

How to apply GARCH model in risk management?

Model diagnosis on GARCH innovations*

Pengfei Sun^{a,†}, Chen Zhou ^{a,b}

a. Erasmus University Rotterdam

b. De Nederlandsche Bank

Abstract

Having accurate estimates on downside risks is a key step in risk management. For financial time series, specific features such as heavy-tailedness and volatility clustering lead to difficulties in downside risk evaluation. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model captures these features regardless of the distributional assumptions on the innovation process. Nevertheless, distribution of the innovations plays an important role when analyzing conditional and unconditional downside risk. We show that the diagnosis method on the heavy-tailedness of GARCH innovations in McNeil and Frey [2000] is not reliable for GARCH processes that are close to Nonstationarity. With comparing different tail index estimates, we provide an alternative approach which leads to a formal test on the distribution of GARCH innovations. Empirical analysis on real data confirms similar finding as in McNeil and Frey [2000] when modeling financial returns with the GARCH model, the downside distribution of innovations possesses heavier tail than the usual normality assumption.

Key words: Dynamic Risk Management, GARCH(1,1), Extreme Value Theory, Hill Estimator.

EFM Classification: 450

*Pengfei Sun is a Marie Curie Fellow at Erasmus University Rotterdam. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7-PEOPLE-ITN-2008 under grant agreement number PITN-GA-2009-237984 (project name: RISK). The funding is gratefully acknowledged.

[†]Corresponding author. Email: sun@ese.eur.nl

1 Introduction

Downside risk of financial investment is of the major concern for investors. Risk managers often assess downside risk measures such as the Value-at-Risk (VaR) to evaluate the potential large loss of their investment portfolios. Regulators also consider downside risk of financial institutions and impose regulation to prevent severe systemic crisis based on quantitative risk measures. Having an accurate estimate of the downside risk measures is not an easy task, due to specific features of financial time series. First, the time series of financial returns exhibit volatility clustering. Second, the distribution of financial returns exhibit heavy-tails: the downside tail of the return distribution decays in a power-law speed instead of the exponential speed as in that of the normal distribution. The two specific features impose a great challenge in producing accurate and time-varying estimates of downside risk measures.

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev [1986] attempts to capture the volatility clustering feature of financial returns by modeling the dynamic of volatility. Thus, it leads to time-varying estimate on downside risk measures such as the VaR. Another surprising fact of the GARCH-type models is that they capture the heavy-tailedness at the same time. Following the result in Kesten [1973], the stationary solution of GARCH (1,1) process follows a heavy-tailed distribution, see Mikosch and Starica [2000], Davis and Mikosch [2009], etc. Therefore, the GARCH-type models turn to be an effective instrument in risk management.

This paper provides a diagnosis framework on the distributional assumption of GARCH innovations. The distribution of GARCH innovations plays an important role for both conditional and unconditional risk measurement. When analyzing conditional risk, it is obvious that the conditional distribution of the future returns is the same as the distribution of innovations. When analyzing unconditional risk, we show that GARCH models with different innovation distributions lead to different shape of the downside tail, hence, different estimates of the tail risk. Therefore, an inappropriate distributional assumption of innovations may lead to either underestimation or overestimation of the downside risk. It is thus important to clarify which type of GARCH innovations should be applied for modeling financial returns and how this influences the risk analysis on the original GARCH process.

The most often applied GARCH (1,1) model with normal distributed innovation assumes that the conditional distribution of financial return given the return level and the volatility of the previous period follows a normal distribution. In risk management, recent literature shows that the conditional normality assumption does not perform well in estimating the downside risk with a low probability, see Danielsson and De Vries [2000]. Mikosch and Starica [2000] show that the GARCH process with normal innovation generates much thinner tail than that obtained from the real data. McNeil and Frey [2000] show that the GARCH models with heavy-tailed innovation is more efficient in estimating and forecasting the downside risk of financial returns, whereas the estimates from GARCH models with normal innovation underestimate the potential downside risk.

McNeil and Frey [2000] detects the heavy-tailedness of innovations based

on estimated innovations. They find that the conditional normality assumption is not valid. The diagnosis procedure can be divided into two steps: first, they use Quasi-Maximum-Likelihood-Estimates (QMLE) to estimate the GARCH coefficients and back out the innovations; second, they use the Generalized Pareto Distribution to model the estimated innovations and consequently estimate the tail shape of the distribution of the innovations. This method has been followed by other studies, see e.g. Hang Chan et al. [2007].

We start by revisiting their method and show that based on the estimated innovations data generated from the GARCH model with normal innovations may still violate the normality assumption. This phenomenon turns apparent when the GARCH process is close to non-stationarity. Thus the McNeil and Frey [2000] method is not robust in diagnosing the heavy-tailedness of GARCH innovations. Next, we provide theoretical reason in explaining this phenomena. We develop an alternative approach based on analyzing the tail index of the GARCH process. Our method yields a formal test on the distributional assumption of the innovations. Taking normal and Student-t distributed innovations as examples, simulations show that our method is robust in the case of *Near-Nonstationarity*. Moreover, this method leads to a robust estimate of the tail index of a GARCH process.

We apply our method to the S&P 500 Composite Index and 12 S&P equity sector indices. The estimated GARCH coefficients indicate that the fitted GARCH models are close to Nonstationarity. Therefore, the McNeil and Frey [2000] method is not valid in this case. With our formal test, we reject the normal innovation in most cases, while can not always reject

hypothesis on the Student-t innovation.

The rest of the paper is organized as follows. In Section 2, we show that using the McNeil and Frey [2000] approach based on estimated innovations is not robust for detecting heavy-tailedness of innovations. Instead, we develop a test based on analyzing the tail index of the GARCH (1,1) model. A discussion on the Hill estimator used in our test is given in Section 3. Simulations and empirical results are presented in Section 4. Section 5 concludes.

2 Theory

2.1 Heavy-tailedness of the GARCH Series

We consider the GARCH (1,1) model in modeling the time series of financial returns. Suppose the returns $\{X_t\}$ satisfies the following model:

$$X_t = \varepsilon_t \sigma_t, \quad (1)$$

$$\sigma_t^2 = \lambda_0 + \lambda_1 X_{t-1}^2 + \lambda_2 \sigma_{t-1}^2, \quad (2)$$

where $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) innovations with zero mean and unit variance, the parameters λ_0 , λ_1 , λ_2 are positive. Moreover, in order to have a stationary solution of the GARCH model, we assume the stationary condition $\lambda_1 + \lambda_2 < 1$.

The heavy-tailedness of the stationary solution of a GARCH model follows from the result of Kesten [1973]. Consider a process $\{Y_t\}_{t=0}^{\infty}$ satisfying

the stochastic difference equation

$$Y_t = Q_t Y_{t-1} + M_t, \quad (3)$$

where $\{(Q_t, M_t)\}$ are i.i.d. R_+^2 -valued random pairs. Kesten [1973] shows that the stationary solution of the stochastic difference equation follows a heavy-tailed distribution. Suppose there exists a positive real number κ , such that

$$E(Q_1^\kappa) = 1, \quad E(Q_1^\kappa \log Q_1) < \infty, \quad 0 < E(M_1^\kappa) < \infty.$$

Moreover, assume that $\frac{M_1}{1-Q_1}$ is non-degenerate and the conditional distribution of $\log Q_1$ given $Q_1 \neq 0$ is nonlattice. Then the stationary solution of $\{Y_t\}$ follows a heavy-tailed distribution as

$$P(Y_t > x) = Ax^{-\kappa}[1 + o(1)], \quad \text{as } x \rightarrow \infty, \quad (4)$$

where κ is the so-called tail index and A is the tail scale.

The GARCH model is associated to a specific stochastic difference equation. By combining the equations (1) and (2), we derive the following stochastic difference equation on the stochastic variance series $\{\sigma_t^2\}_{t=0}^\infty$ as

$$\sigma_t^2 = \lambda_0 + (\lambda_1 \varepsilon_t^2 + \lambda_2) \sigma_{t-1}^2, \quad (5)$$

which satisfies equation (3) with $Q_t = \lambda_1 \varepsilon_t^2 + \lambda_2$ and $M_t = \lambda_0$. Denote

$\varphi_t = \varepsilon_t^2$. Suppose κ is the solution to the equation

$$E[(\lambda_1 \varphi_t + \lambda_2)^\kappa] = 1. \quad (6)$$

The stationary solution of σ_t^2 follows a heavy-tailed distribution with tail index κ . Hence, σ_t follows a heavy-tailed distribution with tail index 2κ . The relation (6) implies that $E[\varepsilon_t^{2\kappa}] < \infty$. Therefore, the tail of σ_t is heavier than that of ε_t . If ε_t follows a thin-tailed distribution such as the normal distribution, σ_t follows a heavy-tailed distribution. If ε_t follows a heavy-tailed distribution, then the tail index of σ_t is lower than that of ε_t .

From Mikosch and Starica [2000], we can derive the following relation

$$P\{|X_t| > x\} = P\{|\sigma_t \varepsilon_t| > x\} \sim E[|\varepsilon_t|]^\kappa P\{\sigma_t > x\}, \text{ as } x \rightarrow \infty.$$

Thus $|X_t|$ has a similar tail behavior as σ_t , in the sense that the tail index of $|X_t|$ is 2κ .

From the discussion, we observe that the general heavy-tailed feature of the GARCH model is irrelevant to the distribution of the innovation. No matter the innovation follows a thin or heavy-tailed distribution, the stationary solution of the GARCH model is always heavy-tailed. However, the shape of the tail distribution of a stationary GARCH series does depend on the distribution of innovations: the solution to equation (6) differs for different distributions of φ_t . Hence different distributional assumptions on the innovations may lead to different risk analysis on a GARCH series. It is thus necessary to verify which innovation model fits the actual financial

returns.

2.2 Diagnosing the Distribution of GARCH Innovations: the McNeil and Frey Approach

McNeil and Frey [2000] (MF2000) confirms that when modeling financial time series by a GARCH model, the innovation is heavy-tailed. They show the heavy-tailedness of the innovations by making QQ plot on the estimated innovations against a normal distribution. Since the innovations are backed out from an estimation procedure, its heavy-tailedness may potentially be imposed by the estimation procedure. We demonstrate this phenomenon by the following simulation.

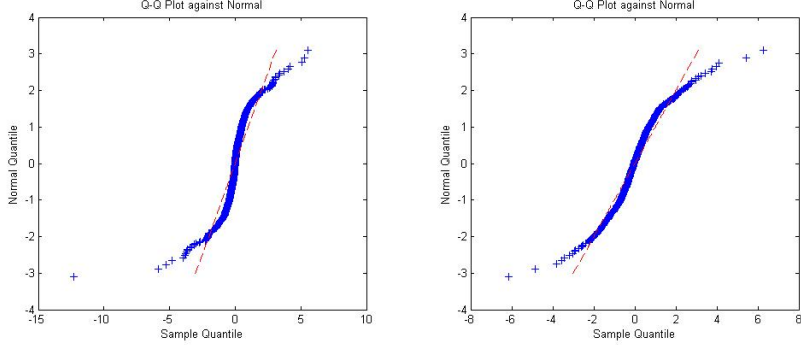
Given the Maximum-Likelihood-Estimates (MLE) $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2$ (see Bollerslev et al. [1986]) on the GARCH coefficients, one can estimate the innovations $\hat{\varepsilon}_t$ by

$$\hat{\varepsilon}_t = X_t / \hat{\sigma}_t, \tag{7}$$

$$\hat{\sigma}_t^2 = \hat{\lambda}_0 + \hat{\lambda}_1 X_{t-1}^2 + \hat{\lambda}_2 \hat{\sigma}_{t-1}^2. \tag{8}$$

Using the same data as in MF2000, we reproduce their QQ plot in Figure 1 (left panel). In addition, we generate a series of observations from the GARCH (1,1)-normal process with the model coefficients equivalent to the MLE obtained from the real data. Then we re-estimate the innovations from (7) & (8) and display the corresponding QQ plot in Figure 1 (right panel). We find that the estimated innovations from the generated data still violate the normality assumption, while exhibit a heavy-tailed feature.

Figure 1: QQ-Plots



Note: The QQ-plots show the estimated innovations against the sample quantiles from a normal distribution. Straight line corresponds to that the estimated innovations follow a normal distribution. The left panel demonstrates the estimated innovations from fitting the S&P 500 returns (from June 1985 to May 1989) to a GARCH (1,1) model by MLE. The right panel demonstrates the estimate innovations from fitting a generated sample, with the same sample size, while data are generated from a GARCH (1,1)-normal model, with the coefficients equivalent to the MLE from S&P 500 returns.

Recall that the generated data follow exactly a GARCH (1,1) model with normal innovations. We observe that using estimated innovations backed out from a finite sample is not a robust method in testing the heavy-tailedness of the innovations. The intuition of the non-robustness is given as follows.

Suppose the GARCH coefficients are perfectly estimated, i.e. $\hat{\lambda}_0 = \lambda_0$, $\hat{\lambda}_1 = \lambda_1$, $\hat{\lambda}_2 = \lambda_2$. By combining equations (2) and (8), we have that

$$\hat{\sigma}_t^2 - \sigma_t^2 = \lambda_2(\hat{\sigma}_{t-1}^2 - \sigma_{t-1}^2) = \dots = \lambda_2^t(\hat{\sigma}_0^2 - \sigma_0^2)$$

Hence, σ_t^2 is not accurately estimated by $\hat{\sigma}_t^2$ for a finite t . This may lead to a misestimation in the innovations. The difference between $\hat{\sigma}_t^2$ and σ_t^2 stems from that between $\hat{\sigma}_0^2$ and σ_0^2 . Moreover, the difference between $\hat{\sigma}_0^2$ and σ_0^2 comes from the fact that $\hat{\sigma}_0^2$ is simply some initial value chosen in estimation,

while σ_0^2 follows the stationary solution of the stochastic difference equation (5), i.e. a heavy-tailed distribution. When $\lambda_1 + \lambda_2$ is close to 1, the parameter κ from equation (6) is close to 1. In the case $\kappa = 1$, we get that $E(\sigma_t^2) = +\infty$. Hence any initial value $\hat{\sigma}_0^2$ may underestimate the potential σ_0^2 , which implies that $\hat{\sigma}_t^2$ underestimates σ_t^2 . Hence $\hat{\varepsilon}_t$ may demonstrate a heavier tail than ε_t . This is the main intuition why given that ε_t follows a normal distribution, it is still possible to obtain heavy-tailedness in the distribution of $\hat{\varepsilon}_t$.

The following lemma shows theoretically that based on finite observations generated from a GARCH model with normal innovations, the estimated innovations follow a heavy-tailed distribution.

Proposition 1 *Consider a GARCH (1,1) model in (1) and (2) with normal distributed innovations $\{\varepsilon_t\}$. Suppose $\hat{\lambda}_0$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ perfectly estimate λ_0 , λ_1 λ_2 , i.e. $\hat{\lambda}_0 = \lambda_0$ $\hat{\lambda}_1 = \lambda_1$, $\hat{\lambda}_2 = \lambda_2$. The estimated innovations from (7) and (8) follow a heavy-tailed distribution for any finite t .*

2.3 Testing Distributional Assumptions on GARCH Innovations

The different distributional assumptions on the GARCH innovations lead to different level of heavy-tailedness of the GARCH series, measured by its tail index. This provides an alternative way to test which innovation fits the data.

We investigate the difference between Student-t and normal innovations as an example. The GARCH series with these two types of innovations

exhibit different tail behavior.

Definition 2 *In a GARCH(1,1)-normal model, the innovation ε_t follows a standard normal distribution. The parameter κ solved from equation (6) is denoted as $\kappa_n(\lambda_1, \lambda_2)$. Then the tail index of the GARCH series is $\alpha_n(\lambda_1, \lambda_2) = 2\kappa_n(\lambda_1, \lambda_2)$.*

Definition 3 *In a GARCH(1,1)-Student model, the innovation ε_t follows a Student-t distribution with degree of freedom ν , normalized to unit variance. The parameter κ solved from equation (6) is denoted as $\kappa_s(\lambda_1, \lambda_2, \nu)$. The tail index of the GARCH series is $\alpha_s(\lambda_1, \lambda_2, \nu) = 2\kappa_s(\lambda_1, \lambda_2, \nu)$.*

Notice that $\kappa_n = \kappa_s = 1$ for $\lambda_1 + \lambda_2 = 1$.

The GARCH coefficients λ_1, λ_2 are connected to the tail index according to $\alpha_n(\lambda_1, \lambda_2)$ or $\alpha_s(\lambda_1, \lambda_2, \nu)$. With assuming the innovations follow a normal or Student-t distribution, we obtain the implied tail indices $\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2)$ or $\alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu})$ after estimating the GARCH coefficients by $\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\nu}$.

The parameters λ_1, λ_2 and ν can be estimated under either the GARCH-normal or Student-t specification by for example the MLE procedure. Once the distribution of the innovations is correctly specified, the estimates are consistent. Together with the fact that both $\kappa_n(\lambda_1, \lambda_2)$ and $\kappa_s(\lambda_1, \lambda_2, \nu)$ are continuous functions with respect to λ_1, λ_2 and ν , we get that the implied tail indices $\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2)$ and $\alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu})$ are consistent estimates of the tail index of the GARCH series.

However, if the distribution of innovations is misspecified, we have the following lemma showing the inconsistency of the implied tail index.

Lemma 4 *Suppose X_1, \dots, X_T follow a GARCH (1,1)-Student process, $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}$ are the consistent estimates of $\lambda_1, \lambda_2, \nu$ when fitting either a GARCH-normal or GARCH-Student model, α is the real tail index of the series. Then as $T \rightarrow \infty$,*

$$P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \rightarrow 1.$$

The relation $P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \rightarrow 1$ indicates that $\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2)$ overestimate the real tail index α when the observations are drawn from a GARCH(1,1)-Student model. A detailed proof is given in Appendix.

Alternatively, the tail index of a GARCH series can be estimated from the so-called Hill estimator in extreme value analysis. Let X_1, \dots, X_T be the observations from a heavy-tailed distribution as in equation (4). The Hill estimator is defined as

$$\hat{\alpha}_H := \left(\frac{1}{k} \sum_{i=1}^T 1_{\{X_i > s\}} [\log(X_i) - \log(s)] \right)^{-1}, \quad (9)$$

where k is the number of observations that exceed the threshold s , satisfying $\frac{k}{T} \rightarrow 0$ as $T \rightarrow \infty$. The Hill estimator is usually applied to i.i.d. sample. Nevertheless, Resnick and Stărică [1998] shows that the consistency of Hill estimator for the solutions of stochastic difference equations of the form in equation (3) holds. For the asymptotic normality of the Hill estimator, it follows from Hsing [1991] and Carrasco and Chen [2002]. Following the result in Carrasco and Chen [2002], the GARCH (1,1) process we are studying satisfies the β -mixing conditions. Following the result in Hsing [1991], the Hill estimator converges to the tail index α with speed of convergence \sqrt{k} .

The asymptotic limit is given as

$$\sqrt{k} \left(\frac{1}{\hat{\alpha}_H} - \frac{1}{\alpha} \right) \xrightarrow{d} N(0, v^2),$$

where the asymptotic variance v^2 is given as $v^2 = \frac{1+\chi+\omega-2\psi}{\alpha^2}$. The parameters χ , ω and ψ refer to measures on the level of serial dependence. The estimators $\hat{\chi}$, $\hat{\omega}$ and $\hat{\psi}$ are given in (3.6) in Hsing [1991]. Hence for GARCH (1,1) process, we can apply the Hill estimator with the asymptotic property as follows:

$$\sqrt{k} \left(\frac{\alpha}{\hat{\alpha}_H} - 1 \right) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi).$$

By testing whether the implied tail indices obtained from the GARCH (1,1)-normal and GARCH (1,1)-Student models differ from the estimated tail index, we can formally distinguish the two models. The following theorem gives the asymptotic properties of the test statistics. With the fact that $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}$ converges to the parameters $\lambda_1, \lambda_2, \nu$ with a faster speed of convergence than that of the Hill estimator. The proof follows from the asymptotic normality of the Hill estimator, thus it is omitted here.

Theorem 5 Denote $\hat{\alpha}_n = \alpha_n(\hat{\lambda}_1, \hat{\lambda}_2)$, $\hat{\alpha}_s = \alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu})$, where $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}$ are consistent estimates of $\lambda_1, \lambda_2, \nu$ with a speed of convergence \sqrt{T} , α is the real tail index of the series. $\hat{\alpha}_H$ is the Hill estimator with sample fraction k , such that $k \rightarrow \infty, \frac{k}{T} \rightarrow 0$ as $T \rightarrow \infty$ and $\sqrt{k}(\frac{\alpha}{\hat{\alpha}_H} - 1) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi)$.

1. If X_1, \dots, X_T follow a GARCH (1,1)-normal process, then

$$\sqrt{k}(\frac{\hat{\alpha}_n}{\hat{\alpha}_H} - 1) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi), \text{ as } T \rightarrow +\infty.$$
2. If X_1, \dots, X_T follow a GARCH (1,1)-Student process, then

$$\sqrt{k}\left(\frac{\hat{\alpha}_s}{\hat{\alpha}_H} - 1\right) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi), \text{ as } T \rightarrow +\infty.$$

Note the MLE satisfies the requirement on the speed of the convergence \sqrt{T} , see Bollerslev et al. [1986].

3 The Asymptotic Bias of the Hill Estimator

Although the Hill estimator is a valid method in estimating the tail index, it bears potential asymptotic bias due to the fact that the tail distribution is not an exact Pareto distribution. In the theoretical setup, the bias is neglected by assuming $\sqrt{k}\left(\frac{\alpha}{\hat{\alpha}_H} - 1\right) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi)$ as $T \rightarrow \infty$. This is difficult to achieve in practice. The following lemma shows how the approximation influences the asymptotic bias of the Hill estimator (see Goldie and Smith [1987]).

Proposition 6 *Suppose observations are obtained from a distribution possessing a density and satisfying the following equation*

$$F(x) = 1 - Ax^{-\alpha} \left[1 + Bx^{-\beta} + o\left(x^{-\beta}\right) \right], \quad \beta > 0, \quad \text{as } x \rightarrow \infty, \quad (10)$$

where β and B are the second order tail index and tail scale respectively. Let the threshold s satisfy $s^\alpha/T \rightarrow 0$, $s \rightarrow \infty$ as $T \rightarrow \infty$. The asymptotic bias of $\hat{\alpha}_H$ is

$$E[\hat{\alpha}_H - \alpha] = \frac{B\alpha\beta}{(\alpha + \beta)} s^{-\beta} + o\left(s^{-\beta}\right), \quad (11)$$

and variance is

$$\text{Var}[\hat{\alpha}_H] = \frac{\alpha^2 s^\alpha}{aT} + o\left(\frac{s^\alpha}{T}\right). \quad (12)$$

Combining the asymptotic bias and variance, we obtain the Asymptotic Mean Squared Error (AMSE) of the estimator $\hat{\alpha}_H$ as follows,

$$AMSE(\hat{\alpha}_H) = \frac{B^2 \alpha^2 \beta^2}{(\alpha + \beta)^2} s^{-2\beta} + \frac{\alpha^2 s^\alpha}{AT}. \quad (13)$$

One can choose the optimal threshold s by minimizing (13). With the optimal threshold, the squared bias and variance vanish at the same rate.

The asymptotic bias problem turns severe for finite sample applications with serial dependence such as the GARCH series. This may potentially contaminate the test procedure introduced in Theorem 5. Therefore, we investigate the asymptotic bias of the Hill estimator when applying to a GARCH series.

We start by clarifying the second order approximation for the tail distribution of the stationary solution of a GARCH model.

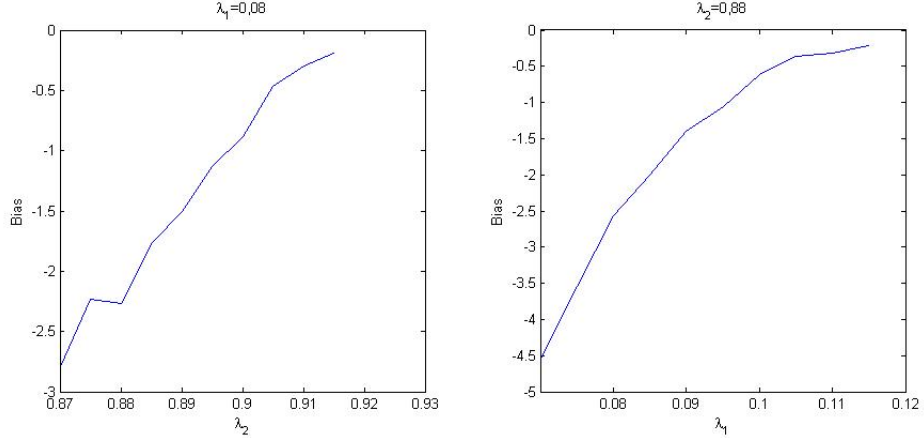
Lemma 7 *For the GARCH (1,1)-normal model, suppose the tail expansion of σ_t^2 satisfies equation (10), where its tail index κ is the solution to the equation $E[(\lambda_1 \varphi_t + \lambda_2)^\kappa] = 1$ for the $\chi^2(1)$ distributed random variable φ . Then, $\beta = 1$ and*

$$B = \frac{\kappa \lambda_0 E[(\lambda_1 \varphi + \lambda_2)^{\kappa+1}]}{1 - E[(\lambda_1 \varphi + \lambda_2)^{\kappa+1}]} < 0. \quad (14)$$

Remark 8 *For the GARCH (1,1)-Student model, the second order index $\beta = 1$ and B has the same form as in (14) if $E[(\lambda_1 \varphi + \lambda_2)^{\kappa+1}] < \infty$.*

Since the constant B is negative for the stationary solution of the GARCH model with normal innovations, the Hill estimator applied to such a GARCH series is downward biased. We demonstrate the Hill bias for the GARCH(1,1)-

Figure 2: Hill Bias of the GARCH (1,1)-normal model



Note: The plots show the Hill bias of the GARCH(1,1)-normal models at the simulated optimal threshold level. In the left panel, we fix $\lambda_1 = 0.08$ and plot the bias against λ_2 , while in the right panel, we fix $\lambda_2 = 0.88$ and plot the bias against λ_1 .

normal model by simulation against various values of λ_1 and λ_2 . We first fix λ_1 at 0.08 and vary λ_2 from 0.87 to 0.91, then fix λ_2 at 0.88 and vary λ_1 from 0.07 to 0.11. The simulation algorithm is as follows.

1. For any fixed λ_1 and λ_2 , calculate the implied tail index $\alpha = 2\kappa$, where κ is solved from equation (6).
2. Generate 2500 samples with sample size 5000 from the GARCH (1,1)-normal model, then calculate the Hill estimator $\hat{\alpha}_i(k)$ of the downside tail for each sample fraction k , where $i = 1, \dots, 2500$, $k = 1, \dots, 500$.
3. For each k , calculate the corresponding Mean Squared Error (MSE) as $MSE(\hat{\alpha}(k)) = \frac{1}{2500} \sum_{i=1}^{2500} (\hat{\alpha}_i(k) - \alpha)^2$.
4. Find the optimal sample fraction k^* through minimizing the MSE,
5. Plot the bias $\frac{1}{2500} \sum_{i=1}^{2500} (\hat{\alpha}_i(k^*) - \alpha)$ against λ_1 or λ_2 .

We observe that the bias problem turns severe for $\lambda_1 + \lambda_2 \ll 1$, while it is of a lower importance when $\lambda_1 + \lambda_2$ is close to 1. Hence our test based on comparing implied tail indices with the estimated tail index by the Hill estimation is particularly effective for the case that $\lambda_1 + \lambda_2$ is close to 1. In other words, when the GARCH process is close to Nonstationarity, our method works better.

4 Simulation and Empirical Studies

4.1 Simulation

We use simulation to validate our new diagnosis method on the distribution of GARCH innovations. We consider two types of distributional assumption, normal and Student-t distributed innovations.

We generate observations from each model with specific coefficients, then fit the generated data to both models. The GARCH coefficients are estimated by the MLE outlined in Bollerslev et al. [1986]. With the estimated GARCH coefficients, we calculate the implied tail indices $\hat{\alpha}_n$ and $\hat{\alpha}_s$ according to the fitted models. Moreover, we obtain the Hill estimates $\hat{\alpha}_H$ by choosing an optimal sample fraction k .¹

Each generated sample consists of 4,000 observations. This is close to the sample size we use later for real data analysis. For each model and fitting procedure, we repeat 100 times to obtain an average estimate for each tail index estimator.

¹The optimal k is chosen from the first stable region of the tail index estimates in the Hill plots, which is the plot of the estimates against various potential k levels, see de Haan and de Ronde [1998].

Table 1: Tail index estimates of GARCH (1,1)-normal Series

$\lambda_1 = 0.08$					$\lambda_1 = 0.88$				
λ_2	α	$\hat{\alpha}_n$	$\hat{\alpha}_s$	$\hat{\alpha}_H$	λ_1	α	$\hat{\alpha}_n$	$\hat{\alpha}_s$	$\hat{\alpha}_H$
0.87	11.88	12.02	11.55	6.15	0.07	13.98	14.44	13.80	6.47
0.88	10.62	11.34	10.88	5.94	0.08	10.62	11.63	11.16	5.86
0.89	9.12	8.82	8.51	5.30	0.09	7.90	7.89	7.63	5.05
0.90	7.28	7.42	7.17	4.86	0.10	5.62	5.88	5.71	4.29
0.91	5.00	5.21	5.07	4.31	0.11	3.70	3.96	3.87	3.65

Note: This table presents the tail index estimation results of the data simulated from a GARCH (1,1)-normal model. For simulating the data, in the left part of the table we fix the GARCH coefficient $\lambda_1 = 0.08$ and vary λ_2 from 0.87 to 0.91, while in the right part, we fix $\lambda_2 = 0.88$ and vary λ_1 from 0.07 to 0.11, $\lambda_0 = 0.5$. For each model, 100 samples are simulated with each sample consisting of 4,000 observations. The simulated data are fitted to a GARCH (1,1)-normal model and a GARCH (1,1)-Student-t model. With the coefficients estimated from the Maximum Likelihood method, we calculate the implied tail indices $\hat{\alpha}_n$, $\hat{\alpha}_s$. $\hat{\alpha}_H$ is the Hill estimate from the simulated data. Numbers reported are the average levels across 100 samples.

Tables 1 presents the simulation results of GARCH (1,1)-normal model. For the GARCH coefficients, we first fix $\lambda_1 = 0.08$ with varying λ_2 from 0.87 to 0.91, then fix $\lambda_2 = 0.88$ with varying λ_1 from 0.07 to 0.11. The constant λ_0 is fixed at 0.5. From the results, we observe that both $\hat{\alpha}_n$ and $\hat{\alpha}_s$ robustly estimate the tail index α . However, the Hill estimates $\hat{\alpha}_H$ considerably underestimate the tail index α , i.e. overestimates the heavy-tailedness. The underestimation is severe for the case that $\lambda_1 + \lambda_2$ is relatively low. This can be explained by the downward bias of the estimate, see Lemma 8. Nevertheless, the Hill estimator performs well in the case that $\lambda_1 + \lambda_2$ is close to 1, in other words, when the GARCH series is close to Nonstationarity.

The simulation results of GARCH (1,1)-Student model are presented in Table 2. We use the same parameters λ_0 , λ_1 and λ_2 as in simulations for the GARCH (1,1)-normal model, while the degree of freedom for Student-t innovation is set to $\nu = 6$. Different from the normal case, only $\hat{\alpha}_s$ robustly

Table 2: Tail index estimates of GARCH (1,1)-Student series

$\lambda_1 = 0.08$					$\lambda_1 = 0.88$				
λ_2	α	$\hat{\alpha}_n$	$\hat{\alpha}_s$	$\hat{\alpha}_H$	λ_1	α	$\hat{\alpha}_n$	$\hat{\alpha}_s$	$\hat{\alpha}_H$
0.87	5.94	12.54	6.00	3.94	0.07	6.48	15.25	6.68	4.03
0.88	5.56	11.50	5.81	3.90	0.08	5.56	11.28	5.74	3.78
0.89	5.08	9.54	5.10	3.55	0.09	4.66	8.21	4.79	3.64
0.90	4.42	8.29	4.78	3.51	0.10	3.78	6.06	3.95	3.26
0.91	3.46	5.42	3.46	3.07	0.11	2.88	3.93	2.91	2.95

Note: This table presents the tail index estimation results of the data simulated from a GARCH (1,1)-Student model. For simulating the data, in the left part of the table we fix the GARCH coefficient $\lambda_1 = 0.08$ and vary λ_2 from 0.87 to 0.91, while in the right part, we fix $\lambda_2 = 0.88$ and vary λ_1 from 0.07 to 0.11, $\lambda_0 = 0.5$. The degree of freedom for the Student-t innovation $\nu = 6$. For each model, 100 samples are simulated with each sample consisting of 4,000 observations. The simulated data are fitted to a GARCH (1,1)-normal model and a GARCH (1,1) - Student-t model. With the coefficients estimated from the Maximum Likelihood method, we calculate the implied tail indices $\hat{\alpha}_n$, $\hat{\alpha}_s$. $\hat{\alpha}_H$ is the Hill estimate from the simulated data. Numbers reported are the average levels across 100 samples.

estimates the tail index of the GARCH series in this case. The normal innovation fitted estimate $\hat{\alpha}_n$ generally overestimates, while the Hill estimator $\hat{\alpha}_H$ underestimates the tail index α for the case that $\lambda_1 + \lambda_2$ is relatively low. Similar to the GARCH (1,1)-normal model, the Hill estimator $\hat{\alpha}_H$ performs well for the case that the GARCH series is close to Nonstationarity.

As a conclusion, comparing the implied tail index with the Hill estimate yield a valid test on the null hypotheses that the GARCH innovations follow a particular distribution, under which the implied tail index is calculated. This method is efficient particularly when the GARCH series is close to Nonstationarity.

Table 3: Tail index estimates of Real Data

	GARCH-normal		GARCH-Student			Tail Index			p -Value	
	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\nu}$	$\hat{\alpha}_n$	$\hat{\alpha}_s$	$\hat{\alpha}_H$	p_n	p_s
S&P	0.0770	0.9174	0.0766	0.9225	6.3377	3.92	2.18	3.05	0.01	0.01
Auto	0.0616	0.9290	0.0655	0.9284	5.9823	6.48	3.34	3.00	0.00	0.29
BioTech	0.0608	0.9346	0.0536	0.9455	5.9461	4.48	2.34	3.40	0.00	0.00
Media	0.0693	0.9257	0.0675	0.9285	7.8790	4.10	3.10	3.00	0.00	0.76
Chemical	0.0818	0.9182	0.0747	0.9238	6.6022	2.02	2.32	2.97	0.01	0.07
Pharm	0.0755	0.9152	0.0798	0.9152	5.9035	5.10	2.80	3.25	0.00	0.12
Retail	0.0649	0.9298	0.0594	0.9373	7.5191	4.48	3.12	3.12	0.00	1.00
SoftWare	0.0596	0.9356	0.0562	0.9438	6.3752	4.66	2.00	3.57	0.00	0.00
Transport	0.0587	0.9365	0.0664	0.9260	6.7542	4.76	3.74	3.38	0.00	0.37
Aero	0.0739	0.9197	0.0727	0.9198	6.7959	4.30	3.50	3.37	0.01	0.73
Steel	0.0525	0.9418	0.0541	0.9433	7.0444	5.84	3.00	2.57	0.00	0.06
Insurance	0.0925	0.9041	0.0891	0.9109	5.5377	2.86	2.02	2.33	0.07	0.30
Bank	0.0870	0.9130	0.0863	0.9137	7.1837	2.02	2.00	2.60	0.06	0.05

Note: This table presents the results of S&P 500 index and 12 equity indices from US market (01.01.1995-31.12.2010). With the coefficients estimated from the Maximum Likelihood method, we calculate the implied tail indices $\hat{\alpha}_n$, $\hat{\alpha}_s$. $\hat{\alpha}_H$ is the Hill estimate. p_n , p_s are the corresponding p -values based on the test statistics in Theorem 5 under 95% confidence level.

4.2 Empirical Study on Equity Indices

In this section, we apply our method to detect the distribution of the GARCH innovations once fitting the GARCH model to real data. The dataset consists of S&P 500 Composite Index and 12 S&P equity sector indices. The time series of daily data runs from 1 January 1995 to 31 December 2010, with a sample of size 4174. All data are collected from Datastream.

We fit the data by both GARCH models with normal and Student-t innovations. With the estimated GARCH coefficients $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\nu}$, we report the implied tail indices $\hat{\alpha}_n$, $\hat{\alpha}_s$. We also estimate the tail index by the Hill estimator $\hat{\alpha}_H$. By applying the test in Theorem 5, we report the corresponding p -values of the hypothesis tests that either the normal innovation or the Student-t innovation is regarded as the null hypothesis.

The results are reported in Table 3.

Firstly, it is notable that for all 13 series, $\hat{\lambda}_1 + \hat{\lambda}_2$ under any model is close to 1. Hence, the GARCH series, no matter which type of innovations, are close to Nonstationarity. This is exactly the case in which the MF2000 method based on estimated innovations fails, while our method based on comparing the implied tail indices with the Hill estimates produces robust testing result. Next we observe that $\hat{\alpha}_n$ is generally higher than $\hat{\alpha}_s$ in all indices except Chemical and Banking sectors. Furthermore, in the p -value columns, we observe that for 10 out of the 13 indices the null hypothesis of Student-t innovations are not rejected under 95% confidence level, whereas for 11 out of 13 indices the null hypothesis of normal innovations are rejected under the same confidence level.

Therefore, we conclude that the GARCH innovations of the equity indices are more likely to follow the Student-t distribution. One potential reason of such a preference is the fact that the Student-t distribution exhibits heavy-tails while the normal distribution does not. Hence, we draw the same conclusion as in MF2000 that assuming conditional heavy-tailedness is necessary when applying GARCH models to financial returns, albeit from a different approach.

5 Conclusion

This paper investigates the diagnosing method on the distribution of GARCH innovations. We first show that the method in McNeil and Frey [2000] based on estimated innovations is not reliable to detect the heavy-tailedness of the

GARCH innovations, particularly for the case that the GARCH process is close to Nonstationarity. Even though the observations are generated from a GARCH model with normal innovations, the estimated innovations from a finite sample follow a heavy-tailed distribution.

We provide an alternative approach on diagnosing the distribution of GARCH innovations by comparing implied tail indices with the estimate from the Hill estimator. A formal test is established from such an approach. In our method, the asymptotic bias of the Hill estimator may potentially contaminate the test. Nevertheless, the potential contamination is of less problematic once the GARCH process is close to Nonstationarity. The simulation results support our proposed approach.

We apply our method to 13 stock indices. The estimated GARCH coefficients in both models indicates *Near-Nonstationarity*. Hence, the MF2000 method is not valid for these data. Nevertheless, the results of our method confirm that a GARCH model with Student-t innovation fits the data better than a GARCH-normal model.

Our findings lead to important implication in dynamic risk management: when applying a GARCH model to evaluate conditional risk measures, such as VaR given the previous return level and volatility, conditional heavy-tailed distribution should be considered. Using a conditional normal model may underestimate the potential dynamic risk. Moreover, it is not proper to evaluate the heavy-tailedness based on the estimated innovations. That may overestimate the conditional risk measures. Lastly, to accurately evaluate the unconditional tail risk, it is better to consider the implied tail index instead of using Hill estimator because of a potential bias. The economic

significance on the difference between the two types of GARCH model is left for future research.

References

- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- Tim Bollerslev, Robert F. Engle, and Daniel B. Nelson. Arch models. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 49, pages 2959–3038. Elsevier, January 1986. URL <http://ideas.repec.org/h/eee/ecochnp/4-49.html>.
- M. Carrasco and X. Chen. Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39, 2002.
- J. Danielsson and C.G. De Vries. Value-at-risk and extreme returns. *Annales d’Economie et de Statistique*, pages 239–270, 2000.
- R.A. Davis and T. Mikosch. Extreme value theory for garch processes. *Handbook of Financial Time Series*, pages 187–200, 2009.
- L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45, 1998.
- C.M. Goldie and R.L. Smith. Slow variation with remainder: Theory and applications. *The Quarterly Journal of Mathematics*, 38(1):45, 1987.

- N. Hang Chan, S.J. Deng, L. Peng, and Z. Xia. Interval estimation of value-at-risk based on garch models with heavy-tailed innovations. *Journal of Econometrics*, 137(2):556–576, 2007.
- T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, pages 1547–1569, 1991.
- H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973.
- A.J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300, 2000.
- J. Meyer. Second degree stochastic dominance with respect to a function. *International Economic Review*, 18(2):477–487, 1977.
- T. Mikosch and C. Starica. Limit theory for the sample autocorrelations and extremes of a garch (1, 1) process. *Annals of Statistics*, pages 1427–1451, 2000.
- S. Resnick and C. Stărică. Tail index estimation for dependent data. *The Annals of Applied Probability*, 8(4):1156–1183, 1998.

6 Appendix

6.1 Proof of Proposition 1

Equation (7) can be rewritten as

$$\hat{\varepsilon}_t = \frac{X_t}{\hat{\sigma}_t} = \frac{\sigma_t}{\hat{\sigma}_t} \varepsilon_t.$$

Since ε_t follows a normal distribution and is independent from $\frac{\sigma_t}{\hat{\sigma}_t}$, it is only necessary to prove that $\frac{\sigma_t}{\hat{\sigma}_t}$ follows a heavy-tailed distribution.

By iteration we get that $\hat{\sigma}_t^2 - \sigma_t^2 = \lambda_2^t (\hat{\sigma}_0^2 - \sigma_0^2)$ and $\sigma_t^2 = A_t + B_t \sigma_0^2$, where A_t and B_t are independent from σ_0^2 and have the following forms

$$\begin{aligned} A_t &= \sum_{j=1}^t \lambda_0 \prod_{i=2}^j (\lambda_1 \varepsilon_{t-i+1}^2 + \lambda_2), \\ B_t &= \prod_{i=1}^t (\lambda_1 \varepsilon_{i-1}^2 + \lambda_2). \end{aligned}$$

Hence, we have that

$$\begin{aligned} P\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x\right) &= P(A_t + B_t \sigma_0^2 > x(\lambda_2^t (\hat{\sigma}_0^2 - \sigma_0^2) + A_t + B_t \sigma_0^2)) \\ &= P([B_t - x(B_t - \lambda_2^t)] \sigma_0^2 > (x-1)A_t + x\lambda_2^t \hat{\sigma}_0^2). \end{aligned}$$

We further derive a lower bound of this probability by conditional on $B_t < \frac{x+1}{x} \lambda_2^t$, which guarantees that $B_t - x(B_t - \lambda_2^t) > \frac{\lambda_2^t}{x}$. We get that

$$P\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x\right) > P\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x, B_t < \frac{x+1}{x} \lambda_2^t\right)$$

$$\begin{aligned}
&> P\left(\frac{\lambda_2^t}{x}\sigma_0^2 > (x-1)A_t + x\lambda_2^t\hat{\sigma}_0^2, B_t < \frac{x+1}{x}\lambda_2^t\right) \\
&> P\left(\frac{\lambda_2^t}{x}\sigma_0^2 > x(A_t + \lambda_2^t\hat{\sigma}_0^2), B_t < \frac{x+1}{x}\lambda_2^t\right) \\
&= P\left(\sigma_0^2 > x^2\left(\frac{A_t}{\lambda_2^t} + \hat{\sigma}_0^2\right), B_t < \frac{x+1}{x}\lambda_2^t\right).
\end{aligned}$$

To continue with the calculation, we study the relation between A_t and B_t as follows. From $A_t = A_{t-1}(\lambda_1\varepsilon_{t-1}^2 + \lambda_2) + \lambda_0$ and $B_t = B_{t-1}(\lambda_1\varepsilon_{t-1}^2 + \lambda_2)$, we get that $A_t = A_{t-1}\frac{B_t}{B_{t-1}} + \lambda_0$. Hence we derive an upper bound for $\frac{A_t}{B_t}$ as

$$\frac{A_t}{B_t} = \frac{A_{t-1}}{B_{t-1}} + \frac{\lambda_0}{B_t} = \dots = \sum_{i=1}^t \frac{\lambda_0}{B_i} \leq \sum_{i=1}^t \frac{\lambda_0}{\lambda_2^i} = \lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{t-1} - \lambda_2}.$$

Therefore, we continue that calculation on the tail distribution function of $\frac{\sigma_t}{\hat{\sigma}_t}$ as

$$\begin{aligned}
P\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x\right) &> P\left(\sigma_0^2 > x^2\left(\lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{2t-1} - \lambda_2^{2t}} B_t + \hat{\sigma}_0^2\right), B_t < \frac{x+1}{x}\lambda_2^t\right) \\
&= E_{B_t} 1_{B_t < \frac{x+1}{x}\lambda_2^t} P\left(\sigma_0^2 > x^2\left(\lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{2t-1} - \lambda_2^{2t}} B_t + \hat{\sigma}_0^2\right) \mid B_t\right) \\
&\sim E_{B_t} 1_{B_t < \frac{x+1}{x}\lambda_2^t} \frac{C}{x^\kappa \left(\lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{2t-1} - \lambda_2^{2t}} B_t + \hat{\sigma}_0^2\right)^\kappa}, \text{ as } x \rightarrow +\infty.
\end{aligned}$$

The last step comes from the facts that σ_0^2 is independent from B_t and follows a heavy-tailed distribution with tail index κ . Here C is the tail scale of σ_0^2 . Under the condition $B_t < \frac{x+1}{x}\lambda_2^t$, we have that

$$E_{B_t} 1_{B_t < \frac{x+1}{x}\lambda_2^t} \frac{C}{x^\kappa \left(\lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{2t-1} - \lambda_2^{2t}} B_t + \hat{\sigma}_0^2\right)^\kappa}$$

$$\begin{aligned}
&\geq \frac{C}{x^\kappa} \left(\lambda_0 \frac{1 - \lambda_2^{t-1}}{\lambda_2^{t-1} - \lambda_2^t} \frac{x+1}{x} + \hat{\sigma}_0^2 \right)^{-\kappa} E_{B_t} 1_{B_t < \frac{x+1}{x} \lambda_2^t} \\
&= \frac{C_1}{x^\kappa} P \left(B_t < \frac{x+1}{x} \lambda_2^t \right),
\end{aligned}$$

where C_1 is a positive constant. For the part $P(B_t < \frac{x+1}{x} \lambda_2^t)$, we use the property of the cumulative distribution function of a χ^2 distributed random variable to simplify the calculation as follows:

$$\begin{aligned}
P \left(B_t < \frac{x+1}{x} \lambda_2^t \right) &= P \left(\log B_t - t \log \lambda_2 < \log \frac{x+1}{x} \right) \\
&= P \left(\sum_{i=1}^t \log \left(1 + \frac{\lambda_1}{\lambda_2} \varepsilon_i^2 \right) < \log \frac{x+1}{x} \right) \\
&> P \left(\sum_{i=1}^t \frac{\lambda_1}{\lambda_2} \varepsilon_i^2 < \log \frac{x+1}{x} \right) \\
&> \frac{\frac{2}{t} \left(\frac{\lambda_2}{2\lambda_1} \log \frac{x+1}{x} \right)^{t/2}}{\Gamma(t/2)} \\
&= C_2 \left(\log \left(1 + \frac{1}{x} \right) \right)^{t/2} \sim C_2 x^{-t/2}, \text{ as } x \rightarrow +\infty.
\end{aligned}$$

Therefore, we have that

$$\liminf_{x \rightarrow +\infty} P \left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x \right) x^{(\kappa+t/2)} \geq C_1 C_2.$$

Hence, $\frac{\sigma_t^2}{\hat{\sigma}_t^2}$ follows a heavy-tailed distribution, which implies that $\frac{\sigma_t}{\hat{\sigma}_t}$ follows a heavy-tailed distribution. This completes the proof of the lemma.

6.2 Proof of Lemma 4

We first show that $\kappa_n(\lambda_1, \lambda_2) > \kappa_s(\lambda_1, \lambda_2, \nu)$ for any degree of freedom ν finite. We start with the following Proposition².

Proposition 9 *Let $\psi = Z_\nu^2$ and $\phi = Z_\mu^2$, where Z_ν and Z_μ follow standardized Student-t distribution with degree of freedoms ν and μ respectively. Then $E(\psi) = E(\phi) = 1$. Denote the cumulative distribution functions of ψ and ϕ as F_ψ and F_ϕ , then $D(x) = \int_0^x F(\psi \leq x)dx - \int_0^x F(\phi \leq x)dx \geq 0$, $\forall t > 0$ and $2 < \nu < \mu < +\infty$.*

Proposition 9 states that ϕ second order stochastically dominates over ψ . By taking $\mu \rightarrow +\infty$, we get the following Corollary.

Corollary 10 *Let φ be a random variable following the $\chi^2(1)$ distribution. φ second order stochastically dominates over ψ for any finite ν .*

Since the function $g(s) = (\lambda_1 s + \lambda_2)^\kappa$ is increasing and convex with respect to s for all $\kappa > 1$, following the property of second order stochastic dominance (see e.g. Meyer [1977]), we have that $E[(\lambda_1 \varphi + \lambda_2)^{\kappa_s}] < E[(\lambda_1 \psi + \lambda_2)^{\kappa_s}] = 1$. Because $E[(\lambda_1 \varphi + \lambda_2)^\kappa] < 1$ for $0 < \kappa < \kappa_n$ and $E[(\lambda_1 \varphi + \lambda_2)^\kappa] > 1$ for $\kappa > \kappa_n$ (see proof of Lemma 7), hence we get that $\kappa_n(\lambda_1, \lambda_2) > \kappa_s(\lambda_1, \lambda_2, \nu)$, which implies that $\alpha_n(\lambda_1, \lambda_2) > \alpha_s(\lambda_1, \lambda_2, \nu)$ for any degree of freedom ν finite. From the consistency that $\alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}) \xrightarrow{P} \alpha(\lambda_1, \lambda_2, \nu)$ as $T \rightarrow +\infty$. It implies that $P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \rightarrow 1$ as $T \rightarrow +\infty$.

²The proof of Proposition 9 is available upon request.

6.3 Proof of Theorem 5

Proof of Statement 1: combining the two facts that the estimated parameters $\hat{\lambda}_1, \hat{\lambda}_2$ have a speed of convergence \sqrt{T} and the intermediate sequence k satisfies $k \rightarrow \infty, \frac{k}{T} \rightarrow 0$ as $T \rightarrow +\infty$, we get that $\sqrt{k}(\hat{\lambda}_1 - \lambda) \xrightarrow{P} 0$ and $\sqrt{k}(\hat{\lambda}_2 - \lambda) \xrightarrow{P} 0$ as $T \rightarrow +\infty$.

With the Taylor expansion

$$\kappa_n(\hat{\lambda}_1, \hat{\lambda}_2) = \kappa(\lambda_1, \lambda_2) + \frac{\partial \kappa}{\partial \lambda_1}(\hat{\lambda}_1 - \lambda_1) + \frac{\partial \kappa}{\partial \lambda_2}(\hat{\lambda}_2 - \lambda_2) + o(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2),$$

we get that $\sqrt{k}(\kappa_n(\hat{\lambda}_1, \hat{\lambda}_2) - \kappa(\lambda_1, \lambda_2)) \xrightarrow{P} 0$ as $T \rightarrow +\infty$ provided that $\left| \frac{\partial \kappa}{\partial \lambda_1} \right|$ and $\left| \frac{\partial \kappa}{\partial \lambda_2} \right|$ are bounded.

We prove the boundedness of the two partial derivatives separately. By taking the partial derivative of equation (6) with respect to λ_1 , we get that

$$E[(\lambda_1 \varphi + \lambda_2)^\kappa] \left(\frac{\kappa}{\lambda_1 \varphi + \lambda_2} \varphi + \log(\lambda_1 \varphi + \lambda_2) \frac{\partial \kappa}{\partial \lambda_1} \right) = 0,$$

which implies that

$$E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1} \kappa \varphi] = -E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)] \frac{\partial \kappa}{\partial \lambda_1}.$$

Hence we get that

$$\begin{aligned} \left| \frac{\partial \kappa}{\partial \lambda_1} \right| &= \left| \frac{E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1} \kappa \varphi]}{E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)]} \right| \\ &= \left| \frac{\frac{\kappa}{\lambda_1} \{E[(\lambda_1 \varphi + \lambda_2)^\kappa] - \lambda_2 E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1}]\}}{E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)]} \right|. \end{aligned}$$

The conditions $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_1 + \lambda_2 < 1$ implies that $\kappa > 1$. The proof of Lemma 7 shows that there exists a lower bound $D_0 > 0$ such that $E[(\lambda_1\varphi + \lambda_2)^\kappa \log(\lambda_1\varphi + \lambda_2)] > D_0$. Thus $\left| \frac{\partial \kappa}{\partial \lambda_1} \right| < \frac{\kappa(1+\lambda_2)}{\lambda_1 D_0}$.

Following the similar procedure, we take the partial derivative of (6) with respect to λ_2 to obtain that

$$E[(\lambda_1\varphi + \lambda_2)^\kappa] \left(\frac{\kappa}{\lambda_1\varphi + \lambda_2} + \log(\lambda_1\varphi + \lambda_2) \frac{\partial \kappa}{\partial \lambda_2} \right) = 0,$$

which implies that

$$\left| \frac{\partial \kappa}{\partial \lambda_2} \right| = \left| \frac{\kappa E[(\lambda_1\varphi + \lambda_2)^{\kappa-1}]}{E[(\lambda_1\varphi + \lambda_2)^\kappa \log(\lambda_1\varphi + \lambda_2)]} \right| < \frac{\kappa}{D_0}.$$

With the bounded partial derivatives, we conclude that

$$\sqrt{k} \left(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) - \alpha \right) \xrightarrow{P} 0 \text{ as } T \rightarrow +\infty. \quad (15)$$

The asymptotic normality of the Hill Estimator follows from Hsing [1991].

The Hill estimator converges to the tail index α with speed of convergence

\sqrt{k} . The asymptotic limit is given as

$$\sqrt{k} \left(\frac{\alpha}{\hat{\alpha}_H} - 1 \right) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi). \quad (16)$$

Therefore,

$$\begin{aligned} \sqrt{k} \left(\frac{\hat{\alpha}_n}{\hat{\alpha}_H} - 1 \right) &= \sqrt{k} \left[\left(\frac{\hat{\alpha}_n}{\hat{\alpha}_H} - \frac{\alpha}{\hat{\alpha}_H} \right) + \left(\frac{\alpha}{\hat{\alpha}_H} - 1 \right) \right] \\ &= \frac{\sqrt{k}(\hat{\alpha}_n - \alpha)}{\hat{\alpha}_H} + \sqrt{k} \left(\frac{\alpha}{\hat{\alpha}_H} - 1 \right). \end{aligned}$$

Notice that $\hat{\alpha}_H \xrightarrow{p} \alpha$ as $T \rightarrow \infty$. With (15) and (16), we get that as $T \rightarrow \infty$,

$$\sqrt{k} \left(\frac{\hat{\alpha}_n}{\hat{\alpha}_H} - 1 \right) \xrightarrow{d} N(0, 1 + \chi + \omega - 2\psi). \quad (17)$$

Proof of Statement 2: the proof follows the similar lines as that in the proof of Statement 1. We only need to check the boundedness of $\left| \frac{\partial \kappa}{\partial \nu} \right|$. We show that for any given values $\nu, \nu_0 > 2, \lambda_1, \lambda_2 > 0$ and $\lambda_1 + \lambda_2 < 1$, $\left| \frac{\kappa(\lambda_1, \lambda_2, \nu) - \kappa(\lambda_1, \lambda_2, \nu_0)}{\nu - \nu_0} \right|$ is bounded.

Without loss of generality, we assume that $\nu > \nu_0$. Suppose that three random variables $X \sim \chi^2(1)$, $\Delta Y \sim \chi^2(\nu - \nu_0)$ and $Y_0 \sim \chi^2(\nu_0)$ are independent. By using the convolution of density functions, we get that $Y := Y_0 + \Delta Y \sim \chi^2(\nu)$. Therefore, we have that $\sqrt{\frac{X}{Y/(\nu-2)}}$ follows a standardized $t(\nu)$ and $\sqrt{\frac{X}{Y_0/(\nu_0-2)}}$ follows a standardized $t(\nu_0)$ distribution, both with variance 1. Recall equation (6), we have that,

$$\begin{aligned} E\left[\left(\lambda_1 \frac{X}{Y/(\nu-2)} + \lambda_2\right)^{\kappa(\lambda_1, \lambda_2, \nu)}\right] &= 1, \\ E\left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2\right)^{\kappa_0(\lambda_1, \lambda_2, \nu_0)}\right] &= 1. \end{aligned}$$

Hence,

$$\begin{aligned} 0 &= E\left[\left(\lambda_1 \frac{X}{Y/(\nu-2)} + \lambda_2\right)^{\kappa(\lambda_1, \lambda_2, \nu)}\right] - E\left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2\right)^{\kappa_0(\lambda_1, \lambda_2, \nu_0)}\right] \\ &= \left\{ E\left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2\right)^{\kappa(\lambda_1, \lambda_2, \nu)}\right] - E\left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2\right)^{\kappa_0(\lambda_1, \lambda_2, \nu_0)}\right] \right\} \\ &\quad + \left\{ E\left[\left(\lambda_1 \frac{X}{Y/(\nu-2)} + \lambda_2\right)^{\kappa(\lambda_1, \lambda_2, \nu)}\right] - E\left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2\right)^{\kappa(\lambda_1, \lambda_2, \nu)}\right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \{E \left[\left(\lambda_1 \frac{X}{Y/(\nu-2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)} \right] - E \left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)} \right]\} \\
& =: I_1 + I_2 + I_3.
\end{aligned}$$

We deal with three parts separately. First, for I_1 , by applying the Mean Value Theorem (MVT), there exists $\tilde{\kappa}$ between κ_0 and κ , such that

$$\begin{aligned}
I_1 &= E \left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2 \right)^{\tilde{\kappa}} \right] \log \left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2 \right) (\kappa(\lambda_1, \lambda_2, \nu) - \kappa_0(\lambda_1, \lambda_2, \nu_0)) \\
&=: J_1 (\kappa(\lambda_1, \lambda_2, \nu) - \kappa_0(\lambda_1, \lambda_2, \nu_0)) \geq D_1 (\kappa(\lambda_1, \lambda_2, \nu) - \kappa_0(\lambda_1, \lambda_2, \nu_0)).
\end{aligned}$$

Secondly, for I_2 , by applying the MVT, there exists ξ between Y_0 and Y , such that

$$\begin{aligned}
-I_2 &= \kappa E \left[\left(\lambda_1 \frac{X}{\xi/(\nu_0-2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)-1} \right] \left(\frac{\lambda_1 X (\nu_0 - 2)}{\xi^2} \right) (Y - Y_0) \\
&\leq \kappa E \left[\left(\lambda_1 \frac{X}{Y_0/(\nu_0-2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)-1} \right] \left(\frac{\lambda_1 X}{Y_0^2/(\nu_0-2)} \right) E(Y - Y_0) \\
&=: \kappa J_2 E(Y - Y_0) \leq D_2 (\nu - \nu_0).
\end{aligned}$$

Thirdly, for I_3 , by applying the MVT, there exists $\tilde{\nu}$ between ν_0 and ν such that

$$\begin{aligned}
I_3 &= \kappa E \left[\left(\lambda_1 \frac{X}{Y/(\tilde{\nu}-2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)-1} \right] \frac{\lambda_1 X}{Y} (\nu - \nu_0) \\
&=: J_3 (\nu - \nu_0) > D_3 (\nu - \nu_0).
\end{aligned}$$

Here D_1 , D_2 and D_3 are the lower bounds of J_1 , J_2 and J_3 ³. Combining

³The proof on the existence of such lower bounds are available upon request.

these three parts, we get that

$$D_1(\kappa(\lambda_1, \lambda_2, \nu) - \kappa_0(\lambda_1, \lambda_2, \nu_0)) + D_3(\nu - \nu_0) \leq I_1 + I_3 = -I_2 \leq D_2(\nu - \nu_0),$$

which implies that

$$\left| \frac{\kappa(\lambda_1, \lambda_2, \nu) - \kappa_0(\lambda_1, \lambda_2, \nu_0)}{\nu - \nu_0} \right| \leq \left| \frac{D_2 - D_3}{D_1} \right|.$$

Therefore, we get that $\left| \frac{\partial \kappa}{\partial \nu} \right|$ is bounded.

6.4 Proof of Lemma 7

Suppose $F(x) = P\{\sigma_t^2 \leq x\}$ satisfies (10). Denote $\bar{F}(x) = 1 - F(x)$. Let $f(\varphi)$ be the density function of the random variable φ . Notice that $\varphi_t = \varepsilon_t^2$ follows a $\chi^2(1)$ distribution. Then,

$$\begin{aligned} P\{\sigma_t^2 > x\} &= P\{\lambda_0 \varphi_t + (\lambda_1 \varphi_t + \lambda_2) \sigma_{t-1}^2 > x\} \\ &= E \left[P\{\sigma_{t-1}^2 > \frac{x - \lambda_0}{\lambda_1 \varphi_t + \lambda_2} \mid \lambda_0, \lambda_1 \varphi_t + \lambda_2\} \right] \\ &= \int_0^\infty \bar{F} \left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right) f(\varphi) d\varphi. \end{aligned}$$

Consider the term $\bar{F} \left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)$ which has the following expansion

$$\begin{aligned} \bar{F} \left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right) &= A \left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)^{-\kappa} \left[1 + B \left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)^{-\beta} + o \left(\left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)^{-\beta} \right) \right] \\ &= A (\lambda_1 \varphi + \lambda_2)^\kappa x^{-\kappa} \left(1 - \frac{\lambda_0}{x} \right)^{-\kappa} \left[1 + B (\lambda_1 \varphi + \lambda_2)^\beta x^{-\beta} \left(1 - \frac{\lambda_0}{x} \right)^{-\beta} \right. \\ &\quad \left. + o \left(\left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)^{-\beta} \right) \right]. \end{aligned}$$

With the Taylor expansion that

$$\left(1 - \frac{\lambda_0}{x}\right)^{-\kappa} = 1 + \kappa \frac{\lambda_0}{x} + \frac{\kappa(1+\kappa)}{2} \left(\frac{\lambda_0}{x}\right)^2 + o\left(\left(\frac{\lambda_0}{x}\right)^2\right),$$

the above equation is simplified as

$$\bar{F}\left(\frac{x - \lambda_0}{\lambda_1\varphi + \lambda_2}\right) = A(\lambda_1\varphi + \lambda_2)^\kappa x^{-\kappa} \left[1 + B(\lambda_1\varphi + \lambda_2)^\beta x^{-\beta} + \kappa\lambda_0 x^{-1} + \max\{o(x^{-\beta-1}), o(x^{-2})\}\right],$$

Hence we get that

$$\begin{aligned} \int_0^\infty \bar{F}\left(\frac{x - \lambda_0}{\lambda_1\varphi + \lambda_2}\right) f(\varphi) d\varphi &= Ax^{-\kappa} [E[(\lambda_1\varphi + \lambda_2)^\kappa] + BE[(\lambda_1\varphi + \lambda_2)^{\kappa+\beta}] x^{-\beta} \\ &\quad + \kappa\lambda_0 E[(\lambda_1\varphi + \lambda_2)^{\kappa+1}] x^{-1} + \max\{o(x^{-\beta-1}), o(x^{-2})\}]. \end{aligned}$$

Comparing with the expansion of $\bar{F}(x)$ as $\bar{F}(x) = Ax^{-\kappa}[1 + Bx^{-\beta} + o(x^{-\beta})]$,

if $\beta > 1$, then the second order terms on both sides are of unequal order; if

$\beta < 1$, then we have that $E[(\lambda_1\varphi + \lambda_2)^{\kappa+\beta}] = 1$, which is contradictory with

the fact that $E[(\lambda_1\varphi + \lambda_2)^\kappa] = 1$.

Hence we conclude that the second order index $\beta = 1$ and the second order tail scale B has the unique form as in (14).

Next we show that the second order tail scale B must be negative. Consider the function $g(s) = E[(\lambda_1\varphi + \lambda_2)^s]$. We have that $g''(s) = E[(\lambda_1\varphi + \lambda_2)^s \log^2(\lambda_1\varphi + \lambda_2)] \geq 0$, thus $g(s)$ is a convex function, with $g(0) = g(\kappa) = 1$. $g'(s)$ is monotonic and continuous, there exists $\bar{s} \in [0, \kappa]$, such that $g'(\bar{s}) = 0$. Because $g'(s)$ is right continuous at 0, we have that $\lim_{s \downarrow 0} g'(s) = E[\log(\lambda_1\varphi + \lambda_2)] \leq \log(\lambda_1 E(\varphi) + \lambda_2) = \log(\lambda_1 + \lambda_2) < 0$, hence $g'(s) > 0$ for $s \geq \bar{s}$. Therefore, $g(s)$ is an increasing function for $s \geq \bar{s}$, which implies

that $g(\kappa + 1) > 1$. This implies that $B < 0$.