right)$$

$$\leq n\mathbb{P}\left(\left|v_0\right| > \left(n\kappa_n\right)^{1/\alpha}\right) = o\left(1\right),$$

due to Assumption 1.a-c. We also have that due to 1.b-d and Cezaro's Theorem

$$\sup_{n\geq 1}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left|\lambda^{Tr}X_i\right|^\alpha \frac{\mathbb{E}\left(v_i^\star\right)}{\kappa_n} \leq \mathbb{E}\left(\left|\lambda^{Tr}K_0\right|^\alpha\right) + \sup_{n\geq 1}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\left|\lambda^{Tr}A_i\right|^\alpha\right) \leq +\infty.$$

Furthermore, for any $\epsilon > 0$

$$\frac{1}{n\kappa_n}\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left(\left(\left|\lambda^{Tr}X_i\right|^\alpha v_i^\star 1_{\{|\lambda^{Tr}X_i|^\alpha v_i^\star > \epsilon n\kappa_n\}}\right) / \sigma\left(X_i\right)\right)\right]$$

$$= \frac{1}{n\kappa_n}\sum_{i=1}^n \mathbb{E}\left[\left|\lambda^{Tr}X_i\right|^\alpha 1_{\{|\lambda^{Tr}X_i|^\alpha > \epsilon\}}\mathbb{E}\left(v_i^\alpha 1_{\left\{\frac{(\epsilon n\kappa_n)^{1/\alpha}}{|\lambda^{Tr}X_i|} < v_i \leq (n\kappa_n)^{1/\alpha}\right\}}\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\left|\lambda^{Tr}X_i\right|^\alpha 1_{\{|\lambda^{Tr}X_i|^\alpha > \epsilon\}}\frac{h^\star\left(\left(n\kappa_n\right)^{1/\alpha}\right) - h^\star\left(\frac{(\epsilon n\kappa_n)^{1/\alpha}}{|\lambda^{Tr}X_i|}\right)}{\kappa_n}\right]$$

and the latter converges to zero due to dominated convergence, Assumption 3 and the fact that $h^\star$ is slowly varying at infinity. Hence there exists some $\epsilon_n \to 0$ for which

$$\lim_{n\to\infty} \frac{1}{n\kappa_n} \sum_{i=1}^{n} \mathbb{E}\left[\left|\lambda^{Tr} X_i\right|^{\alpha} v_i^\star 1_{\{|\lambda^{Tr} X_i|^{\alpha} v_i^\star > \epsilon_n n \kappa_n\}}\right] = 0$$

and thereby, due to Theorem 1 of Sung [14]

$$\frac{1}{n} \sum_{i=1}^{n} \left(\left|\lambda^{Tr} X_i\right|^{\alpha} \frac{v_i^\star - \mathbb{E}\left(v_i^\star\right)}{\kappa_n} \right) = o_p\left(1\right).$$

The result then follows from Lemma 2. A similar consideration provides the second result. □

Lemma 5. *Under Assumption 1.a-d, for any $M > 0$*

$$\mathbb{P}\left(\max_{i\le n} \left|\lambda^{Tr} X_i v_i\right| > M \left(n r_n \kappa_n\right)^{1/\alpha} \right) \to 0.$$

Proof. See Cline and Samorodnitsky [3]. □

Lemma 6. *Under Assumption 1, for any $\delta > 0$,*

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left[\left(\xi_i v_i\right)^2 - \left(\epsilon_i\right)^2 \right] \right| \geq \delta \right) \to 0$$

where $\epsilon_i \doteqdot y_i - X_i^{Tr} \beta_n$.

Proof. For arbitrary $\delta > 0$ and $\beta_{0,n}$ as in Remark 1.6, due to submultiplicativity

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left[\left(\xi_i v_i\right)^2 - \left(y_i - X_i^{Tr} \beta_n\right)^2 \right] \right| \geq \delta \right)$$

$$\leq \mathbb{P}\left(\left| \left\|\beta_n - \beta_{0,n}\right\|^2 \frac{1}{n} \sum_{i=1}^{n} \left\|X_i\right\|^2 \right| \geq \frac{\delta}{2} \right) + \mathbb{P}\left(\left| \frac{\left\|\beta_n - \beta_{0,n}\right\|}{n} \sum_{i=1}^{n} \left\|X_i^{Tr} v_i \xi_i\right\| \right| \geq \frac{\delta}{4} \right),$$

and the latter probability converges to zero due to Proposition 1 and Lemma 3. □