Diagnosing and treating the fat tails in financial returns data

Stefan Mittnik a, *, Marc S. Paolella a, Svetlozar T. Rachev b

a Institute of Statistics and Econometrics, University of Kiel, Olshausenstr. 40, D-24098 Kiel, Germany
b Institute of Statistics and Mathematical Economics, University of Karlsruhe, Kollegium am Schloss Bau II, D-76128, Karlsruhe, Germany

Abstract

We address an open and important question regarding the nature of the fat tails found in financial-return data, which has been raised by Ghose and Kroner [Journal of Empirical Finance, 2 (1995) 225]. These authors find that two classes of models used for modeling financial returns, namely the independent and identically distributed (iid) stable Paretoian and the GARCH assumption, have several features in common, with the latter being preferred. We advocate models that combine the two allegedly disjoint paradigms, i.e., GARCH processes driven by stable Paretoian innovations, and investigate some of their theoretical and small-sample properties. Finally, we demonstrate the plausibility of the new models for several exchange-rate series involved in the Asian crisis. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Asian crisis; GARCH; Stable Paretoian distribution; Summability test

1. Introduction

Starting with the work of Fama (1963, 1965) and Mandelbrot (1963), there have been numerous studies investigating the appropriateness of the stable Paretoian distribution for modeling the unconditional distribution of asset returns (cf. Mittnik and Rachev, 1993a; McCulloch, 1997a; Rachev and Mittnik, 2000).
Despite the theoretical appeal and empirical goodness of fit of the stable class of models for describing highly leptokurtic, oftentimes skewed, financial return data, there is evidence that they are not always suitable, most notably in the form of significant volatility clustering. The class of GARCH models (cf. Gourieroux, 1997; Palm, 1997) has been very successful in modeling such behavior; yet for financial time series, they are still inadequate for explaining residual behavior inconsistent with the normal distribution. These have given rise to GARCH models driven by generalized exponential (Nelson, 1991), Student’s t (Bollerslev, 1987) and even distributions that nest these (Bollerslev et al., 1994; Paolella, 1997).

One arguable disadvantage of such models is that the innovations are not closed under addition. This seemingly innocuous property is quite important: One of the most ubiquitous assumptions common to virtually all statistical models is that the innovations, or residuals, are composed of sums of (usually unobservable) independent and identically distributed (iid) effects too numerous and difficult to incorporate into the model. From the Generalized Central Limit Theorem, the only non-degenerate distribution arising as such a sum are members of the stable Paretian class, of which the normal distribution is a special case.

It appears as though a dichotomy has arisen: Is the source of the fatter-tailed nature of asset returns due to GARCH effects or are the data realizations from a stable Paretian distribution? This question has been recently posed by Ghose and Kroner (1995), hereafter GK, who present evidence that the former, i.e., GARCH, is more capable of explaining the fat-tailedness. In some respects, this result is not overly surprising. A cursory look at plots of several asset return series immediately reveals the strong volatility clustering, or lack of independence between observations, which is reinforced by inspection of sample autocorrelation functions of absolute and squared returns. Since, rather obviously, iid stable Paretian observations will not exhibit such behavior, a standard GARCH model, driven even by normal innovations, will perform much better with respect to almost any goodness-of-fit test and lead to rejection of the stable hypothesis.

The GARCH and stable Paretian hypotheses are not, however, mutually exclusive: a rather natural consideration would be to specify GARCH models driven by innovations from the stable class, as has been done by McCulloch (1985), Liu and Brorsen (1995), Panorska et al. (1995), and Mittnik et al. (1999c).

GK have, however, brought to light another, more interesting, point by empirically demonstrating that data generated by IGARCH models with non-normal innovations also exhibit stability under summation or, in short, SuS. That is, when summing the observations of consecutive non-overlapping subsamples of length $S$ of simulated series, the estimate of the tail index appropriate for stable Paretian data does not move towards two as $S$ increases, but remains approximately constant. The SuS property is frequently viewed as supporting the stable Paretian assumption. Their findings indicate that the tail behavior of IGARCH and iid-stable processes possess some similar characteristics. In addition, as emphasized in
GK, this shows via a concrete and popular model that although stable Paretian data imply summability, the converse is not true. Passing the SuS criterion is, therefore, a necessary, but not sufficient condition for the data being stable Paretian.

By allowing for a mechanism to account for the clustering effect, fair comparisons could be made to assess the extent to which the innovations process is adequately described by a stable law. In particular, while the results of SuS tests performed on serially dependent data might serve exploratory data analysis purposes, SuS tests on model residuals are in order if the objective is to test for stable Paretian innovations. SuS based rejections of the stable Paretian hypothesis for asset returns could be the consequence of either non-Paretianity or serial dependence or, in fact, both. This draws into question the results of various previous studies, which, based strictly on SuS tests of unconditional return data, rejected the stable hypothesis see Paolella, 2000, and the references therein. Conversely, the finding of GK that SuS cannot be rejected for data generated from an IGARCH process with non-normal innovations (precisely the model appropriate for the majority of financial time series—see, e.g., Baillie and Bollerslev, 1989; Hsieh, 1989; Lumsdaine, 1995)—also implies that those studies, which, based on SuS, do not reject the stable hypothesis, could be misleading.

One of the major objections to the use of either the conditional or unconditional non-normal stable Paretian distribution in this context is that it has infinite variance; this contradicts alleged empirical evidence suggesting the existence of third or higher moments in various financial return data (cf. Loretan and Phillips, 1994; and Pagan, 1996). Since these findings were almost exclusively arrived at by use of the Hill (1975) or related tail estimators, which are known to be horrendously unreliable in-even large-finite samples (cf. Mittnik and Rachev, 1993b; Kratz and Resnick, 1996; Adler, 1997; McCulloch, 1997b; Resnick, 1997; Mittnik et al., 1998a; Paolella, 2000) and even worse for data with GARCH structures (Kearns and Pagan, 1997), they need to be heavily discounted and, hence, do not provide an answer to the open question of the maximally existing moment of financial return data.\footnote{In fact, Adler (1997) went so far as to say that “[m]any of the problems faced by the Hill and related estimators of the tail decay parameter $\alpha$ can be overcome if one is prepared to adopt a more parametric model and assume, e.g., stable innovations” and concluded by noting that “[o]verall, it seems that the time may have come to relegate Hill-like estimators to the *Annals of Not-Terribly-Useful Ideas*. Further flaws inherent in insisting upon finite-variance models for data modeling are discussed in Nolan (in press).}

In fact, empirical evidence presented in Mittnik et al. (1999b) suggests that stable-GARCH models outperform their finite-variance competitors with respect to prediction of down-side market risk. Indeed, modern risk management, particularly in conjunction with higher frequency data now easily available, are shifting from classic mean-variance techniques to more modern methods such as extreme-
value theory and stochastic volatility, which do not, a-priori, rely on the existence of second moments. See also McCulloch (1997a) and Rachev and Mittnik (2000) for further discussion.

This paper makes two contributions. We extend previously proposed stable-GARCH models to incorporate more general GARCH structures driven by stable Pareto innovations and use these in conjunction with SuS tests performed on estimated model residuals to appropriately test for the stable Pareto hypothesis. By focusing on the properties of the innovations driving GARCH processes, namely heavy-tailedness and summability, our work differs from related work such as de Vries (1991), Davis and Mikosch (1998), and Mikosch and Starica (2000), who investigate properties of the observed output variable of GARCH-type processes driven by finite-variance innovations.

The remainder of this paper is organized as follows. Section 2 details the stable-GARCH model including two new extensions and provides simulation results on the small-sample behavior of the maximum likelihood estimator for the parameters. Section 3 examines the GARCH SuS issue raised by GK in more detail and demonstrates analogous findings for stable-GARCH models. By applying a formal test for stable Pareto variates, Section 4 provides empirical evidence that the estimated GARCH-filtered innovations driving East Asian currency-return series comply with the stable law. Section 5 contains the conclusion.

2. Stable-GARCH processes

We first introduce the GARCH process with stable Pareto innovations and briefly discuss the required modification of the sufficient condition for stationarity detailed in Mittnik et al. (1999c). The model is then generalized to support more general GARCH structures, which have been recently shown to provide a significant advantage in terms of model fit and forecasting. Simulation results are then shown, demonstrating the acceptable small sample properties of the estimated parameters of stable-GARCH models.

2.1. Standard stable-GARCH process

We refer to sequence \( y_t \) as an (asymmetric) stable Pareto power-GARCH process or, in short, an \( S_{\alpha, \beta} \) GARCH\((r, s)\) process, if

\[
y_t = \mu_t + \epsilon_t,
\]

where \( \epsilon_t = \sigma_t z_t, z_t \sim S_{\alpha, \beta}(0, 1) \) and:

\[
\sigma_t^\beta = \sigma_0 + \sum_{i=1}^{r} c_i |\epsilon_{t-i}|^\beta + \sum_{j=1}^{s} d_j \sigma_{t-j}^\beta,
\]

...
with $S_{\alpha,\beta}(0,1)$ denoting the standard asymmetric stable Paretian distribution with stable index $0 < \alpha \leq 2$, skewness parameter $-1 \leq \beta \leq 1$, zero location parameter, and unit scale parameter. Recall that for $\alpha < 2$, $z_j$ does not possess moments of order $\alpha$ or higher, and that for $\alpha = 2$, $z_j$ is normally distributed. By letting location parameter $\mu$, in Eq. (1) be time-varying, a range of mean equations, including regression and/or ARMA structures, is possible.

Mittnik et al. (1999c) show that the $S_{\alpha,\beta}^\delta$ GARCH($r, s$) process with $1 < \alpha \leq 2$ and $0 < \delta < \alpha$ has a unique strictly stationary solution if $c_0 > 0$, $c_i \geq 0$, $i = 1, \ldots, r$, $r \geq 1$, $d_j \geq 0$, $j = 1, \ldots, s$, $s \geq 0$, and

$$V = \mathbb{E}|z_j|^\delta \sum_{i=1}^{r} c_i + \sum_{j=1}^{s} d_j \leq 1$$

holds. For $1 < \alpha \leq 2$ and $0 < \delta < \alpha$, quantity $\mathbb{E}|z_j|^\delta$ can be expressed in closed form as

$$\mathbb{E}|z_j|^\delta = \frac{1}{\psi_\delta} \Gamma\left(1 - \frac{\delta}{\alpha}\right)\left(1 + \tau_{\alpha,\beta}^2\right)^{\frac{\delta}{2\alpha}} \cos\left(\frac{\delta}{\alpha} \arctan(\tau_{\alpha,\beta})\right),$$

where $\tau_{\alpha,\beta} := \beta \tan(\alpha \pi/2)$ and

$$\psi_\delta = \begin{cases} \Gamma(1 - \delta) \cos\left(\frac{\pi \delta}{2}\right), & \text{if } \delta \neq 1, \\ \pi/2, & \text{if } \delta = 1. \end{cases}$$

As $\delta$ approaches $\alpha$, $\mathbb{E}|z_j|^\delta$ increases without bound and, in the limit, leads to explosive behavior uncommon in practice.

This implies that for integrated $S_{\alpha,\beta}^\delta$ GARCH($r, s$) processes, denoted as $S_{\alpha,\beta}^\delta$ IGARCH($r, s$), the usual “sum-to-one” condition, i.e., does not hold. In fact, this is true for GARCH processes driven by any non-normal innovations process for which $\mathbb{E}|z_j|^\delta \neq 1$.\(^2\)

Inspection of generated $S_{\alpha,\beta}^\delta$ GARCH($r, s$) processes, i.e., with $\delta = \alpha$, as assumed in Liu and Broersen (1995), suggests that they are non-stationary: they exhibit extreme volatility clustering and eventually increase without bound. This behavior is highly dependent on the closeness to the IGARCH border implied by Eq. (3), with simulated IGARCH processes “exploding” for very small $T$ while models with volatility persistence parameter $\overline{V} \leq 0.9$ often appear stationary for $T > 5000$. Thus, estimated $S_{\alpha,\beta}^\delta$ GARCH($r, s$) models with $\overline{V} \approx 1$ and $\delta \approx \alpha$.

\(^2\) For example, an IGARCH(1,1) process driven by Student’s $t$ innovations with $\nu$ degrees of freedom and $\delta = 2$ would set $d_1 = 1 - c_1 \nu/\nu$ for $0 < c_1 < (\nu - 2)/\nu$. 

could be the sign of model misspecification. Modeling returns on several foreign exchange rates, Mittnik et al. (1999b) show that in all cases considered, the resulting models were stationary and $\hat{d}$ was significantly less than $\hat{a}$.

By juxtaposing graphs of iid stable Paretoian data with those generated from a stationary normal-GARCH model, GK made the point that particularly when $c_1 + d_1$ is far from the IGARCH border, the two processes exhibit similar characteristics and cannot be easily distinguished from one another. In Fig. 1, we also show simulations from such processes, but accompany them with the sample autocorrelation function (SACF) of the absolute values. For neither process is the small-sample behavior of the SACF well known, although it is evident that between the two processes, the SACFs behave quite differently.

Fig. 2 plots realizations from six stationary $S_{\alpha,\beta}^d$ GARCH(1, 1) processes for different $\alpha$ and $\beta$ combinations, using the same seed value. Their SACF plots (not shown) were relatively similar and show unequivocally that the absolute data possess strong correlation and are, thus, far from being iid. As $\alpha$ decreases, the relative severity of the large shocks increase, while with asymmetric (left-skewed) shocks, that of the negative innovations increase, the magnitude of which is dependent on $\alpha$ (as $\alpha$ approaches 2, $\beta$ loses its effect).

2.2. Extending the GARCH-stable process

As mentioned in the Introduction, the use of the summability property to effectively test the stable Paretoian hypothesis in finite samples requires the data to be independent. Certainly, the residuals from an ARMA-GARCH filter applied to financial return data will be much closer to iid than their unfiltered counterparts. Nevertheless, few would insist that observed data series are really generated by an ARMA-GARCH process, let alone from the particular handful of models chosen to be fitted by the researcher. Thus, SuS tests may be jeopardized by the extent to which the filtered residuals deviate from iid-ness. Although attaining genuine iid residuals is an ideal which, at least for financial data with their inherently complex data generating mechanisms, will hardly be reached, it is essential that one uses models that filter the data as effectively as possible.

To this end, we entertain some of the numerous extensions of the standard GARCH equation which have succeeded in adding only a small number of parameters to the model, but which are highly significant and easily favored by even the harshest of penalty measures. Two such models, both of which nest a variety of previously proposed extensions, have proven very successful with regard to in-sample and out-of-sample fit. These are the Asymmetric Power ARCH, or A-PARCH model, proposed by Ding et al. (1993) and the Generalized

\footnote{For correlated stable data, see, however, Brockwell and Davis (1991), Adler et al. (1998) and Davis and Mikosch (1998); while, for normal-GARCH data, Diebold (1986) and Bera et al. (1992).}
Fig. 1. Left upper panel is iid $S_{i.i.d.}$ data; right upper is GARCH with $\alpha = 0.2$, $\beta = 0.6$ and iid normal innovations. Bottom panels are their respective SACFs of the squared data.
Fig. 2. $S^1_{\alpha, \beta}$ GARCH(1,1) processes with $c_0 = 0.01$, $c_1 = 0.2$ and $d_1$ chosen such that $V = 0.95$. From left to right, $\alpha = 2.0, 1.8, 1.4$; top (bottom) is for $\beta = 0$ ($\beta = -0.9$).
Quadratic ARCH, or Q-GARCH model, introduced by Sentana (1991, 1995). Both models allow current conditional volatility to react asymmetrically to the size of the previous periods’ innovations, which is necessary to capture the well-known leverage effect as originally noted by Black (1976) (see also Christie, 1982; French et al., 1987; Pagan and Schwert, 1990). Whether or not the leverage effect is present, the additional parameters required to incorporate these asymmetries are quite often found to be highly significant for a variety of data sets (see Ding et al., 1993; Franses and van Dijk, 1996; Lane et al., 1996; Paolella, 1997).

The A-PARCH(r, s) model is given by

\[
\sigma_t^g = c_0 + \sum_{i=1}^{r} c_i \left( |\epsilon_{t-i} - \gamma_i \epsilon_{t-i}| \right)^g + \sum_{i=1}^{s} d_i \sigma_{t-i}^g,
\]

for \(|\gamma_i| < 1\) with the special case \(g = 2\), \(\gamma = 0\) corresponding to the standard GARCH model. Negative values of \(\gamma_i\) result in relatively higher values of \(\sigma_t\) when \(z_{t-i}\) is negative. By assuming the existence of a stationary solution to Eq. (6), the unconditional expectation of \(\sigma^g\) is \(c_0 = (1 - V_g)\), where

\[
V_g = \sum_{i=1}^{r} \kappa_i c_i + \sum_{j=1}^{s} d_j
\]

is a generalization of \(V\) specified in Eq. (3) and

\[
\kappa_i := E\left(\left| z_{t-i} - \gamma_i z_{t-i} \right|^g \right) = E\left(|z| - \gamma_i z \right)^g
\]

depends on the distributional specification of \(z_i\) and exists only for values of \(g > 0\) for which \(E(z^g) < \infty\). For example, with Student’s \(t\) innovations with \(\nu\) degrees of freedom, elementary integration yields

\[
\kappa_i = \nu^2 \frac{1}{2\nu} \left( (1 + \gamma_i)^g + (1 - \gamma_i)^g \right) \Gamma\left( \frac{\delta + 1}{2} \right) \Gamma^{-1}\left( \frac{\nu}{2} \right) \Gamma\left( \frac{\nu - \delta}{2} \right).
\]

which exists when \(\delta < \nu\). For \(S_{0,1}\) innovations for which \(1 < \alpha < 2\), constraint \(\delta < \alpha\) is required for \(\kappa_i\) to exist. For this latter model, say \(S_{\alpha,\beta}\)

A-PARCH(r, s) no closed-form expression for \(\kappa_i\) is known, leaving numerical integration the only alternative. This is easier said than done.

---

It is worth mentioning that neither model nests the well-known Exponential GARCH (EGARCH) from Nelson (1991), as also considered in GK, nor its generalizations proposed by Hentschel (1995). Although GK simulated from this model, they did not estimate it, which is known to be numerically demanding. In particular, Franses and van Dijk (1996) noted that “…only with starting-values with a precision of eight digits could we obtain convergence”, while Kaiser (1996) observed “…severe estimation problems” associated with the EGARCH model. Similar findings have been reported by Frachot (1995) as well. For this reason, we do not entertain the EGARCH model further.
In particular, recalling the thickness of the stable Pareto tails, which are then compounded with the factor \((|z_{t-1}| - \gamma_{t-1})^\delta\), the integration needs to be performed over a very large range, with accurate values of the pdf far into the tails. We accomplish this by using the pdf of Doganoglu and Mittnik (1998), which couples a polynomial approximation to the mode-centered asymmetric stable density over \([-10, 10]\) with a series-approximation for the tails. They show that for \(\alpha > 1.2\), the approximation delivers a maximum relative inaccuracy of \(10^{-6}\) over the whole real line. This is accurate (and fast) enough to be used in conjunction with numerical integration routines to approximate the \(\kappa_t\). The restriction that \(\alpha > 1.2\) is of little importance in this context, since the estimated value of tail index \(\alpha\) for conditional GARCH residuals rarely dips below 1.5, even for highly volatile data (Liu and Borsen, 1995; Mittnik et al., 1998b,c, 1999b).

The Q-GARCH\((r, s)\) model is given by

\[
\sigma_t^2 = c_0 + \sum_{i=1}^{r} c_i \epsilon_{t-i}^2 + \sum_{i=1}^{r} c_{ij} \epsilon_{t-i} \epsilon_{t-j} + 2 \sum_{i=1}^{r} \sum_{j=i+1}^{r} c_{ij} \epsilon_{t-i} \epsilon_{t-j} + \sum_{j=1}^{s} d_j \sigma_{t-j}^2. \tag{9}
\]

With \(s = 0\), it has the natural interpretation as being a second order approximation to the unknown conditional volatility function and can be shown to be the most general quadratic version of ARCH-type models. An obvious generalization is to relax the exponent in Eq. (9), i.e.,

\[
\sigma_t^\delta = c_0 + \sum_{i=1}^{r} c_i \epsilon_{t-i}^\delta + \sum_{i=1}^{r} c_{ij} \epsilon_{t-i} \epsilon_{t-j}^\delta + 2 \sum_{i=1}^{r} \sum_{j=i+1}^{r} c_{ij} \epsilon_{t-i} \epsilon_{t-j}^\delta + \sum_{j=1}^{s} d_j \sigma_{t-j}^\delta. \tag{10}
\]

This model, denoted \(S_{\alpha,\beta}^\delta\) Q-GARCH\((r, s)\) will be useful in cases where the second moment of \(\epsilon_t\) does not exist, e.g., with non-normal stable Pareto innovations. Taking iterated expectations shows that \(E(\epsilon_{t-i} \epsilon_{t-j}) = 0\) for \(i \neq j\), so that \(E(\sigma_t^\delta) = c_0/(1 - V_0)\), where

\[
V_0 = \sum_{i=1}^{r} c_i E[|\epsilon_{t-i}|^\delta] + \sum_{j=1}^{s} d_j. \tag{11}
\]

Note that \(E[|\epsilon_{t-i}|^\delta]\) in Eq. (11) is just the \(\kappa\) defined in Eq. (8) with \(\gamma_t = 0\), for which the closed-form expression Eq. (4) can be used when \(z_t \sim S_{\alpha,\beta}^\delta\), \(0 < \delta < \alpha\) and \(1 < \alpha \leq 2\).

2.3. Small sample behavior of point estimates

The estimation (and even simulation) of joint GARCH-stable models is, however, not straightforward and has been only very recently considered. In the
unconditional case, where only the four parameters (tail-thickness, $\alpha$; skewness, $\beta$; and the location and scale parameters) need to be estimated, fast and reliable methods have been in use for quite some time, the most successful of these being the quantile-based estimator of McCulloch (1986). In the context of more general conditional models, however, this estimator is inapplicable and one would rather rely on the maximum likelihood (ML) estimator (see, e.g., Nolan, 2000, and the references therein). The greatest hindrance for ML estimation involves the complex calculation of the stable Paretoian density function $pdf$, which is not available in closed form. For symmetric pdfs, McCulloch (1998) has provided an accurate polynomial approximation; while Doganoglu and Mittnik (1998) and Mittnik et al. (1999a) pursue fast and reliable methods for the general, asymmetric case.

To gain some understanding of the small-sample performance of the model, we simulate 100 replications of an $S^1_{2,0}$-PARCH(1, 1) process for sample size $T = 5000$, with parameter values $c_0 = 0.01$, $c_1 = 0.2$, $\gamma_1 = -0.5$ and $d_1 = 0.95 - \kappa_1 c_1$, i.e., that value $d_1$ such that $V_\gamma = 0.95$ and $\kappa_1$ depends on the chosen value of $\alpha$.\footnote{We restrict $\delta = 1$, since, as $E(z) = 0$ for $\alpha > 1$, it follows that $E(|z| - \gamma; z) = E|z|$, which is available in closed form and, thus, avoids the otherwise necessary numerical integration. In addition, we use the scaled innovations $S_{\alpha, \beta}/\sqrt{2}$, which has the benefit that the $\alpha = 2$ case coincides with $N(0, 1)$ instead of $N(0, 2)$. It is simple to show that the corresponding parameter values for $c_1$ and $\gamma_1$ are remarkably similar, while, as $\alpha$ moves towards two, the estimate of $c_0$ exhibits a very slight increase in variance; that of $V_\gamma$ tends to exhibit fatter tails (more mass in the center and more outliers); and the variance of $V_\beta$ increases greatly. For the latter, this is to be expected, recalling that the effect of $\beta$ diminishes as $\alpha$ increases. Finally, regarding $\delta$, for smaller values of the actual $\alpha$ (1.5 and 1.6), the resulting small-sample distribution is thin-tailed; but as the true value in-}

Fig. 3 and the upper plots in Fig. 4 show boxplots of the estimated parameters using actual tail-index values $\alpha = 1.5$, 1.6, ..., 2.0.\footnote{The quasi-Newton BFGS algorithm as implemented in Matlab is used for estimation, with starting values arbitrarily chosen as values typical for financial time-series. On a 200 MHz Pentium PC, most models are estimated in less than 30 s. Further simulation results as well as estimation with actual data from a variety of sources (Mittnik et al., 1999b) demonstrate that in the vast majority of cases, use of different (and oftentimes quite poor) starting values leads, without difficulty, to the same log-likelihood maximum.} With one exception, the estimates appear virtually unbiased with a symmetric distribution not dominated by outliers. This does not hold for the estimate of $\alpha$ when $\alpha = 2$, which, owing to the necessary constraint that $1 < \hat{\alpha} \leq 2$, is highly skewed. In this case, however, it is interesting to note how close the estimates are to the correct value of 2.0. All 100 values were greater than 1.99, with 90 being greater than 1.996. With respect to the different $\alpha$-values, each of the small-sample distributions of $\hat{\beta}$, $\hat{c}_1$, $\hat{d}_1$ and $\hat{\gamma}_1$ are remarkably similar, while, as $\alpha$ moves towards two, that of $\hat{c}_0$ exhibits a very slight increase in variance; that of $V_\gamma$ tends to exhibit fatter tails (more mass in the center and more outliers); and the variance of $\hat{\beta}$ increases greatly. For the latter, this is to be expected, recalling that the effect of $\beta$ diminishes as $\alpha$ increases. Finally, regarding $\delta$, for smaller values of the actual $\alpha$ (1.5 and 1.6), the resulting small-sample distribution is thin-tailed; but as the true value in-
Fig. 3. Estimated parameters from 100 simulated $S_{\alpha}^\sigma$-A-PARCH(1, 1) processes for $\alpha = 1.5, 1.6, \ldots, 2.0$ (x-axis) with $S_{\alpha}/\sqrt{2}$ innovations, $c_0 = 0.01$, $\epsilon_1 = 0.2$ and $d_1$ chosen such that $V_\gamma = 0.95$. Boxes have lines at the 25%, 50% and 75% quantiles.

creases, it becomes very fat-tailed and, as $\alpha$ approaches two, very skewed. Somewhat oddly, the behavior at $\alpha = 1.7$ and 1.8 is not quite consistent with the general pattern. It should be, nevertheless, emphasized that for all replications and all $\alpha$-levels considered, $\hat{\alpha}$ was 65% and 98.5% of the time within 0.01 and 0.05 of the true value, respectively.

Finally, the lower half of Fig. 4 shows the small-sample performance of two useful goodness-of-fit criteria based on the empirical distribution function of the
residuals; the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) statistics, defined by

$$ KS = 100 \times \sup_{x \in \mathbb{R}} |F_x(x) - \hat{F}(x)| \quad \text{and} \quad AD = \sup_{x \in \mathbb{R}} \frac{|F_x(x) - \hat{F}(x)|}{\sqrt{\hat{F}(x)(1-\hat{F}(x))}}. $$

(12)

respectively, where $\hat{F}(x)$ and $F_x(x)$ are the cdf of the estimated parametric density and the empirical sample distribution

$$ F_x(x) = \frac{1}{T} \sum_{t=1}^{T} \mathcal{J}_{-x, x} \left( \frac{r_t - \hat{\mu}_t}{\hat{\sigma}_t} \right), $$

respectively, with $\mathcal{J}(\cdot)$ denoting the usual indicator function. The AD statistic differs from KS in that it is given by the deviation between empirical and fitted cdf standardized by the standard deviation of $F_x(x)$.

In the empirical application in Section 4, we will apply the AD statistic as a goodness-of-fit measure for the estimated models. Since the AD statistic weights discrepancies more appropriately across the whole support of the distribution, most
notably so in the tails, it is particularly suitable for assessing the conditional probability of large investment losses, where one focuses on accurate inference from the left tail of the conditional return distribution.

From Fig. 4, we see that the statistics’ performance is virtually independent of the true $\alpha$, with the median KS value falling almost precisely at one, while that of AD at 0.03. Although it is comforting to know that these measures are not dependent on $\alpha$ (over the range considered and for the given model), it should be kept in mind that this is of only limited practical interest, since one is usually concerned not with a test of model adequacy based on the observed KS or AD value per se (for which small sample results are lacking), but rather their comparison across models. Nevertheless, by detailing the finite-sample performance of these measures, such a simulation exercise can help decide if two or more competing values of KS or AD are significantly different from one another.

2.4. Small sample behavior of asymptotic variance estimates

Although there is no formal proof, we conjecture that—if the parameter vector belongs to the interior of the admissible parameter space, and other suitable regularity conditions hold—the ML estimator is consistent and asymptotically normal. DuMouchel (1973, p. 952ff) shows that this holds for the iid case with $\alpha \geq \epsilon > 0$, $\epsilon$ arbitrarily small. The boxplots previously discussed show that except for $\hat{\alpha}$ and, as $\alpha$ approaches two, $\hat{\beta}$, the small-sample distributions of the parameters, including $\hat{V}$, are thin-tailed and essentially symmetric, so that normality is not an unrealistic assumption. For this reason, we concentrate on the quality of the approximate standard errors obtained by inverting the Hessian matrix obtained from the quasi-Newton numerical optimization, as compared with the empirical variance observed from the Monte Carlo simulations.

Inference on the persistence parameter $V$ is greatly facilitated if the asymptotic distribution of $\hat{V}$ is known and can be reliably calculated. Provided that the ML estimator $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}_0, \ldots, \hat{\gamma}_c, \hat{\delta}_1, \ldots, \hat{\delta}_d, \hat{\delta})$ is asymptotically normally distributed with mean $\theta$ and covariance matrix given by the inverse of the information matrix, $I_{\theta}$, then $\hat{V}_{V}^{\text{asymp}} \approx N(\tau V_{V}^{\tau} I_{\theta}^{-1} \tau')$, where $\tau = (\partial V_{V}/\partial \theta)$. As closed-form expressions for $\kappa$ do not exist when $\delta \neq 1$ and $\gamma_i \neq 0$, this method cannot be used for the general $A$-PARCH model.

Clearly, asymptotic normality will fail as we approach the boundary of the parameter space where the distribution becomes degenerate. In the likelihood maximization, we appropriately constrain the parameter space to ensure numerical stability. Note that $\hat{V}$ slightly exceeded unity in three out of eight cases (see Table 1 below), which is the lower bound for stationarity implied by the sufficient—but not necessarily necessary—condition stated in Section 2. To what extent this is of consequence for theoretical or applied work needs to be investigated further.
When $\delta = 1$, the $\gamma_i$ drop out, i.e., $\partial V/\partial \gamma_i = 0$, while $\partial V/\partial c_i = \kappa(\alpha, \beta) = E[Z]$, $i = 1, \ldots, r$. Clearly, $\partial V/\partial d_j = 1$, $j = 1, \ldots, s$, $\partial V/\partial \mu = \partial V/\partial c_0 = 0$ and $\partial V/\partial \beta = \sum_{i=1}^r c_i \partial \kappa(\alpha, \beta)/\partial \beta$, where, with $m = \alpha\pi/2$,

$$
\frac{\partial \kappa(\alpha, \beta)}{\partial \beta} = m^{-1}(-1)^{(1/2\alpha)}\left[(\beta^2 - 1)\cos^2(m) - \beta^2\right]^{1/2\alpha} - 1 \cos m \left[\beta \cos(\alpha^{-1}\tan \phi)(\cos^2(m) - 1)
+ \cos(m)\sin(m)\sin(\alpha^{-1}\tan \phi)\right],
$$

which is real for $1 < \alpha \leq 2$ and $|\beta|$ and, with $\phi = \beta \tan(\alpha\pi/2)$, $\partial \kappa/\partial \beta$ is equal to zero for $\beta = 0$. Similarly, $\partial V/\partial \alpha = \sum_{i=1}^r c_i \partial \kappa(\alpha, \beta)/\partial \alpha$ but the latter derivative does not admit an amenable closed-form expression. As such, we directly approximate the derivative numerically, which appears quite reliable. Fig. 5 plots $\partial \kappa(\alpha, \beta)/\partial \alpha$ and $\partial \kappa(\alpha, \beta)/\partial \beta$ over a portion of the parameter space.

For the $S_d^{\beta \gamma}$ Q-GARCH$(r, s)$ model, $\text{Var}(\hat{V}_t)$ can also be readily approximated by $\tau \Gamma_{-\gamma^2}^{1/\tau}$, $\tau = (\partial V/\partial \gamma)$, where values $\partial \kappa(\alpha, \beta, \delta)/\partial \alpha$, $\partial \kappa(\alpha, \beta, \delta)/\partial \beta$, and $\partial \kappa(\alpha, \beta, \delta)/\partial \delta$ need to be numerically approximated; here, $\kappa(\alpha, \beta, \delta)$ is given in Eq. (8), in which $\gamma_I = 0$ and $z$ denotes a $S_d^{\beta \gamma}(0,1)$ random variable.

Fig. 6 is similar to Fig. 3 but shows boxplots of the 100 approximate standard errors, which accompany the estimated model parameters returned from the ML estimation procedure, along with the calculated value of $\text{SE}(\hat{V}_t)$. For each parameter and value of $\alpha$, the inscribed circle indicates the empirical standard

---

To prevent the boxplots from being too small, we removed several of the largest occurring values. The numbers removed from each boxplot are given in the square brackets of the upper caption of each plot. On average, they amount to about 4 out of the 100 observations.
Fig. 6. Estimated standard errors for the parameters from the 100 simulated $S^{\alpha/2}_t$-A-PARCH(1, 1) processes. The circle indicates the empirical standard deviation of the 100 estimates. The numbers in brackets on top of each plot indicate the number of removed values to enhance readability (see footnote 8).

deviation computed from the 100 estimates, and serves as the “true” standard deviation appropriate for this model and sample size. The boxplots show that the approximate standard errors are quite accurate for $\alpha < 2$, while, for $\alpha$ approaching 2, they tend to underestimate the true variability. All variance estimates are, however, themselves quite fat-tailed and, due to the occasional occurrence of values larger than the depicted range shows, are skewed upwards, i.e., except for the $\alpha = 2$ case, have a slight tendency to overestimate the true variability.
3. Empirical examination of stability under summability

It is a remarkable feature of simulated realizations from near-IGARCH processes driven by fat-tailed innovations that they tend to exhibit SuS. GK demonstrated this property by estimating the tail index $\alpha$ of simulated GARCH and IGARCH observations for several values of $S$, the number of consecutive non-overlapping summed observations. In their subsequent note, Groenendijk et al. (1995) provided some theoretical insight explaining this phenomenon. It should be noted that, rather than using a general tail estimator, GK adopted that of McCulloch (1986), which is specifically designed to estimate the tail index for (iid) stable Paretian data.

Specifically, GK make the following observations. First, for a normal-GARCH process, the closer the model comes to being IGARCH (i.e., the closer $c_1 + d_1$ is to one), the higher the value of $S$ has to be before the majority of estimated stable tail index values reach two. Second, they find that not only does the sum $c_1 + d_1$ influence the convergence towards two, but also the values themselves, with, for a given value of sum $c_1 + d_1$, larger values of $c_1$ corresponding to slower convergence (i.e., higher $S$). Finally, using Student’s $t$ innovations with $5\, df$, in short, $t(5)$, they show that these results still hold, but convergence is, compared to normal innovations, much slower. In fact, for $t(5)$-IGARCH models, evidence from their Table 2 and their further Monte Carlo runs suggests that the estimated stable tail index does not converge at all as $S$ increases. It seems safe to say that this is not a phenomenon of just the $t(5)$ distribution, but rather of general violation of normality, in particular, a “fattening of the tails”.

We now consider to what extent such results hold for the stable Paretian A-PARCH model using parameter values similar to ones estimated using real financial data. In particular, for each of the 24 combinations resulting from using $c_1 = 0.1, 0.3, \gamma = 0, -0.5, 0.5, d_1$ such that $V_y = 0.95, 0.99$, and $\alpha = 1.6, 1.8, 2$ (with constant values $c_0 = 0.01, \beta = 0$ and $\delta = 1$), we generate $S_{\alpha, \beta}^{A-PARCH(1,1)}$ time series, each of length 5000. For each of these series, the stable tail index is estimated. Fig. 7 shows, separate for each of the 24 parameter constellations, the means of the estimated stable tail index for the ranges $S = 1, 2, \ldots, 50$ computed over the 100 simulated series. In comparison to GK, who used the McCulloch (1986) estimator, we employ a modified Hill estimator (Paolella, 1997; Mittnik and Paolella, 1999), which is also designed specifically for the tail index of iid stable Paretian observations, but possesses better small-sample properties than the McCulloch estimator (see Appendix A).

Recall that, while the innovations driving the simulated A-PARCH processes are iid stable Paretian, the observed output series are clearly not, so that the estimated stable tail index should not be expected to recover the tail index used to generate the innovations process. We would instead expect such an estimate to (highly) underestimate the actual tail index of the innovations, given that the GARCH effects induce a fattening of the tails in the output process. Initial
inspection of Fig. 7 immediately verifies that (i) the unconditional estimate of $\alpha$ is indeed far below the true value; and (ii) the SuS phenomenon holds quite strongly for $S_{\alpha}^g$-A-PARCH(1, 1) processes, with a near-linear downward shift accompanying a decrease in $\alpha$. Notice further that for given values of $V_\gamma$, $\gamma$ and $\alpha$, the estimated stable tail index using $c_1 = 0.1$ is much higher than that for $c_1 = 0.3$, which, for $\alpha = 2$, agrees with the findings of GK for their normal-IGARCH models.

The main difference between the plots corresponding to $V_\gamma = 0.95$ from those of $V_\gamma = 0.99$ is a downward shift, i.e., all things being equal, the closer $V_\gamma$ is to
one, the lower the estimated tail index is. In addition, the variability of the estimated tail index increases with increasing $V_s = 0.99$ as well as increasing $c_1$.

The effect of decreasing is demonstrated by comparing the left and right panels; all other factors being equal, the estimated tail index undergoes a downward shift, which parallels the finding of GK using EGARCH models with non-zero asymmetry parameter. Note that this shift is not linear in $\alpha$, with the drop being far more pronounced for large $\alpha$-values (near two), and minimal for smaller values of $\alpha$.

4. An application to East Asian currencies

We now examine whether returns on East Asian currencies can be modeled by the aforementioned stable-GARCH modeled and whether the estimated innovations are adequately described by the stable law. During the Asian financial crises in the late 1990s, the exchange rates of many Asian countries experienced violent fluctuations, rendering the resulting daily return series particularly recalcitrant. The empirical results for one currency, the Thailand bhat, will be discussed in some detail. Results for other series, namely the Singapore dollar, the South Korean won and the Sri Lanka rupee (all against the US dollar) are briefly summarized toward the end of this section.

The Thai bhat is particularly interesting because its massive devaluation in July 1997 is considered to have triggered the financial crisis that led to the collapse of foreign exchange and equity prices in many East Asian countries. During the 1980s, Thailand’s capital markets became increasingly liberalized and opened up for foreign investors. As in other countries in the region, the Thailand economy experienced high economic growth and large inflows of private capital throughout the 1990s, while maintaining an effectively pegged exchange nominal rate. The capital inflow led to an appreciation of the real exchange rate, which, in turn, negatively affected firms’ exports and profit margins and, ultimately, led to cash flow shortages. From 1990 to 1997, Thai companies increased the US-denominated foreign debt outstanding from international bond market activities by a factor of about 10, to US$12.9 billion (cf. Harvey and Roper, 1999). In February 1997, Somprasong became the first Thai company which failed to serve foreign debt. In May of that year, the Thai bhat came under heavy attacks from speculators acting on the economic slow down as well as political instability. The largest finance company, Finance One, failed shortly thereafter. In June, Thailand’s finance minister, who had strongly resisted a devaluation of the bhat, resigned. On July 2, the Bank of Thailand announced a managed float of the bhat and requested “technical assistance” from the International Monetary Fund. The resulting deval-

---

9 For a more detailed account of the Asian financial arises, see, e.g., Corsetti et al. (1999a,b) and Harvey and Roper (1999).
The 2205 daily returns of the Thai bhat–US dollar exchange rate from the period January 1, 1990 to December 31, 1998 are illustrated in the top panel of Fig. 8. Visibly striking is the massive increase in volatility occurring at the onset of the crisis. Although at first sight, it might appear that separate models for the two distinct periods would be necessary (or the use of more complicated models, which incorporate regime switching, structural breaks, etc.), we find that the proposed class of stable-GARCH models are adequate. In particular, we fit an $S^a_{\alpha,\beta}$ A-PARCH(1, 1) and an $S^a_{\alpha,\beta}$ Q-GARCH(2, 1) model, each with an MA(1)
component for the mean. Summary statistics are provided in Table 1 and include the estimated tail index and skewness parameter of the innovations process, $\hat{\alpha}$, and $\hat{\beta}$, respectively, along with approximate standard errors, $\hat{SE}$. Similarly, $\hat{V}$ stems from either Eqs. (7) or (11), with calculation of $\hat{SE}$ as discussed in Section 2.3. The corrected AIC (AICC) criterion:

$$AICC = -2\hat{\mathcal{L}} + \frac{2T(k + 1)}{T - k - 2},$$

where $k$ denotes the number of estimated parameters and $T$ the number of observations and the AD statistic (Eq. (12)) are reported as comparative goodness-of-fit measures between the two fitted models.

Given the considerably lower value of the AICC statistic for the Q-GARCH specification, we restrict discussion just to the results of this model. This yielded a stable tail index of $\hat{\alpha} = 1.65$ and persistence parameter $\hat{V} = 0.996$. The residuals from this model are plotted in the middle panel of Fig. 8; they show no sign of a structural break at the crisis onset and appear to be consistent with iid stable realizations with $\hat{\alpha} = 1.65$. Correlogram graphs (not shown) for the residuals, absolute residuals and squared residuals do not provide any evidence against the independence hypothesis.

We are now interested in testing whether these residuals are adequately described by a stable law. Under the null hypothesis, the $\alpha$-estimates of sums of iid stable variates should be invariant with respect to the number of summands, $S$, while, for non-stable data, the index should increase towards two as $S$ increases. A test based on this property has been proposed in Paolella (2000) and shown to have not only appropriate nominal size, but also reasonable power against both Student's $t$ and mixed normal alternatives. It is based on the linear increase of $\hat{\alpha}$ vs. $S$, where $\hat{\alpha}$ is calculated using the Hill-intercept tail estimator (see Appendix A for details). The lower panel of Fig. 8 plots $\hat{\alpha}$ vs. $S$ along with $\pm 2$ times the estimated standard error. The small increase in $\hat{\alpha}$ turns out to be well within the range expected from simulated iid stable variates with $\alpha = 1.65$. In particular, the stability test delivers both the test statistic, $\tau_T(\alpha)$, and the appropriate cutoff values (as a function of the estimated stable tail index $\alpha$ and sample size $T$) at the 90, 95 and 99 percent levels. The bhat Q-GARCH residuals yield $\tau_T(\alpha) = 3.23$, for which the null hypothesis of stability cannot be rejected at the 90 (or higher) percent level. Thus, to the extent to which the chosen model captures the dominating signal in the data and recovers approximately iid innovations, the stable hypothesis seems quite plausible.

A similar analysis is conducted for the returns on three other East Asian currency series, namely the Singapore dollar, South Korean won and Sri Lanka rupee, with estimation results also shown in Table 1. As with the Thai bhat, the AICC clearly favors the Q-GARCH over the A-PARCH model for these three currencies as well. The estimates for the tail index obtained with the Q-GARCH model range from 1.58 to 1.62 and are marginally lower than for the Thai bhat.
Table 1
Selected parameter estimates, goodness-of-fit and summability inference measures

<table>
<thead>
<tr>
<th>Return series</th>
<th>$S_{n, eta}^2$ A-PARCH(1, 1)</th>
<th>$S_{n, eta}^2$ Q-GARCH(2, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}$ (SE) $\hat{\beta}$ (SE) $\hat{\gamma}$ (SE) AICC AD $\tau_\gamma$ ((\alpha)) $\hat{\alpha}$ (SE) $\hat{\beta}$ (SE) $\hat{\gamma}$ (SE) AICC AD $\tau_\gamma$ ((\alpha))</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>1.54 (0.0034) 0.0303 (0.044) 0.982 (0.0037) -877.35 0.0862 3.11 * 1.65 (0.011) -0.0094 0.996 (0.0067) -913.5 0.0670 1.65</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>1.51 (0.011) -0.149 (0.068) 0.992 (0.0039) 240.83 0.0668 5.35 ** 1.62 (0.030) 0.214 1.011 (0.0033) 198.12 0.0542 4.44 **</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>1.51 (0.0026) 0.083 (0.016) 0.981 (0.0053) -1137.1 0.0644 5.70 *** 1.58 (0.0071) 0.0600 0.995 (0.00080) -1150.9 0.0944 3.30</td>
<td></td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>1.50 (0.009) 0.043 (0.044) 1.053 (0.011) 82.289 0.1244 3.82 * 1.58 (0.021) 0.0805 1.002 (0.0028) 59.724 0.1540 1.71</td>
<td></td>
</tr>
</tbody>
</table>

* Thailand, Singapore, Korea and Sri Lanka denote the Thai bhat–US dollar, Singapore dollar–US dollar, South Korean won–US dollar (January 1, 1990 to December 31, 1998) and Sri Lanka rupee–US dollar (January 1, 1992 to December 31, 1998), respectively. Column $\tau_\gamma$ (\(\alpha\)) is the summability test statistic; no stars means we cannot reject the null of stability at the 90% (and then also the 95% and 99%).

** The null of stability at 95% level cannot be rejected here.

*** The null of stability at 99% level cannot be rejected here.
Only for the Singapore dollar is the estimate of the skewness parameter, $\hat{\beta} = -0.214$, significantly different from zero, implying left skewness. For the persistence measure $\hat{V}$, we obtain values around unity—slightly above (below) one for the Singapore dollar (Sri Lanka rupee). The values for the $\tau_\alpha (\alpha)$ statistic are insignificant; an exception is the Singapore dollar, for which we reject the SuS hypothesis at the 95% level. With respect to sum stability, the A-PARCH model comes to somewhat different results. For two of the four currencies (South Korean won and Singapore dollar), we would reject the SuS hypothesis at the 95% level and for the two other currencies at the 90% level. This discrepancy for the two models indicates the importance of the chosen specification for the GARCH equation. It appears that the higher degree of non-linearity and the higher lag length of the Q-GARCH(2, 1) specification is more capable of producing iid-like residuals than the A-PARCH(1, 1) model. This coincides to some extent with other findings. Lane et al. (1996) demonstrated the superiority of the Q-GARCH model over that of bilinear models in terms of both model fit and forecasting ability for data sets exhibiting significant non-linearities; while Franses and van Dijk (1996) showed that the Q-GARCH model markedly outperformed several rival GARCH models in an extensive forecasting comparison. For either model, however, we strongly reject the normality assumption for the innovation process.

5. Conclusions

By combining GARCH models with (conditional) stable Paretian innovations, we provide a resolution to the seeming dichotomy discussed in GK as to the source of fat tails in financial returns data. We demonstrate the favorable small sample properties of point estimates and approximate standard errors for our proposed model, as well as detailing a method for calculation of the standard error of $\hat{V}$, the estimated volatility—persistence parameter.

We also reinforce the results of GK, demonstrating that data generated from (generalized) GARCH models with conditional stable Paretian innovations using parameters typical of those observed in practice give rise to series with virtual summability—which is clearly misleading in light of the data generating process.

Finally, we apply the suggested models to exchange-rate returns from several East Asian countries before and after the currency crisis and, using an unbiased summability test with adequate power against non-stable alternatives, show that the proposed models are plausible.

Acknowledgements

We thank Franz Palm, Jon Danielsson, two referees and the participants of the 1999 Journal of Empirical Finance Conference on Risk Management for constructive comments.
Appendix A

We briefly outline the computation of the stable tail index estimator and the summability test used. For a more detailed account see Paolella (1997), Mittnik and Paolella (1999) and Paolella (2000). The tail index estimator $\hat{\alpha}_{\text{Hill}}$ is specifically designed for stable Paretoian data and given by:

$$\hat{\alpha}_{\text{Hill}} = -0.8110 - 0.3079 \hat{b} + 2.0278 \hat{b}^{0.5},$$

(14)

where $\hat{b}$ is the intercept in the simple linear regression of $\hat{\alpha}_{\text{Hill}}(k)$ on $k/1000$; the elements of vector $k$ are such that $0.2T \leq k \leq 0.8T$ in steps of max $\{\lfloor T/100 \rfloor, 1\}$, and $\hat{\alpha}_{\text{Hill}}(k)$ is the popular Hill (1975) estimator:

$$\hat{\alpha}_{\text{Hill}}^{-1} = k^{-1} \sum_{j=1}^{k} \ln(Z_{T+1-jT}) - \ln Z_{n-kT}$$

(15)

with $Z_{jT}$ denoting the $j$th order statistic of sample $Z_1, \ldots, Z_T$. An accurate approximation to its standard error is given by:

$$\sqrt{\text{SE}(\hat{\alpha}_{\text{Hill}})} \approx 0.0322 - 0.00205T + 0.02273T_{s}^{-1} - 0.0008352T_{s}^{-2}.$$  \hspace{1cm} (16)

Estimator $\hat{\alpha}_{\text{Hill}}$ is unbiased and virtually exactly normally distributed in samples as small as 50, with estimated small-sample variance both lower and more accurate than that given in McCulloch (1986). A general tail estimator designed for Pareto-type distributions, which also exploits the approximate linearity of the Hill estimator in $k$, is proposed by Huisman et al. (1997).

Consider condensing the $T$-length vector of (presumably iid stable Paretoian) realizations into summed non-overlapping $S$-length segments. Let $\hat{\alpha}(s)$ denote the vector of tail-index estimates using the tail estimator $\hat{\alpha}_{\text{Hill}}$ evaluated at each element in $s = \{1, 2, \ldots, S_{\text{max}}(T)\}$, with $S_{\text{max}}(T) = \min(10, \lfloor T/200 \rfloor)$. The test statistic is given by:

$$\tau_T = (\alpha) = \hat{b}/\sqrt{\text{SE}(\hat{b})},$$

(17)

where $\hat{b}$ denotes the weighted least-squares estimate of the slope of $\hat{\alpha}(s)$ regressed on $s$ (and a constant) and $\sqrt{\text{SE}(\hat{b})}$, its corresponding standard error. The weights are taken to be the inverse of the estimated standard error of $\hat{\alpha}(s)$ as given in Eq. (16).

Under the null hypothesis of iid stable Paretoian realizations, we have $\tau_T(\alpha) = 0$ while, under the alternative, $\tau_T(\alpha) > 0$. The sampling distribution of $\tau_T(\alpha)$ is non-standard; simulated $\gamma$-level cutoff values under the null hypothesis are approximated by:

$$C_T(\alpha, \gamma) = c_{0,\gamma} + c_{1,}\gamma \alpha + c_{2}\gamma \alpha^2$$

(18)
for $1 < \alpha < 2$ and $1500 \leq T \leq 10,000$. In practice, an estimated value of $\alpha$ (obtained by using the whole sample) is required. For $\gamma = 0.90, 0.95$ and 0.99, the required coefficients in Eq. (18) can be expressed as functions of $T$:

$$
c_{0.90} = 6.891 - 0.22591 T_s^\alpha + 0.11807 T_s^\gamma
$$

$$
c_{1.90} = -3.405 + 0.08045 T_s^\alpha + 1.479 T_s^\gamma
$$

$$
c_{2.90} = 0.4986 + 0.05480 T_s^\alpha - 0.86807 T_s^\gamma
$$

$$
c_{0.95} = 8.377 - 0.1007 T_s^\alpha - 0.05117 T_s^\gamma
$$

$$
c_{1.95} = -1.408 + 0.3862 T_s^\alpha + 0.0431 T_s^\gamma
$$

$$
c_{2.95} = -1.124 - 0.2334 T_s^\alpha + 0.3542 T_s^\gamma
$$

$$
c_{0.99} = 14.28 - 0.02562 T_s^\alpha + 0.2462 T_s^\gamma
$$

$$
c_{1.99} = -4.081 + 0.02769 T_s^\alpha + 0.6514 T_s^\gamma
$$

$$
c_{2.99} = -0.5839 - 0.005483 T_s^\alpha - 0.3507 T_s^\gamma
$$

where $T_s = T/1000$.

A Matlab program to compute both the estimator Eq. (14) and the test statistic Eq. (18) with their cutoff values is available from the authors upon request.

References


Lumsdaine, R.L., 1995. Finite-sample properties of the maximum likelihood estimator in GARCH(1, 1)
and IGARCH(1, 1) models: a Monte Carlo investigation. Journal of Business and Economic
of Banking and Finance 9, 137–156.
McCulloch, J.H., 1997b. Measuring tail thickness in order to estimate the stable index of the stable
Mittnik, S., Paolella, M.S., 1999. A simple estimator for the characteristic exponent of the stable
Reviews 12, 261–330.
Mittnik, S., Rachev, S.T., 1993b. Reply to comments on ‘Modeling Asset Returns with Alternative
Mittnik, S., Paolella, M.S., Rachev, S.T., 1998a. A tail estimator for the index of the stable Paretian
Mittnik, S., Paolella, M.S., Rachev, S.T., 1998b. Unconditional and conditional distributional models
Feldman, R.E., Taqqu, M.S. (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques for
Analyzing Heavy Tailed Distributions. Birkhäuser, Boston.
Mittnik, S., Doganoglu, T., Chenyao, D., 1999a. Computing the probability density function of the
University, Kiel, Germany.
report, Institute of Statistics and Econometrics, Christian Albrechts University, Kiel, Germany.
347–370.
thesis, Institute of Statistics and Econometrics, Christian Albrechts University, Kiel, Germany.
in press.