

## TAIL INDEX ESTIMATION FOR A FILTERED DEPENDENT TIME SERIES

Jonathan B. Hill

*University of North Carolina at Chapel Hill*

*Abstract:* We prove Hill's (1975) tail index estimator is asymptotically normal when the employed data are generated by a stationary parametric time series  $\{x_t(\theta^0) : t \in \mathbb{Z}\}$  and  $\theta^0$  is an unknown  $k \times 1$  vector. We assume  $x_t(\theta^0)$  is unobservable but  $\theta^0$  is estimable with estimator  $\hat{\theta}_n$  and sample size  $n \geq 1$ , and that the filtered series  $x_t(\hat{\theta}_n)$  is observed and used to estimate the tail index. Natural applications include regression residuals, GARCH filters, and weighted sums based on an optimization problem like optimal portfolio selection. Our main result substantially extends Resnick and Stărică's (1997) theory for estimated AR i.i.d. errors and Ling and Peng's (2004) theory for estimated ARMA i.i.d. errors to a wide range of filtered time series since we do not require  $x_t(\theta^0)$  to be i.i.d., nor generated by a linear process with geometric dependence. We assume  $x_t(\theta^0)$  is  $\beta$ -mixing with possibly hyperbolic dependence, covering ARMA-GARCH filters, ARMA filters with heteroscedastic errors of unknown form, nonlinear filters like threshold autoregressions, and filters based on mis-specified models, as well as i.i.d. errors in an ARMA model. Finally, as opposed to Resnick and Stărică (1997) and Ling and Peng (2004) we do not require  $\hat{\theta}_n$  to be super- $\sqrt{n}$ -convergent when  $x_t(\theta^0)$  has an infinite variance. We allow a far greater variety of plug-ins, including those that are slower than  $\sqrt{n}$ , such as QML-type estimators for GARCH models.

*Key words and phrases:* GARCH filter, regression residuals, tail index estimation, weak dependence.

### 1. Introduction

In this paper we establish the asymptotic normality of Hill's (1975) seminal tail index estimator for a stationary ergodic, filtered process  $\{x_t(\theta^0) : -\infty < t < \infty\}$ . We assume  $x_t(\theta)$  maps

$$x_t : \Theta \rightarrow \mathbb{R} \quad \text{and } \Theta \text{ is a compact subset of } \mathbb{R}^k \text{ for finite } k \geq 1,$$

and  $x_t(\theta)$  is thrice continuously differentiable with a continuous distribution for each  $\theta$ .

We assume  $\theta^0$  is an unknown unique point in  $\Theta$ , and that  $x_t(\theta^0)$  is not observed, but that  $\theta^0$  is estimable and  $x_t(\hat{\theta}_n)$  is observable, where  $\hat{\theta}_n$  is a plug-in estimator of  $\theta^0$  and  $n \geq 1$  is the sample size. We therefore use  $x_t(\hat{\theta}_n)$  to estimate the tail index of  $x_t := x_t(\theta^0)$ .

If  $x_t$  is observed then the reader can refer to Hsing (1991) and Hill (2010, 2011a) for what appears to be the most general limit theory for Hill's (1975) estimator in terms of allowed dependence and non-stationarity. See also Ilić (2012) for an extension to the missing data case. We assume  $x_t$  is not observed and  $\theta^0$  must be estimated.

The dominant example is the use of regression model residuals for tail index estimation, including use of a GARCH model to control for conditional heteroskedasticity. This offers an advantage for stationary ARMA processes since the observed series and the i.i.d. errors  $x_t$  have the same tail index, and the Hill (1975) estimator is more efficient if residuals are used for computation provided they are based on an estimator that is super- $n^{1/2}$ -convergent when  $E[x_t^2] = \infty$  (Resnick and Stărică (1997)).

It is standard practice in macroeconomics and finance to use pre-filtering to control for heteroskedasticity, while GARCH-type feedback implies heavy tails (e.g., Basrak, Davis, and Mikosch (2002), Liu (2006), Cline (2007)). GARCH filters are particularly relevant for tests of volatility spillover in financial markets, and knowing whether the GARCH error is heavy tailed has major repercussions on the test approach and estimator used (for theory and references see Hill and Aguilar (2013) and Aguilar and Hill (2014)). As another example, GARCH filters require a parameter estimator and if the GARCH error has an infinite fourth moment then QML has a slow convergence rate (Hall and Yao (2003)), while robust non-Gaussian QML estimators in Berkes and Horvath (2004) and Zhu and Ling (2011) and robust Gaussian QML in Hill (2014a) converge faster and may have better small sample properties. Further, in finance the return  $x_t$  on an optimal portfolio involves a weighted sum of asset returns  $\sum_{i=1}^k \theta_i^0 y_{i,t}$  with unknown but estimable weights  $\theta^0$ , and  $y_{i,t}$  is the  $i^{\text{th}}$  asset return. Typically returns  $y_{i,t}$  are dependent, conditionally heteroscedastic, asymmetrically distributed or generated by a nonlinear process, and heavy tailed (cf., Embrechts, Klüppelberg, and Mikosch (1997)).

We assume  $x_t$  has support  $[0, \infty)$  and has, for each  $t$ , a common regularly varying distribution tail with tail index  $\kappa > 0$ :

$$P(x_t > a) = a^{-\kappa} \mathcal{L}(a), \quad \text{where } a > 0 \text{ and } \mathcal{L}(a) \text{ is slowly varying.} \quad (1.1)$$

If  $\{y_t\}$  is the process of interest then  $x_t$  simply represents a tail-specific version:  $|y_t|$ ,  $-y_t I(y_t < 0)$  or  $y_t I(y_t > 0)$ . See Resnick (1987) for a compendium treatment of regular variation. Let  $x_{(i)}$  be the order statistics of the sample  $\{x_t\}_{t=1}^n$ :  $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)}$ , and let  $\{m_n : n \in \mathbb{N}\}$  be an intermediate order sequence:  $1 \leq m_n < n$ ,  $m_n \rightarrow \infty$  as  $n \rightarrow \infty$  and  $m_n/n \rightarrow 0$ . Hill's (1975) estimator is the inverse of the mean log peak-over-threshold

$$\hat{\kappa}_{m_n}(\theta) = \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \ln \left( \frac{x_{(i)}(\theta)}{x_{(m_n+1)}(\theta)} \right) \right)^{-1} \quad \text{and write } \hat{\kappa}_{m_n} := \hat{\kappa}_{m_n}(\theta^0).$$

Thus,  $\hat{\kappa}_{m_n}(\theta)$  may estimate the left or right index of some  $y_t$ , or the tail index of  $|y_t|$ .

We prove

$$m_n^{1/2} \frac{(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \kappa^{-1})}{\sigma_{m_n}} \xrightarrow{d} N(0, 1),$$

with the mean-squared-error [mse]  $\sigma_{m_n}^2 := E(m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \kappa^{-1}))^2$ , as long as  $x_t$  is absolutely regular (i.e.  $\beta$ -mixing) with summable coefficients, and the slowly varying tail component  $\mathcal{L}(a)$  satisfies a second order property. See Theorem 1 in Section 2. We also require the plug-in  $\hat{\theta}_n$  to converge to  $\theta^0$  sufficiently fast. This is intuitive since  $m_n^{1/2}$  is the rate of convergence of  $\hat{\kappa}_{m_n}^{-1}$  under general conditions of memory and heterogeneity (Hill (2010)): if  $\hat{\theta}_n \xrightarrow{p} \theta^0$  slightly faster than  $m_n^{1/2}$  then  $\hat{\theta}_n$  will not affect  $\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n)$  asymptotically. In particular we require  $m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) \xrightarrow{p} 0$ . The latter is easy to satisfy since the number of tail observations  $m_n$  is chosen by the analyst, while setting  $m_n^{1/2} \ln(n) = o(n^{1/2})$  implies sub- $n^{1/2}$ -convergent estimators  $\hat{\theta}_n$  are allowed. This is important since some QML-type estimators for GARCH are sub- $n^{1/2}$ -convergent if the error has an infinite fourth moment (Hall and Yao (2003), Hill (2014a)); and nonparametric estimators based on kernel smoothing, GMM estimators with nearly weak instruments, and estimators with in-fill asymptotics may be sub- $n^{1/2}$ -convergent in general (see Antoine and Renault (2012)).

The mse  $\sigma_{m_n}^2$  of  $\hat{\kappa}_{m_n}^{-1}$  is the proper scale for  $\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n)$ : by assuming  $\hat{\theta}_n \xrightarrow{p} \theta^0$  faster than  $m_n^{1/2}$  it is as though  $\theta^0$  were known. The same result in the i.i.d. case for linear filters is derived in Resnick and Stărică (1997) and Ling and Peng (2004) since they assume  $\hat{\theta}_n$  is super- $n^{1/2}$ -convergent and, rather trivially,  $m_n$  must be  $o(n)$ . Consult Hill (2010) for a non-parametric estimator of  $\sigma_{m_n}^2$  for a broad class of dependent processes that covers the present environment. If  $x_t$  is i.i.d., or is stochastic volatility, or in general exhibits a sufficient degree of serial extremal orthogonality, then  $\sigma_{m_n}^2 \rightarrow \kappa^{-2}$  (Hall (1982), Hill (2011b), Hill and Shneyerov (2013)). Nevertheless, a non-parametric estimator of  $\sigma_{m_n}^2$  may lead to better small sample inference since the filtered series  $\{x_t(\hat{\theta}_n)\}_{t=1}^n$  may be serially dependent even when  $x_t$  is independent (cf., Hill and Shneyerov (2013)). See the simulation study in Section 4. The omitted case where  $\hat{\theta}_n \xrightarrow{p} \theta^0$  *not faster* than  $m_n^{1/2}$  is certainly possible but requires handling how  $\hat{\theta}_n$  is computed and therefore impacts  $\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n)$  asymptotically.

In terms of stationarity and dependence, greater generality is certainly possible. This includes short range dependent mixing and mixingale-like properties, like near epoch dependence assigned only to extreme values (e.g., Hsing (1991), Hill (2009, 2010, 2011a)), and long range dependence for linear processes with

i.i.d. errors (Beran, Das, and Schnell (2012)) or stochastic volatility (Hill (2011b), Kulik and Soulier (2011)). In order to assess how  $\hat{\theta}_n$  impacts  $\hat{\kappa}_{m_n}(\hat{\theta}_n)$ , however, we exploit a first order expansion around  $\theta^0$  that requires a Gaussian uniform central limit theorem for a tail empirical process on a compact neighborhood of  $\theta^0$ . While the setting here is not the most general, we aim for compactness by using an elegant result due to Doukhan, Massart, and Rio (1995) for a stationary weakly dependent  $\beta$ -mixing process, the conditions of which are easily satisfied if  $x_t(\theta)$  is continuous with a continuous bounded distribution. A more general pointwise limit theory for  $\alpha$ -mixing processes without reference to mixing coefficient decay can be found in Peligrad (1996), for example. This, however, requires an additional latticed correlation property that is difficult to verify in practice (cf., Bradley (1993)). A limit theory for a simple tail empirical process of a long range dependent stochastic volatility process is given in Kulik and Soulier (2011). They show by application that Hill's (1975) estimator is still asymptotically normal. In order to extend that result here, however, we would need a more general weak limit theory in view of the mapping  $x_t : \Theta \rightarrow \mathbb{R}$  that occurs in the tail empirical process.

Although tail exponents are frequently computed from regression model residuals, there are few results in the literature to justify the presumed asymptotic properties that typically ignore plug-in sampling error. Resnick and Stărică (1997), however, develop a Hill-estimator theory for estimated AR errors and Ling and Peng (2004) extend their results to an ARMA filter under fewer assumptions. In both cases the true unobserved errors are assumed i.i.d., and the plug-in is assumed to be super- $n^{1/2}$ -convergent when  $E[x_t^2] = \infty$ , e.g., OLS and LAD (Davis, Knight, and Liu (1992)). Each result limits the type of filter, dependence, and plug-in allowed, and each presumes we know the true model. For example, we cannot use a GARCH filter estimated by QML or QML-type estimators in Hill (2014a), Berkes and Horvath (2004), and Zhu and Ling (2011) since these estimators are not super- $n^{1/2}$ -convergent in any case. We also cannot use a mis-specified model, for example an AR model, when the true model is ARMA with an i.i.d. error, since the regression error is then dependent. Similarly we cannot use an ARMA filter if the error is non-i.i.d. as with an heteroscedastic error of unknown form, or a GARCH(1,1) error that is intrinsically heavy tailed (Mikosch and Stărică (2000)). The theory presented here allows for dependent data, nonlinear filters, regression models with non-i.i.d. errors, mis-specified models, and sub- $n^{1/2}$ -convergent plug-in estimators.

Hill-estimator asymptotics are grounded on a tail empirical process that arises from the use of an intermediate order statistic  $x_{(m_n+1)}(\theta)$ , cf., Hsing (1991). An asymptotic theory for any tail index estimator based on such order statistics follows similarly. Examples include Pickands (1975) and Dekkers, Einmahl, and de Haan (1989), but there are many more.

In Section 2 we present the main result, and examples follow in Section 3. Section 4 contains a simulation study. Proofs are in the Supplementary Material, Hill (2014c).

Throughout,  $[z]$  denotes the integer part of  $z \in \mathbb{R}$ .

**2. Estimation**

We state all assumptions and then present the main result. Drop  $\theta^0$  everywhere, e.g.,  $x_t = x_t(\theta^0)$  and  $\hat{\kappa}_{m_n} = \hat{\kappa}_{m_n}(\theta^0)$ . The first and second derivatives of  $x_t(\theta)$  are

$$g_t(\theta) = [g_{i,t}(\theta)] := \frac{\partial}{\partial \theta} x_t(\theta) \in \mathbb{R}^k \quad \text{and} \quad h_t(\theta) = [h_{i,j,t}(\theta)] := \frac{\partial^2}{\partial \theta \partial \theta'} x_t(\theta) \in \mathbb{R}^{k \times k}.$$

Derivatives at a point are written  $(\partial/\partial \theta)x_t(\tilde{\theta}) = (\partial/\partial \theta)x_t(\theta)|_{\tilde{\theta}}$  and  $(\partial/\partial \theta)x_t = (\partial/\partial \theta)x_t(\theta)|_{\theta^0}$ . Let  $\|\cdot\|$  denote the Euclidean norm of a vector or matrix, and  $\|\cdot\|_2$  the  $L_2$ -norm.

**2.1. Assumptions and main result**

**Assumption 1** (Smoothness and Moments).

- a. Let  $\{\mathfrak{S}_t\}_{t \in \mathbb{Z}}$  be a sequence of  $\sigma$ -fields that do not depend on  $\theta$  and define  $\mathcal{F} := \sigma(\cup_{t \in \mathbb{Z}} \mathfrak{S}_t)$ .  $x_t(\theta)$  lies on a complete probability measure space  $(\Omega, \mathcal{F}, P)$  and is  $\mathfrak{S}_t$ -measurable. Borel functions of  $x_t(\theta)$  satisfy Pollard (1984, Appendix C)'s permissibility criteria.
- b.  $x_t(\theta)$  is stationary, ergodic, and thrice continuously differentiable with  $\mathfrak{S}_t$ -measurable stationary and ergodic derivatives  $g_t(\theta)$  and  $h_t(\theta)$ .
- c. Each  $w_t(\theta) \in \{x_t(\theta), g_{i,t}(\theta), h_{i,j,t}(\theta)\}$  is governed by a non-degenerate distribution that is absolutely continuous with respect to Lebesgue measure, with uniformly bounded derivatives:  $\sup_{\theta \in \Theta} \sup_{a \in \mathbb{R}} \|(\partial/\partial \theta)P(w_t(\theta) \leq a)\| < \infty$  and  $\sup_{\theta \in \Theta} \sup_{a \in \mathbb{R}} \{(\partial/\partial a)P(w_t(\theta) \leq a)\} < \infty$ . Further  $E[\sup_{\theta \in \Theta} |w_t(\theta)|^\iota] < \infty$  for some tiny  $\iota > 0$ .
- d.  $\inf_{\theta \in \Theta} x_t(\theta) \geq \delta$  a.s. for some  $\delta > 0$ .

**Remark 1.** Assumption 1.a ensures probabilities and expectations of majorants of functions of  $x_t(\theta)$  are well defined, in which case it is understood that we use outer probabilities and expectations. Cf., Dudley (1978).

**Remark 2.** Differentiability permits standard expansions around  $\theta^0$ . The bound  $\inf_{\theta \in \Theta} x_t(\theta) \geq \delta > 0$  a.s. ensures  $g_{i,t}(\theta)/x_t(\theta)$  is uniformly  $L_\iota$ -bounded for tiny  $\iota > 0$ , and is trivially satisfied for tail index estimation simply by letting  $x_t(\theta)$  denote  $x_t(\theta) + \delta$  for any chosen  $\delta > 0$ , if necessary. The  $L_\iota$ -boundedness of  $g_{i,t}(\theta)/x_t(\theta)$  is useful in a first order expansion of  $1/m_n \sum_{i=1}^{m_n} \ln(x_{(i)}(\hat{\theta}_n))$  around  $\theta^0$ .

We assume  $x_t(\theta)$ , for all  $\theta$  in some neighborhood of  $\theta^0$ , has a distribution tail that satisfies first and second order regular variation properties. The need to impose higher order regular variation is treated in depth in Haeusler and Teugels (1985), Goldie and Smith (1987), and de Haan and Stadtmuller (1996). The second order property we use is *slow variation with remainder*, in particular the condition (SR1) in Goldie and Smith (1987). See also Hsing (1991) and Hill (2010). Otherwise we do not restrict the tails of  $x_t(\theta)$  outside a neighborhood of  $\theta^0$ .

Let the sequence  $\{c_n(\theta)\}$  of positive real mappings  $c_n : \Theta \rightarrow (0, \infty)$  satisfy the following for any intermediate order sequence  $\{m_n\}$ :

$$P(x_t(\theta) \geq c_n(\theta)) = \frac{m_n}{n}. \quad (2.1)$$

The existence of  $c_n(\theta)$  for any  $\theta$  is assured by distribution smoothness. Let  $\mathcal{N}_0(\delta)$  be a  $\delta$ -neighborhood of  $\theta$ :  $\mathcal{N}_0(\delta) = \{\theta \in \Theta : \|\theta - \theta^0\| \leq \delta\}$  for  $\delta > 0$ .

**Assumption 2** (Regular Variation and Fractile Bound).

a. *There exists a neighborhood  $\mathcal{N}_0(\delta)$  such that*

$$\lim_{a \rightarrow \infty} \sup_{\theta \in \mathcal{N}_0(\delta)} \left| \frac{a^{\kappa(\theta)}}{\mathcal{L}(a, \theta)} P(x_t(\theta) > a) - 1 \right| = 0. \quad (2.2)$$

*Note  $\mathcal{L}(a, \theta^0) = \mathcal{L}(a)$  in (1.1). The tail component  $\mathcal{L}(a, \theta)$  is slowly varying with remainder in  $a$ , uniformly on  $\Theta$ , that is  $\sup_{\theta \in \mathcal{N}_0(\delta)} |\mathcal{L}(\lambda a, \theta)/\mathcal{L}(a, \theta) - 1| = O(h(a))$  as  $a \rightarrow \infty$  for any  $\lambda > 0$  where  $h$  is a measurable function on  $(0, \infty)$  with bounded increase: there exist  $0 < D, z_0 < \infty$ , and  $\tau \leq 0$  such that  $h(\vartheta z)/h(z) \leq D\vartheta^\tau$  some for  $\vartheta \geq 1$  and  $z \geq z_0$  (Goldie and Smith (1987)). Further,  $m_n^{1/2} h(c_n) \rightarrow 0$ . Moreover, the tail index  $\kappa(\theta)$  is locally bounded,  $\inf_{\theta \in \mathcal{N}_0(\delta)} \kappa(\theta) > 0$  and  $\sup_{\theta \in \mathcal{N}_0(\delta)} \kappa(\theta) < \infty$ , and is twice differentiable with locally bounded derivatives and a Lipschitz first derivative:  $\|(\partial/\partial\theta)\kappa(\theta)\| < \infty$ ,  $\|(\partial/\partial\theta)^2\kappa(\theta)\| < \infty$ , and  $\|(\partial/\partial\theta)\kappa(\theta) - (\partial/\partial\theta)\kappa(\tilde{\theta})\| \leq K\|\theta - \tilde{\theta}\|$  for each  $\theta, \tilde{\theta} \in \mathcal{N}_0(\delta)$ .*

b.  $m_n \rightarrow \infty$  and  $m_n = o(n/\ln(n))$ .

**Remark 3.** Property (a) says any  $x_t(\theta)$  on some neighborhood  $\mathcal{N}_0(\delta)$  of  $\theta^0$  has a regularly varying tail. We use the property to derive a uniform limit for the intermediate order statistic  $x_{(m_n+1)}(\theta)$ . The property is satisfied by various classes of Markov chains, including ARMA with heavy tailed errors, and stochastic recurrence equations like GARCH and Random Coefficient Autoregressions, when the errors satisfy a second order Paretian expansion. See Section 3 for examples.

**Remark 4.** The property  $m_n^{1/2}h(c_n) \rightarrow 0$  implicitly limits how big  $m_n$  can be, ensuring that observations are taken from far enough out in the tails. If  $x_t(\theta)$  is exactly Pareto distributed for any  $\theta$  then  $L(a, \theta)$  depends only on  $\theta$  hence  $h(a) = 0$ . Thus, for a Pareto law there are no restrictions on how many observations are valid, other than  $m_n = o(n)$ , cf., Haeusler and Teugels (1985) and Goldie and Smith (1987). We impose  $m_n = o(n/\ln(n))$  in (b), even in the exact Pareto case, to simplify working with the derivative of  $x_t(\theta)$  and therefore an expansions of  $\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n)$  around  $\theta^0$ . In many cases once  $m_n^{1/2}h(c_n) \rightarrow 0$  is assured then (b) automatically holds. See Section 3 for an example.

**Assumption 3** (mixing). Define  $\mathcal{G}_s^t(\delta) := \sigma(\cup_{\theta \in \mathcal{N}_0(\delta)} \sigma(x_\tau(\theta) : s \leq \tau \leq t))$ .  $x_t(\theta)$  is  $\beta$ -mixing on  $\mathcal{N}_0(\delta)$  with summable coefficients:  $\beta_l := \sup_{\mathcal{A} \subset \mathcal{G}_{t+l}^\infty(\delta)} E|P(\mathcal{A}|\mathcal{G}_{-\infty}^t(\delta)) - P(\mathcal{A})|$ , where  $\sum_{l=1}^\infty \beta_l < \infty$ .

**Remark 5.** Allowed dependence decay is at least geometric  $\beta_l = O(\rho^l)$  where  $\rho \in (0, 1)$ , and hyperbolic, e.g.,  $\beta_l = O(l^{-1}/\ln(l))$ . Similar to Assumption 2, we only need to consider dependence on a neighborhood of  $\theta_0$ . This is key, however, since some processes  $\{x_t(\theta)\}$  can only be shown to be mixing on a small neighborhood of  $\theta_0$ , including GARCH processes. See Section 3.

**Remark 6.** Our limit theory hinges on the tail empirical process  $\{\mathcal{I}_{n,t}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$  based on the scaled indicator  $I(|x_t(\theta)| \leq c_n(\theta))$  defined in (2.3) below. We show in Lemma 1 that the summability  $\sum_{l=1}^\infty \beta_l < \infty$  ensures  $\{\mathcal{I}_{n,t}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$  satisfies a Gaussian uniform central limit theorem, while a Gaussian limit can be used to argue that  $\sum_{l=1}^\infty \beta_l < \infty$  implies short-range dependence in  $x_t(\theta)$  on  $\mathcal{N}_0(\delta)$ , an argument dating to Rosenblatt (1956). Interestingly, Kulik and Soulier (2011) show if in  $I(|x_t| \leq c_n)$  we replace  $c_n$  with  $x_{(m_n+1)}$  as we do in the Hill (1975) estimator, then a Gaussian limit theory exists for  $I(|x_t| \leq x_{(m_n+1)})$  irrespective of long memory. That result, however, does not cover the uniform central limit theorem for  $\mathcal{I}_{n,t}(\theta)$  that we require.

**Assumption 4** (Plug-In). There exists a unique point  $\theta^0 \in \Theta$  such that  $m_n^{1/2} \ln(n)(\hat{\theta}_n - \theta^0) = o_p(1)$ .

**Theorem 1.** Under Assumptions 1 – 4,  $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \kappa^{-1})/\sigma_{m_n} \xrightarrow{d} N(0, 1)$ , where  $\sigma_{m_n}^2 := E(m_n^{1/2}(\hat{\kappa}_{m_n}^{-1} - \kappa^{-1}))^2$ .

**Remark 7.** The proper scale is  $\sigma_{m_n}^2$  as if the true value  $\theta^0$  were used instead of the plug-in  $\hat{\theta}_n$ . Therefore, if  $x_t$  is i.i.d. then  $m_n^{1/2}(\hat{\kappa}_{m_n}^{-1}(\hat{\theta}_n) - \kappa^{-1}) \xrightarrow{d} N(0, \kappa^{-2})$ , cf., Hsing (1991, eq. (2.8)) and Hill (2010, Thm. 2). See Hall (1982, Thm. 2) for a seminal proof for i.i.d.  $x_t$  with a second order Paretian tail  $P(x_t > a) = da^{-\kappa}(1 + O(a^{-\beta}))$ , where  $\beta, d, \kappa > 0$ .

**Remark 8.** Hill (2010) presents a nonparametric estimator  $\hat{\sigma}_{m_n}^2$  of  $\sigma_{m_n}^2$  using non-filtered data, e.g.,  $x_t$ . Under regular variation (1.1), a general weak dependence property that covers mixing Assumption 3, and regularity conditions on  $m_n$ , the estimator is consistent:  $\hat{\sigma}_{m_n}^2/\sigma_{m_n}^2 \xrightarrow{P} 1$ . Although we do not provide a proof here, it is readily shown that  $\hat{\sigma}_{m_n}^2/\sigma_{m_n}^2 \xrightarrow{P} 1$  applies using filtered data  $x_t(\hat{\theta}_n)$ , provided Assumptions 1-4 hold. Indeed, since only consistency for  $\hat{\sigma}_{m_n}^2$  is at stake, the plug-in need only be consistent  $\hat{\theta}_n = \theta^0 + o_p(1)$ .

The proof of Theorem 1 is contained in Hill (2014c), and exploits weak limit theory for tail and non-tail arrays. In particular, we require a result concerning trimming indicators and order statistics, that may be of independent interest. Take

$$\mathcal{I}_{n,t}(\theta) := \left(\frac{n}{m_n}\right)^{1/2} \{I(|x_t(\theta)| \leq c_n(\theta)) - E[I(|x_t(\theta)| \leq c_n(\theta))]\}, \quad (2.3)$$

where  $I(\cdot)$  is an indicator function:  $I(A) = 1$  (or 0) if  $A$  is true (or false). By construction and (2.1):  $E[\mathcal{I}_{n,t}(\theta)] = 0$  and  $E[\mathcal{I}_{n,t}^2(\theta)] = (n/m_n)P(|x_t(\theta)| > c_n(\theta)) \times P(|x_t(\theta)| \leq c_n(\theta)) = P(|x_t(\theta)| \leq c_n(\theta)) \rightarrow 1$ .

In the following we exploit the concept of weak convergence on a Polish space, denoted  $\implies^*$ , cf., Hoffmann-Jørgensen (1991). A Polish space is a separable and completely metrizable space. This generality helps sidestep difficult measurability issues that arise when proving weak convergence of functions (cf., Dudley (1978), Hoffmann-Jørgensen (1991)). In particular, let  $\{\mathcal{Z}_n(\theta) \in \mathcal{T}\}$  be a stochastic process on compact  $\mathcal{T} \subseteq \Theta$ , and  $\{\mathcal{Z}(\theta) \in \mathcal{T}\}$  be a Gaussian process with uniformly bounded and uniformly continuous sample paths with respect to  $\|\cdot\|_2$ . Let  $P^*$  denote outer probability. Then  $\{\mathcal{Z}_n(\theta) \in \mathcal{T}\} \implies^* \{\mathcal{Z}(\theta) \in \mathcal{T}\}$  if  $\mathcal{Z}_n(\theta)$  converges in finite dimensional distributions, and  $\{\mathcal{Z}_n(\theta) \in \mathcal{T}\}$  is tight in the sense that  $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P^*(\sup_{\|\theta - \tilde{\theta}\| \leq \delta} |\mathcal{Z}_n(\theta) - \mathcal{Z}_n(\tilde{\theta})| > \varepsilon) = 0 \forall \varepsilon > 0$ . This approach for proving uniform central limit theorems dates at least to Dudley (1978).

See the Supplementary Material, Hill (2014c), for a proof.

**Lemma 1.** *Under Assumptions 1 – 3 there exists a Gaussian process  $\{\mathcal{I}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$  with uniformly bounded and uniformly continuous sample paths with respect to  $\|\cdot\|_2$  such that*

- a.  $\{n^{-1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta) : \theta \in \mathcal{N}_0(\delta)\} \implies^* \{\mathcal{I}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$ ,
- b.  $\sup_{\theta \in \mathcal{N}_0(\delta)} |m_n^{1/2} \ln(x_{(m_n+1)}(\theta)/c_n(\theta)) - \kappa^{-1} n^{-1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta)| \xrightarrow{P} 0$ ,
- c.  $\{m_n^{1/2} \ln(x_{(m_n+1)}(\theta)/c_n(\theta)) : \theta \in \mathcal{N}_0(\delta)\} \implies^* \{\kappa^{-1} \mathcal{I}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$ .

**Remark 9.** Apply the Mean Value and Continuous Mapping theorems to deduce a uniform probability bound for the intermediate order statistic:  $\sup_{\theta \in \mathcal{N}_0(\delta)} |x_{(m_n+1)}(\theta)/c_n(\theta) - 1| = O_p(1/m_n^{1/2})$ .

**Remark 10.** By construction,  $\{1/n^{1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta) : \theta \in \mathcal{N}_0(\delta)\}$  is a type of tail empirical process. The standard construction is  $\{1/n^{1/2} \sum_{t=1}^n \mathcal{J}_{n,t}(u) : u \geq 0\}$ , where

$$\mathcal{J}_{n,t}(u) := \left(\frac{n}{m_n}\right)^{1/2} \{I(|x_t| > c_n + u\vartheta_n) - E[I(|x_t| > c_n + u\vartheta_n)]\},$$

and  $\{\vartheta_n\}$  are positive norming constants (Rootzén (2009, p.469)). Since trivially

$$\mathcal{I}_{n,t}(\theta) = -\left(\frac{n}{m_n}\right)^{1/2} \{I(|x_t(\theta)| > c_n(\theta)) - E[I(|x_t(\theta)| > c_n(\theta))]\},$$

a tail empirical process includes a non-tail process (see Hill (2011a)). Lemma 1 extends Theorem 2.1 in Rootzén (2009) for  $\{1/n^{1/2} \sum_{t=1}^n \mathcal{J}_{n,t}(u) : u \in \mathcal{R}\}$  and compact  $\mathcal{R} \subset [0, \infty)$  to a larger function class, since  $|x_t(\theta)| > c_n(\theta)$  for  $\theta \in \Theta$  generalizes  $|x_t| > c_n + u\vartheta_n$  for  $u \in \mathcal{R}$ . Rootzén (2009), however, only imposes  $\beta$ -mixing on tail information  $x_t I(|x_t| > c_n)$  as  $n \rightarrow \infty$ , but by trivial added steps the same generality applies here. See also Hill (2009, 2010, 2011a). Similarly, by the proof of Lemma 1.a, we can easily derive weak convergence for  $\{1/n^{1/2} \sum_{t=1}^n \mathcal{I}_{n,t}(\theta, u) : \theta, u \in \mathcal{N}_0(\delta) \times \mathcal{R}\}$ , where

$$\mathcal{I}_{n,t}(\theta, u) := \left(\frac{n}{m_n}\right)^{1/2} \{I(|x_t(\theta)| \leq c_n(\theta) + u\vartheta_n) - E[I(|x_t(\theta)| \leq c_n(\theta) + u\vartheta_n)]\}.$$

### 3. Examples

We study filters and plug-ins for ARMA, GARCH and nonlinear GARCH models with errors  $\epsilon_t$ . Throughout  $\epsilon_t$  is an i.i.d. random variable with an absolutely continuous distribution that is positive on  $\mathbb{R}$ , with  $\sup_{a \in \mathbb{R}} (\partial/\partial a)P(\epsilon_t \leq a) < \infty$ . In each example we impose a second order tail expansion for  $\epsilon_t$  (or a similar error) for brevity (cf., Hall (1982), Haeusler and Teugels (1985)):

$$P(|\epsilon_t| > a) = da^{-\kappa} \left(1 + ca^{-\beta}\right), \quad \beta, c, d, \kappa \in (0, \infty). \tag{3.1}$$

Let the fractile sequence  $\{m_n\}$  satisfy

$$m_n \rightarrow \infty \quad \text{and} \quad m_n = o(n^{2\tilde{\beta}/(2\tilde{\beta}+\kappa)}), \quad \text{where} \quad \tilde{\beta} := \min\{\beta, 1\}. \tag{3.2}$$

If a higher order regular variation property is used in place of (3.1), then (3.2) generally needs to be changed to ensure all parts of Assumption 2 hold. See

Haeusler and Teugels (1985) and Goldie and Smith (1987) for theory and examples. In practice  $m_n = o(n^{2\tilde{\beta}/(2\tilde{\beta}+\kappa)})$  is easily satisfied by setting  $m_n = O(\ln(n))$ , e.g.,  $m_n = \lceil \lambda \ln(n) \rceil$  for some  $\lambda > 0$ .

### 3.1. ARMA error

Consider estimating the tail index of the i.i.d. error  $\epsilon_t$  in an ARMA( $p, q$ ) process  $y_t = \sum_{i=1}^p a_i^0 y_{t-1} + \sum_{i=1}^q b_i^0 \epsilon_{t-i} + \epsilon_t$ , where  $\epsilon_t$  has tail (3.1), cf., Ling and Peng (2004). Define polynomials  $a(z) := 1 - \sum_{i=1}^p a_i z^i$  and  $b(z) := 1 + \sum_{i=1}^q b_i z^i$ , and assume  $a^0(z)$  and  $b^0(z)$  have no common roots, and that all roots lie outside the unit circle. Although  $y_t$  has the same tail index as  $\epsilon_t$  (Brockwell and Cline (1985)), use of the unobserved  $\epsilon_t$  for tail index estimation leads to an efficiency gain for tail index estimation (Ling and Peng (2004)).

Let  $\mathcal{A} \subset \mathbb{R}^p$  and  $\mathcal{B} \subset \mathbb{R}^q$  be compact sets of vectors  $a$  and  $b$  with polynomial roots outside the unit circle, and let  $\Theta = \mathcal{A} \times \mathcal{B}$ . Define recursively  $\epsilon_t(\theta) := y_t - \sum_{i=1}^p a_i y_{t-1} - \sum_{i=1}^q b_i \epsilon_{t-i}(\theta)$  hence  $\epsilon_t = \epsilon_t(\theta^0)$ . We restrict attention to tail-sum  $P(|\epsilon_t| > a)$  tail index estimation, hence the filter  $x_t(\theta)$  is a smoothed absolute error bounded from zero:  $x_t(\theta) = (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$  for any small  $\varepsilon > 0$ . Trivially  $x_t = (\epsilon_t^2 + \varepsilon)^{1/2}$  has the same tail index as  $|\epsilon_t|$ , while  $x_t(\theta) \geq \varepsilon^{1/2} > 0$  a.s. and  $x_t(\theta)$  is three-times continuously differentiable. In this, and each subsequent example, if an estimate of the left or right tail index is desired then we use  $(\epsilon_t^2(\theta)I(\epsilon_t(\theta) < 0) + \varepsilon)^{1/2}$  or  $(\epsilon_t^2(\theta)I(\epsilon_t(\theta) > 0) + \varepsilon)^{1/2}$ .

Valid plug-ins include a large variety of M-estimators since under mild conditions these are at least  $n^{1/2}$ -convergent. Examples include smooth M-estimators and LAD (Davis, Knight, and Liu (1992)), least tail-trimmed squares (Hill (2013)), and weighted LAD (Zhu and Ling (2012)).

**Lemma 2.** *Assumptions 1 – 3 hold. In the general ARMA case Assumption 4 holds for estimators in Davis (1996), Mikosch et al. (1995) and Zhu and Ling (2012). In the AR case Assumption 4 holds for estimators in Hill (2013) and Davis, Knight, and Liu (1992).*

### 3.2. Mis-specified autoregression error

Assume  $y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ , where  $\sum_{i=0}^{\infty} |\psi_i| < \infty$ , and  $\epsilon_t$  has a zero mean and tail (3.1) with index  $\kappa > 1$ . Assume  $y_t$  is  $\beta$ -mixing with summable coefficients. The mixing assumption is mild since  $y_t$  is  $\alpha$ -mixing with summable coefficients ((Doukhan, 1994, p.74)), and in the stationary ARMA case  $y_t$  is geometrically  $\beta$ -mixing (Mokkadem (1988), Doukhan (1994, p.99)). By independence and coefficient summability  $y_t$  has the same tail index as  $\epsilon_t$  (cf., Brockwell and Cline (1985)).

We can always model  $y_t$  as a finite order autoregression  $y_t = \sum_{i=1}^p \theta_i^0 y_{t-i} + v_t$  with a zero mean error  $v_t$ . For example, there exists  $\theta^0$  such that  $\sum_{i=1}^p \theta_i^0 y_{t-i}$  is the minimum  $L_r$ -norm predictor for  $r > 1$ , hence  $E[v_t^{<r-1>} y_{t-i}] = 0$  where  $v_t^{<r-1>}$  is the signed power  $sign(v_t) \times |v_t|^r$  (cf., Giles (1967)). Further, although a minimum  $\|\cdot\|_2$ -predictor is not well defined, the least squares estimator identifies  $\theta^0 := ([\sum_{i=0}^\infty \psi_{i-j_1} \psi_{i-j_2}]_{j_1, j_2=1}^p)^{-1} \times [\sum_{i=0}^\infty \psi_i \psi_{i-j}]_{j=1}^p$ , cf., Davis and Resnick (1986).

A finite order AR may not be the true data generating process, which means  $v_t$  may not be i.i.d. Nevertheless, use of a mis-specified model still leads to valid inference on the tail index of  $y_t$  since  $v_t$  is also a linear convolution of i.i.d.  $\epsilon_t$ , hence it has the same tail index as  $\epsilon_t$  and therefore  $y_t$  (Brockwell and Cline (1985)).

Define  $v_t(\theta) := y_t - \sum_{i=1}^p \theta_i y_{t-i}$  and consider the filter  $x_t(\theta) = (v_t^2(\theta) + \varepsilon)^{1/2}$  for small  $\varepsilon > 0$ . Let (3.2) hold.

**Lemma 3.** *Assumptions 1–3 hold, and the OLS estimator satisfies Assumption 4.*

**Remark 11.** Other estimators are evidently valid for AR models when the true DGP is  $y_t = \sum_{i=0}^\infty \psi_i \epsilon_{t-i}$  with possibly heavy tailed  $\epsilon_t$ , including each estimator cited in Lemma 2 as well as a generalization to linear processes of the heavy tail robust method of moments estimator in Hill (2014a). OLS, however, is easily shown to satisfy Assumption 4 by exploiting limit theory results for sample autocorrelations in Davis and Resnick (1986).

### 3.3. GARCH error

Consider estimating the tail index of the error  $\epsilon_t$  in a GARCH(1,1) model  $y_t = \sigma_t \epsilon_t$ , where  $\sigma_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 \sigma_{t-1}^2$ , and  $\omega^0 > 0, \alpha^0, \beta^0 \in (0, 1)$ . Assume the error  $\epsilon_t$  has a zero mean, unit variance, tail (3.1) with index  $\kappa > 2$ , and  $E[\ln(\alpha^0 \epsilon_t^2 + \beta^0)] < 0$ . Define  $\theta := [\omega, \alpha, \beta]'$  and assume  $\theta$  lies in a compact subset  $\Theta$  of  $(0, \infty) \times (0, 1) \times (0, 1)$  with  $\theta^0$  in the interior such that  $E[\ln(\alpha \epsilon_t^2 + \beta)] < 0 \forall \theta \in \Theta$ . Define the GARCH error function

$$\epsilon_t(\theta) = \frac{y_t}{\sigma_t(\theta)} = \frac{y_t}{(\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta))^{1/2}}.$$

The filter is  $x_t(\theta) = (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$  for small  $\varepsilon > 0$ . Let (3.2) hold.

Under the stated properties  $\{y_t, \sigma_t\}$  are strictly stationary and geometrically  $\beta$ -mixing (Nelson (1990), Carrasco and Chen (2002)). Valid plug-ins include at least QML, Log-LAD (Peng and Yao (2003)), Quasi-Maximum Tail-Trimmed Likelihood (Hill (2014a)), and weighted Laplace QML (Zhu and Ling (2011)). Other non-Gaussian QML estimators in Berkes and Horvath (2004) are valid provided moment conditions hold other than, and possibly in place of, the conventional GARCH moment conditions  $E[\epsilon_t] = 0$  and  $E[\epsilon_t^2] = 1$  (see Hill (2014a)).

**Lemma 4.** *Assumptions 1–3 hold, and Log-LAD, Quasi-Maximum Tail-Trimmed Likelihood and weighted Laplace QML satisfy Assumption 4. QML satisfies Assumption 4 when  $\kappa > 4$ , and when  $\kappa \in (2, 4]$  provided  $m_n = o(n^{2-4/\kappa})$ .*

**Remark 12.** If  $\kappa \in (2, 4]$  then the QML rate is  $n^{1-2/\kappa}/\mathcal{L}(n) \leq n^{1/2}/\mathcal{L}(n)$  for some slowly varying  $\mathcal{L}(n) \rightarrow \infty$  (Hall and Yao (2003)). QML therefore satisfies Assumption 4 only if  $m_n = o(n^{2-4/\kappa})$ , in addition to (3.2). It suffices to set  $m_n = O(\ln(n))$ .

### 3.4. AR-ARCH: ARCH error filter

Consider an AR(1)-ARCH(1) model  $y_t = \phi^0 y_{t-1} + (\omega^0 + \alpha^0 y_{t-1}^2)^{1/2} \epsilon_t$ , where  $|\phi^0| < 1$ ,  $\omega^0 > 0$ , and  $\alpha^0 \in (0, 1)$ , and  $\epsilon_t$  is i.i.d. with a zero mean, unit variance, and tail (3.1) with index  $\kappa > 2$ . Cf., Borkovec and Klüppelberg (2004). Define  $\theta = [\phi, \omega, \alpha]'$  and the ARCH error  $\epsilon_t(\theta) := (y_t - \phi y_{t-1})/(\omega + \alpha y_{t-1}^2)^{1/2}$  on any compact subset  $\Theta$  of  $(-1, 1) \times (0, \infty) \times (0, 1)$ . Consider a filter derived from the ARCH error  $x_t(\theta) = (\epsilon_t^2(\theta) + \varepsilon)^{1/2}$  for small  $\varepsilon > 0$ . Let (3.2) hold.

There are few results in the literature concerning estimation of ARMA-GARCH models with heavy tailed errors. We therefore only consider Zhu and Ling's (2011) estimator. Note that Hill's (2014a) Quasi-Maximum Tail-Trimmed Likelihood and method of moments estimators easily extend to ARMA-GARCH models.

**Lemma 5.** *Assumptions 1–3 hold and weighted Laplace QML satisfies Assumption 4.*

### 3.5. AR-GARCH: AR error filter

The model is an AR(1)  $y_t = \theta^0 y_{t-1} + u_t$  with GARCH(1,1) error  $u_t = \sigma_t \epsilon_t$  and  $\sigma_t^2 = \omega^0 + \alpha^0 u_{t-1}^2 + \beta^0 \sigma_{t-1}^2$ , where  $|\theta^0| < 1$ ,  $\omega^0 > 0$ ,  $\alpha^0, \beta^0 \in (0, 1)$ ,  $E[\ln(\alpha^0 \epsilon_t^2 + \beta^0)] < 0$  and  $\epsilon_t$  is i.i.d. with a zero mean and unit variance. In this case the filter is based on the AR error  $x_t(\theta) = (u_t^2(\theta) + \varepsilon)^{1/2}$ . Distribution smoothness of  $\epsilon_t$  and stationarity ensure  $u_t$  has a power law tail, irrespective of whether  $\epsilon_t$  has tail (3.1).

**Lemma 6.** *Assumptions 1 and 3 hold, and  $x_t$  has tail  $P(|x_t| > a) = da^{-\kappa}(1 + o(1))$ . If  $\epsilon_t$  has a symmetric distribution then Hill's (2014b) Generalized Empirical Likelihood estimator satisfies Assumption 4.*

Assumption 2.a remains to be resolved in general, while Assumption 2.b can always be enforced. Although GARCH  $u_t$  (and therefore  $x_t$ ) has a first order power law tail  $P(|u_t| > a) = da^{-\kappa}(1 + o(1))$  under very general conditions on  $\epsilon_t$ , we are not aware of a set of conditions that ensure a second order property

like (3.1). See Basrak, Davis, and Mikosch (2002) for references, and see Geluk et al. (1997) for second order regular variation that extends to linear processes  $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$  when i.i.d.  $\epsilon_t$  has second order tail (3.1). If, however,  $u_t$  has tail (3.1) then so does  $u_t(\theta)$  by the proof of Lemma 4, cf., Lemma 2. In order to apply Theorem 1 we would have to assume a second order property on  $u_t$  such that all aspects of Assumption 2.a hold.

Plug-in choices are limited due to the incomplete theory for parametric estimation with a non-i.i.d. heavy tailed error. If we assume the GARCH error has a symmetric distribution then Hill's (2014b) class of GEL estimators is  $n^{1/2}$ -convergent and therefore valid for Assumption 4. Hill (2013, 2014a) develops tail-trimmed M-estimators for AR and GARCH models with i.i.d. errors. It is straightforward to combine technical arguments in those papers for a tail-trimmed least squares estimator for an AR model with a geometrically  $\beta$ -mixing error that covers GARCH. In this case at worst tail-trimming leads to an estimator with the rate  $n^{1/2}/g_n$  for some sequence of numbers  $\{g_n\}$  that satisfies  $1 \leq g_n < n$  and  $g_n \rightarrow \infty$  as slowly as desired (see the proof of Lemma 4, above). Zhu and Ling (2011, 2012) present estimators for ARMA models with an i.i.d. error and ARMA-GARCH where all parameters must be estimated. Evidently their theory can be extended to cover AR parameter estimation when the error is non-i.i.d.

### 3.6. AR with heteroscedastic error: AR error filter

In the previous example we imposed a GARCH structure on the error, but in practice we may only assume  $y_t = \theta^0 y_{t-1} + u_t$  with error  $u_t = \sigma_t \epsilon_t$  and not know the form of heteroskedasticity  $\sigma_t$ . In view of Lemma 6 and the subsequent remark, Assumptions 1-3 all hold if the AR error  $u_t$  has tail (3.1) and is a geometrically  $\beta$ -mixing martingale difference. In terms of Assumption 4, Hill's (2014b) GEL estimator is  $n^{1/2}$ -convergent and therefore valid provided  $\epsilon_t$  has a symmetric distribution.

### 3.7. Nonlinear or asymmetric GARCH error

Consider a nonlinear or asymmetric GARCH(1,1) model  $y_t = \sigma_t \epsilon_t$ , where  $\sigma_t^2 = f(\sigma_{t-1}^2, y_{t-1}, \theta^0)$  for some mapping  $f : [0, \infty) \times (-\infty, \infty) \times \Theta \rightarrow [0, \infty)$  that is twice differentiable in each argument, and  $\Theta$  is a compact subset of  $\mathbb{R}^k$  for some  $k \geq 1$ . As long as  $\epsilon_t$  has a finite fourth moment then, under mild smoothness and boundedness conditions on  $f$ , the QML estimator is  $n^{1/2}$ -convergent. See Francq and Zakoian (2010) and Meitz and Saikkonen (2011) for extensive references covering Threshold GARCH, GJR-GARCH, Switching GARCH, Absolute Value GARCH, and so on.

In the case of heavy tailed errors, however, very little theory exists for estimation. If  $\epsilon_t$  has a symmetric distribution then Hill's (2014b) GEL estimator is  $n^{1/2}$ -convergent and therefore valid. In principle Zhu and Ling's (2011) non-Gaussian QML estimator for ARMA-GARCH models can be used on nonlinear models. Hill's (2014a) Quasi-Maximum Tail-Trimmed Likelihood estimator covers a large class of symmetric and asymmetric GARCH models with heavy tailed errors, without any change in asymptotic theory, as long as  $(\partial/\partial\theta)\sigma_t^2/\sigma_t^2$  is  $L_{2+\iota}$ -bounded for tiny  $\iota > 0$ , as it is for linear GARCH with  $\alpha^0 + \beta^0 > 0$ .

## 4. Simulation Study

### 4.1. Set-up

We drew 10,000 samples of size  $n \in \{100, 250, 500\}$  of AR and GARCH random variables. In the AR case  $y_1 = \epsilon_1$  and  $y_t = \theta^0 y_{t-1} + \epsilon_t$  for  $t \geq 2$ , where  $\theta^0 \in \{0.4, 0.9\}$  and  $\epsilon_t$  is distributed i.i.d. Pareto,  $P(\epsilon_t < -a) = P(\epsilon_t > a) = 0.5(1+a)^{-\kappa}$  with  $\kappa = 1.5$ . In the GARCH case,  $y_t = h_t \epsilon_t$  where  $h_1^2 = \omega^0$ ,  $h_t^2 = \omega^0 + \alpha^0 y_{t-1}^2 + \beta^0 h_{t-1}^2$  for  $t \geq 2$ ,  $\{\omega^0, \alpha^0\} = \{1, 0.6\}$ , and  $\beta^0 \in \{0.2, 0.4\}$ ;  $\epsilon_t$  was distributed i.i.d. Pareto as above, with  $\kappa = 2.5$ . In this case the complete parameter set is  $\theta^0 = \{\omega^0, \alpha^0, \beta^0\}$ . We drew  $20n$  observations and retained the last  $n$  for analysis. We estimated  $\kappa$  for  $x_t = (\epsilon_t^2 + 10^{-10})^{1/2}$  using either the true  $\epsilon_t$  or a filter  $\epsilon_t(\hat{\theta}_n)$  for the sake of comparison.

In the AR case we computed  $\theta^0$  by least squares. Since  $\epsilon_t$  and  $y_t$  have the same tail index, for comparison we also estimated  $\kappa$  using the observed data  $|y_t|$ .

In the GARCH case we computed  $\theta^0$  on  $\Theta = [0, 2] \times [0.01, 0.99] \times [0.01, 0.99]$  using Hill's (2014a) Quasi-Maximum Tail-Trimmed Likelihood estimator. Take an iterated volatility process  $\hat{h}_1^2(\theta) = \omega$  and  $\hat{h}_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta \hat{h}_{t-1}^2(\theta)$  for  $t \geq 2$ , and define  $\hat{\epsilon}_t(\theta) := y_t/\hat{h}_t(\theta)$  and  $\hat{\mathcal{E}}_t(\theta) := \hat{\epsilon}_t^2(\theta) - 1$ . Let  $\hat{\mathcal{E}}_t^{(-)}(\theta) := \hat{\mathcal{E}}_t(\theta)I(\hat{\mathcal{E}}_t(\theta) < 0)$  and  $\hat{\mathcal{E}}_t^{(+)}(\theta) := \hat{\mathcal{E}}_t(\theta)I(\hat{\mathcal{E}}_t(\theta) \geq 0)$ , with  $\{\hat{\mathcal{E}}_t^{(-)}(\theta), \hat{\mathcal{E}}_t^{(+)}(\theta)\}$  the order statistics  $\hat{\mathcal{E}}_{(1)}^{(-)}(\theta) \leq \hat{\mathcal{E}}_{(2)}^{(-)}(\theta) \leq \dots \leq \hat{\mathcal{E}}_{(n)}^{(-)}(\theta) \leq 0$  and  $\hat{\mathcal{E}}_{(1)}^{(+)}(\theta) \geq \hat{\mathcal{E}}_{(2)}^{(+)}(\theta) \geq \dots \geq \hat{\mathcal{E}}_{(n)}^{(+)}(\theta) \geq 0$ . Define trimming indicators  $\mathcal{I}_{n,t}^{(\mathcal{E})}(\theta) := I(\hat{\mathcal{E}}_{(k_1, n)}^{(-)}(\theta) \leq \hat{\mathcal{E}}_t(\theta) \leq \hat{\mathcal{E}}_{(k_2, n)}^{(+)}(\theta))$  and  $\mathcal{I}_{n,t}^{(y)} := I(|y_t| \leq y_{(k_n^{(y)})}^{(a)})$ . The criterion is  $Q_n(\theta) := \sum_{t=2}^n \{\ln \hat{h}_t^2(\theta) + \hat{\epsilon}_t^2(\theta)\} \times \mathcal{I}_{n,t}^{(\mathcal{E})}(\theta)\mathcal{I}_{n,t-1}^{(y)}(\theta)$ , where  $k_{2,n}^{(\mathcal{E})} = [0.025n/\ln(n)]$ ,  $k_{1,n}^{(\mathcal{E})} = 35k_{2,n}^{(\mathcal{E})}$ , and  $k_n^{(y)} = [0.2 \ln(n)]$ . In view of the Pareto error form, QMTTL is asymptotically unbiased, consistent and normal, and satisfies Assumption 4, cf., (Hill, 2014a, Sec. 2.3). We randomized 100 initial values to obtain 100 estimates, and picked the one that minimized  $Q_n(\theta)$ .

Least squares and Quasi-Maximum Tail-Trimmed Likelihood are at least  $n^{1/2}/\mathcal{L}(n)$ -convergent for slowly varying  $\mathcal{L}(n) \rightarrow \infty$ . Hence, by Theorem 1,  $\hat{\theta}_n$

does not impact the limit distribution of  $\hat{\kappa}_{m_n}(\hat{\theta}_n)$  when  $m_n$  is regularly varying, i.e.  $m_n = \lfloor \lambda n^\xi \rfloor$  for  $\lambda > 0$  and  $\xi \in (0, 1)$ . Since  $\epsilon_t$  is i.i.d., it follows by Remark 7 and the Mean Value Theorem that  $m_n^{1/2}(\hat{\kappa}_{m_n}(\hat{\theta}_n) - \kappa) \xrightarrow{d} N(0, \kappa^2)$ . The appropriate scale for  $m_n^{1/2}(\hat{\kappa}_{m_n}(\hat{\theta}_n) - \kappa)$ , technically, is  $\sigma_{m_n} \kappa^2$ , where by independence,  $\sigma_{m_n}^2 \rightarrow \kappa^{-2}$ .

In Figures 1 and 2 we plot simulation averages of Hill’s (1975) estimator over fractiles  $m_n \in \{5, \dots, 100\}$  for the case  $n = 250$ . Results using  $n = 100$  or  $n = 500$  were qualitatively similar and are therefore omitted. In Table 1 we report two optimal  $m'_n$ s: one is the average  $m_n$  across samples that minimizes  $|\hat{\kappa}_{m_n} - \kappa|$  over  $m_n \in \{5, \dots, 150\}$ , denoted  $m_n^*$ ; the other minimizes the simulation mse of  $\hat{\kappa}_{m_n}$  over  $m_n \in \{5, \dots, 150\}$ , denoted  $\hat{m}_n$ . In this study plotting  $\hat{\kappa}_{m_n}$  past  $m_n = 100$  was redundant, except in the AR case where  $\theta^0 = 0.9$  and  $y_t$  was used for tail index estimation. In this case the optimal fractile was above 100 in many samples. Hence in all cases we computed  $m_n^*$  over a greater range than plotted. See also Table 1 for the average  $\hat{\kappa}_{m_n^*}$  across samples, and associated mse estimates.

Let  $\hat{\kappa}_{m_n}$  denote the simulation average estimate for any case. We include two 95% confidence bands. One is based on the assumption that the data are i.i.d., hence we report  $\hat{\kappa}_{m_n} \pm 1.96\hat{\kappa}_{m_n}/m_n^{1/2}$ . This is only correct asymptotically in the case of using residuals or the true errors since  $\sigma_{m_n}^2 \rightarrow \kappa^{-1}$ . It is incorrect asymptotically when  $\hat{\kappa}_{m_n}$  is estimated using AR data  $y_t$ . Further, in small samples, even when  $\sigma_{m_n}^2 \rightarrow \kappa^{-1}$  is true,  $\hat{\kappa}_{m_n}^2$  may underrepresent the true sampling dispersion of  $\hat{\kappa}_{m_n}$  since  $\epsilon_t(\hat{\theta}_n)$  is serially dependent. See Hill and Shneyerov (2013, Sec. 3.3) for discussion. We therefore also computed Hill’s (2010) consistent kernel estimator of  $\sigma_{m_n}^2$ . Writing  $\hat{x}_t = x_t(\hat{\theta}_n)$  and  $(z)_+ := \max\{z, 0\}$ , the estimator is:

$$\hat{\sigma}_{m_n}^2 = \frac{1}{m_n} \sum_{s,t=1}^n \mathcal{K}\left(\frac{s-t}{b_n}\right) \left\{ \ln\left(\frac{\hat{x}_s}{\hat{x}_{(m_n+1)_+}}\right) - \frac{m_n}{n} \hat{\kappa}_{m_n}^{-1} \right\} \\ \times \left\{ \ln\left(\frac{\hat{x}_t}{\hat{x}_{(m_n+1)_+}}\right) - \frac{m_n}{n} \hat{\kappa}_{m_n}^{-1} \right\},$$

with Bartlett kernel  $\mathcal{K}(x) = (1 - |x|)_+$  and bandwidth  $b_n = n^{0.25}$ . If  $\hat{\sigma}_{m_n}^2$  denotes the simulation average, then by Remark 7 and the Mean Value Theorem the asymptotic bands are  $\hat{\kappa}_{m_n} \pm 1.96\hat{\sigma}_{m_n} \hat{\kappa}_{m_n}^2/m_n^{1/2}$ , and a robust mse estimator for  $m_n^{1/2}(\hat{\kappa}_{m_n}(\hat{\theta}_n) - \kappa)$  is  $\hat{\sigma}_{m_n}^2 \hat{\kappa}_{m_n}^4$ .

**4.2. Summary of results**

AR or GARCH filters  $\epsilon_t(\hat{\theta}_n)$  led to roughly the same tail index estimate  $\hat{\kappa}_{m_n}$  as if the true unobserved errors  $\epsilon_t$  were used. The optimal  $m_n$  were roughly

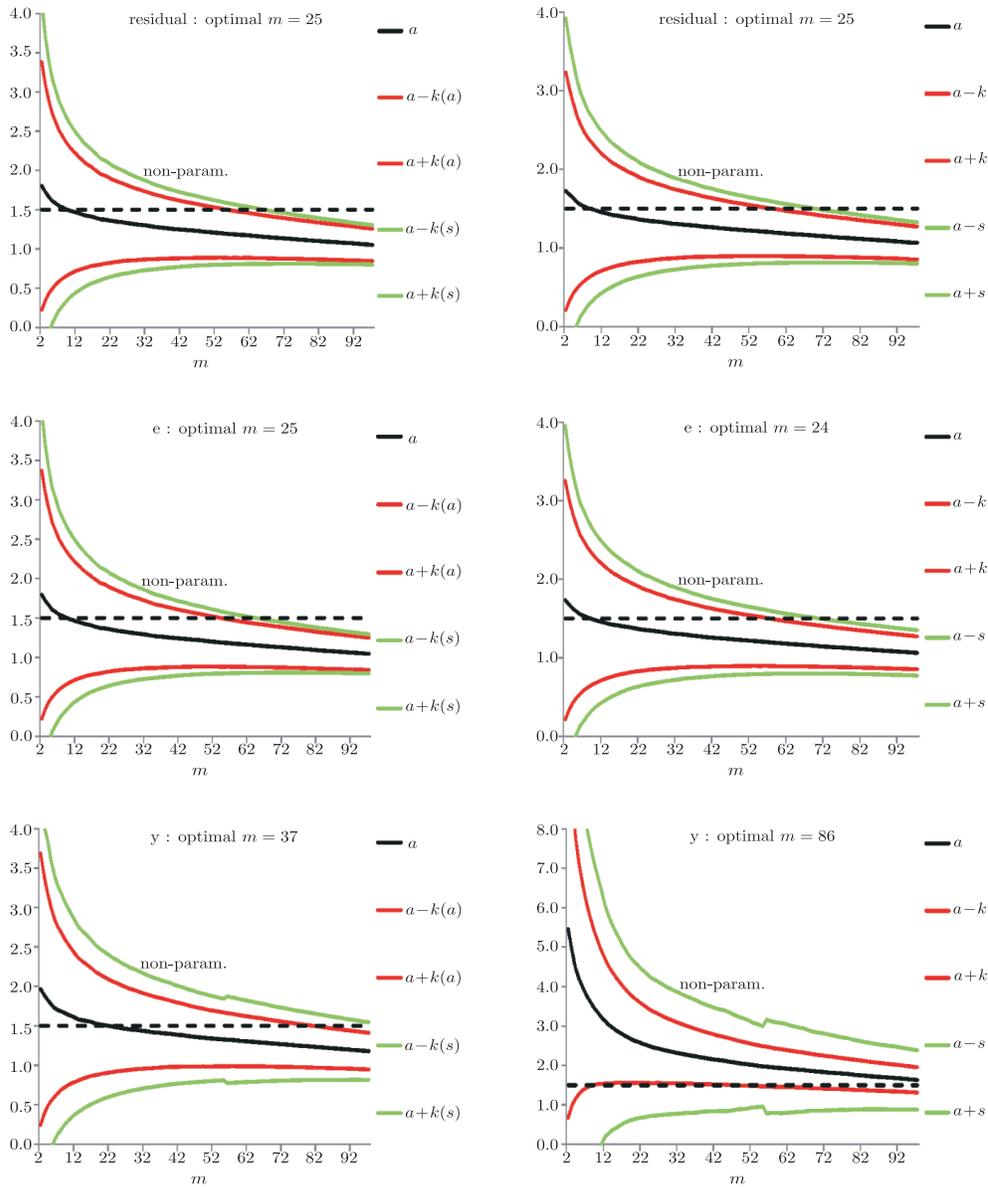


Figure 1. Hill-Plots for AR,  $\kappa = 1.5$ . The model is  $y_t = \theta^0 y_{t-1} + \epsilon_t$ . Left panels:  $\theta^0 = 0.4$ , and right panels:  $\theta^0 = 0.8$ . The top row is for the filter  $\epsilon_t(\hat{\theta}_n)$ , the middle for the true error  $\epsilon_t$  and the bottom for  $y_t$ . The plotted lines are  $\hat{\kappa}_{m_n}$  and two 95% confidence bands. The outer "non-param" bands use the robust nonparametric mse estimator. The other band uses the mse estimator in the i.i.d. case. The optimal  $m_n$  minimizes  $|\hat{\kappa}_{m_n} - \kappa|$  over  $\{1, \dots, 150\}$ . The reported  $m_n$  is the average over 10,000 samples.

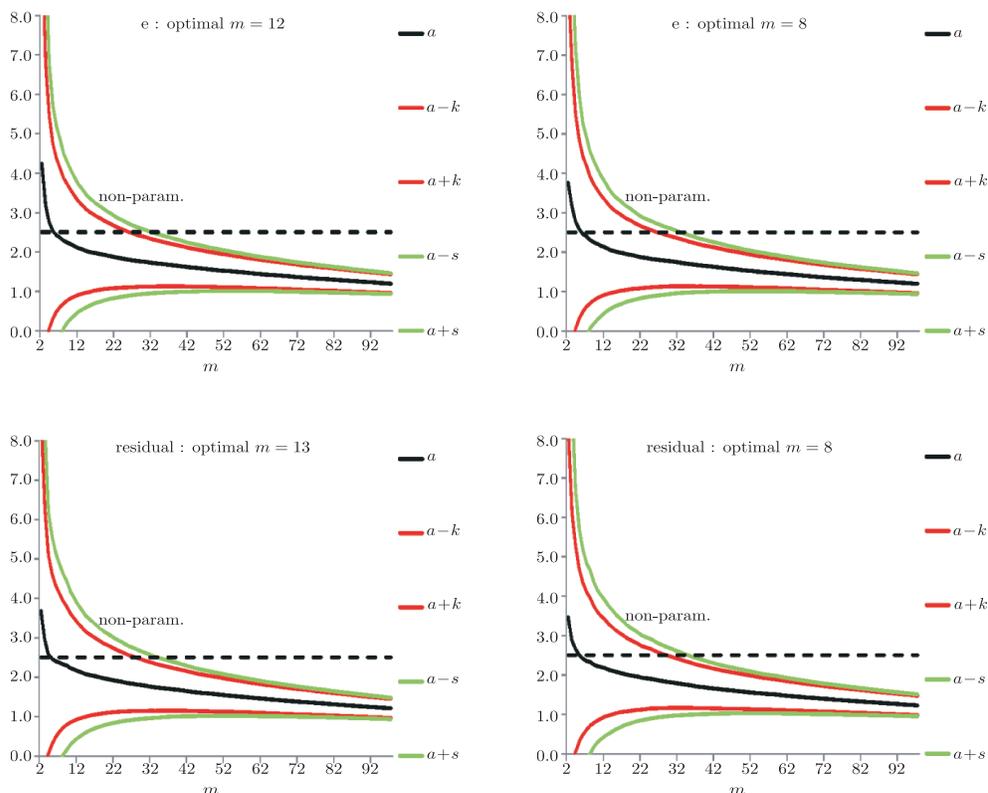


Figure 2. Hill-Plots for GARCH,  $\kappa = 2.5$ . The model is  $y_t = h_t \epsilon_t$  where  $h_t^2 = 1 + 0.6y_{t-1}^2 + \beta^0 h_{t-1}^2$ . Left panels:  $\beta^0 = 0.2$ , and right panels:  $\beta^0 = 0.4$ . The top row is for the true error  $\epsilon_t$  and the bottom row is for the filter  $\epsilon_t(\hat{\theta}_n)$ . The plotted lines are  $\hat{\kappa}_{m_n}$  and two 95% confidence bands. The outer “non-param” bands use the robust nonparametric mse estimator. The other band uses the mse estimator in the i.i.d. case. The optimal  $m_n$  minimizes  $|\hat{\kappa}_{m_n} - \kappa|$  over  $\{1, \dots, 150\}$ . The reported  $m_n$  is the average over 10,000 samples.

identical for  $\epsilon_t$  and  $\epsilon_t(\hat{\theta}_n)$  in all cases, and were 1.5 to 3.5 times larger for AR  $y_t$  depending on  $\theta^0 = 0.4$  or  $0.9$ .

Figures 1 and 2 and Table 1 also reveal several important traits of Hill’s (1975) estimator. First, in the AR case  $\theta^0 = 0.9$ , strong positive serial dependence inflated the tail index value at each fractile  $m_n$  giving the appearance of thinner tails.

Second, in the AR case the use of a filter  $\epsilon_t(\hat{\theta}_n)$  with a least squares plug-in  $\hat{\theta}_n$ , as opposed to the observed data  $y_t$ , led to an efficiency improvement. See Resnick and Stărică (1997), cf., Ling and Peng (2004). Table 1 shows the average robust non-parametric mse estimates  $\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4$  at the optimal fractile  $m_n^*$

Table 1.

	AR: $\kappa = 1.5$						GARCH: $\kappa = 2.5$			
	$\theta^0 = 0.4$			$\theta^0 = 0.9$			$\beta^0 = 0.2$		$\beta^0 = 0.4$	
	$\epsilon_t(\hat{\theta}_n)$	$\epsilon_t$	$y_t$	$\epsilon_t(\hat{\theta}_n)$	$\epsilon_t$	$y_t$	$\epsilon_t(\hat{\theta}_n)$	$\epsilon_t$	$\epsilon_t(\hat{\theta}_n)$	$\epsilon_t$
$\hat{m}_n$	31 <sup>a</sup>	25	47	32	28	142	16	13	11	9
$m_n^*$	25 <sup>b</sup>	25	37	25	24	86	13	12	8	8
$\hat{\kappa}_{m_n^*}$	1.48 <sup>c</sup>	1.48	1.49	1.48	1.48	1.54	2.57	2.51	2.58	2.52
$\hat{\kappa}_{m_n^*}^2$	2.19 <sup>d</sup>	2.19	2.25	2.19	2.19	2.46	6.60	6.30	6.66	6.25
$\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4$	1.98 <sup>e</sup>	1.91	6.29	2.10	2.02	67.3	8.62	8.10	11.2	10.4

- The fractile that minimizes the simulation mse over  $m_n \in \{1, \dots, 150\}$ .
- The average fractile that minimizes  $|\hat{\kappa}_{m_n} - \kappa|$  over  $m_n \in \{1, \dots, 150\}$ .
- The average closest tail index estimate  $\hat{\kappa}_{m_n}$  to  $\kappa$  for each sample.
- The average asymptotic mse estimate of  $\hat{\kappa}_{m_n^*}$  under the i.i.d. assumption.
- The average robust non-parametric mse estimate of  $\hat{\kappa}_{m_n^*}$ .

were nearly identical when  $\epsilon_t$  or  $\epsilon_t(\hat{\theta}_n)$  was used, but were larger, and possibly much larger, when  $y_t$  was used.

Third, in the GARCH case we used an iterated volatility sequence  $\hat{h}_1^2 = \hat{\omega}_n$  and  $\hat{h}_t^2 = \hat{\omega}_n + \hat{\alpha}_n y_{t-1}^2 + \hat{\beta}_n \hat{h}_{t-1}^2$  for  $t \geq 2$ , hence an iterated estimated error sequence  $y_t/\hat{h}_t$  for computing the tail index of  $\epsilon_t$ . By comparison,  $\epsilon_t = y_t/h_t$  was drawn using a large burn-in of  $19n = 4,750$  (discarded) observations. This implies  $y_t/\hat{h}_t$  involved both sampling error from  $\hat{\theta}_n$  and a potential under approximation of the true volatility at small  $t$ , hence  $y_t/\hat{h}_t$  appeared more heavy tailed than  $\epsilon_t$ . In fact,  $y_t/\hat{h}_t$  appeared even more heavy tailed for larger GARCH parameter values  $\beta^0$  due to a monotonically larger under-representation of volatility.

Fourth, there was a comparatively large discrepancy between Hill-estimator mse estimates  $\hat{\kappa}_{m_n^*}^2$  and  $\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4$  when the data were dependent. Specifically  $\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4 > \hat{\kappa}_{m_n^*}^2$ , suggesting dependence increased Hill-estimator dispersion. The difference was massive in the AR case when  $y_t$  with  $\theta^0 = 0.9$  was used for tail index estimation, but also arose with residuals since sampling error adds serial dependence. However, in the AR case, the nonparametric mse  $\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4$  was close to  $\hat{\kappa}_{m_n^*}^2$  for  $\epsilon_t$  and  $\epsilon_t(\hat{\theta}_n)$ , as it should be, since  $\hat{\sigma}_{m_n^*}^2 \xrightarrow{p} \kappa^{-2}$ . In the GARCH case this should also be true, but  $\hat{\sigma}_{m_n^*}^2 \hat{\kappa}_{m_n^*}^4$  was comparatively larger than  $\hat{\kappa}_{m_n^*}^2$  since the residuals  $y_t/\hat{h}_t$  exhibited a greater degree of serial dependence due to sampling error compounded with the fact that  $\hat{h}_t^2$  under-represented true volatility.

Fifth,  $\hat{m}_n$  and  $m_n^*$  roughly coincided for i.i.d.  $\epsilon_t$ , but otherwise did not coincide for dependent data. There was a small difference  $\hat{m}_n > m_n^*$  for residuals  $\epsilon_t(\hat{\theta}_n)$ , and substantial  $\hat{m}_n > m_n^*$  for AR data when  $y_t$  was used for estimation.

The latter matches evidence for a large dispersion in  $\hat{\kappa}_{m_n}$  for highly dependent data: variance dominates the mse, hence minimizing the mse leaves a high degree of bias. This suggests that conventional adaptive tail index estimation methods for i.i.d. data may not be appropriate for filtered data (e.g., Hill (1975), Hall (1982), Hall and Welsh (1985), Huisman et al. (2001) Groeneboom, Lopuhaä and de Wolf (2003)).

### Acknowledgements

We thank a referee and Editor Qiwei Yao for helpful comments and suggestions.

### References

- Aguilar, M. and Hill, J. B. (2014). Robust score and portmanteau tests of volatility spillover. *J. Economet.*, forthcoming.
- Antoine, B. and Renault, E. (2012). Efficient minimum distance estimation with multiple rates of convergence. *J. Economet.* **170**, 350-367.
- Basrak, B., Davis, R. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99**, 95-115.
- Beran, J., Das, B. and Schnell, D. (2012). On robust tail index estimation for linear long-memory processes. *J. Time Ser. Anal.* **33**, 406-423.
- Berkes, I. and Horvath, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32**, 633-655.
- Borkovec, M. and Klüppelberg, C. (2004). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220-1241.
- Bradley, R. C. (1993). Equivalent mixing conditions for random fields. *Ann. Probab.* **21**, 1921-1926.
- Brockwell, P. J. and Cline, D. B. H. (1985). Linear prediction of ARMA processes with infinite variance. *Stochastic Process. Appl.* **19**, 281-296.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Economet. Th.* **18**, 17-39.
- Cline, D. B. H. (2007). Regular variation of order 1 nonlinear AR-ARCH models. *Stochastic Process. Appl.* **117**, 840-861.
- Davis, R. A. (1996). Gauss-newton and M-estimation for ARMA processes with infinite variance. *Stochastic Process. Appl.* **63**, 75-95.
- Davis, R. A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40**, 145-180.
- Davis, R. A. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.
- de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc.* **61**, 381-395.
- Dekkers, A. L. M., Einmahl, J. H. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833-1855.
- Doukhan, P. (1994). *Mixing Properties and Examples*. Springer-Verlag, New York.

- Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. P.* **31**, 393-427.
- Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6**, 899-929.
- Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press, Cambridge, UK.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.
- Franco, C. and Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, West Sussex, UK.
- Geluk, J., de Haan, L., Resnick, S. and Stărică, C. (1997). Second-order regular variation, convolution and the central limit theorem. *Stochastic Process. Appl.* **69**, 139-159.
- Giles, J. R. (1967). Classes of semi-inner-product spaces. *Trans. Amer. Math. Soc.* **129**, 436-446.
- Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: Theory and applications. *Q. J. Math.* **38**, 45-71.
- Groeneboom, P., Lopushaä, H. P. and de Wolf, P. P. (2003). Kernel-type estimators for the extreme value index. *Ann. Statist.* **31**, 1956-1995.
- Haeusler, E. and Teugels, J. L. (1985). On asymptotic normality of hill's estimator for the exponent of regular variation. *Ann. Statist.* **13**, 743-756.
- Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37-42.
- Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.
- Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285-317.
- Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.
- Hill, J. B. (2009). On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *J. Statist. Plann. Inference* **139**, 2091-2110.
- Hill, J. B. (2010). On tail index estimation for dependent heterogeneous data. *Economet. Th.* **26**, 1398-1436.
- Hill, J. B. (2011a). Tail and non-tail memory with applications to extreme value and robust statistics. *Economet. Th.* **27**, 844-884.
- Hill, J. B. (2011b). Extremal memory of stochastic volatility with an application to tail shape inference. *J. Statist. Plann. Inference* **141**, 663-676.
- Hill, J. B. (2013). Least tail-trimmed squares for infinite variance autoregressions. *J. Time Ser. Anal.* **34**, 168-186.
- Hill, J. B. (2014a). Robust estimation for heavy tailed GARCH. *Bernoulli*, forthcoming.
- Hill, J. B. (2014b). Robust generalized empirical likelihood for heavy tailed autoregressions with conditionally heteroscedastic errors. Working Paper, Dept. of Economics, University of North Carolina.
- Hill, J. B. (2014c). Supplementary Material for *Tail Index Estimation for a Filtered Dependent Time Series*.
- Hill, J. B. and Aguilar, M. (2013). Moment condition tests for heavy tailed time series. *J. Economet.* **72**, 255-274.

- Hill, J. B. and Shneyerov, A. (2013). Are there common values in first-price auctions? A tail-index nonparametric test. *J. Economet.* **174**, 144-164.
- Hoffmann-Jørgensen, J. (1991). Convergence of stochastic processes on polish spaces. *Various Publication Series* **39**, Matematisk Institute, Aarhus University.
- Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19**, 1547-1569.
- Huisman, R., Koedijk, K., Kool, C. and Palm, F. (2001). Tail-index estimates in small samples. *J. Bus. Econom. Statist.* **19**, 208-216.
- Ilić, I. (2012). On tail index estimation using a sample with missing observations. *Statist. Probab. Lett.* **82**, 949-958.
- Kulik, R. and Soulier, P. (2011). The tail empirical process for long memory stochastic volatility sequences. *Stochastic Process. Appl.* **121**, 109-134.
- Ling, S. and Peng, L. (2004). Hill's estimator for the tail index of an ARMA model. *J. Statist. Plann. Inference* **123**, 279-293.
- Liu, J.-C. (2006). On the tail behaviors of a family of GARCH processes. *Economet. Th.* **22**, 852-862.
- Meitz, M. and Saikkonen, P. (2011). Parameter estimation in nonlinear AR-GARCH models. *Economet. Th.* **27**, 1236-1278.
- Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427-1451.
- Mikosch, T., Gadrich, T., Klüppelberg, C. and Alder, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.* **23**, 305-326.
- Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Process. Appl.* **29**, 309-315.
- Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Economet. Th.* **6**, 318-334.
- Peligrad, M. (1996). On the asymptotic normality of sequences of weak dependent random variables. *J. Theoret. Probab.* **9**, 703-715.
- Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967-975.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.
- Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, Berlin.
- Resnick, S. and Stărică, C. (1997). Asymptotic behavior of Hill's estimator for autoregressive data. *Comm. Statist. Stochastic Models* **13**, 703-721.
- Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.* **119**, 468-490.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.* **4**, 43-47.
- Zhu, K. and Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.* **39**, 2131-2163.
- Zhu, K. and Ling, S. (2012). The global weighted LAD estimators for finite/infinite variance ARMA( $p, q$ ) models. *Economet. Th.* **28**, 1065-1086.

Department of Economics, University of North Carolina at Chapel Hill, NC 27514, USA.

E-mail: jbhill@email.unc.edu

(Received July 2012; accepted March 2014)