Diagnosing the Distribution of GARCH Innovations

Pengfei SUN
Department of Economics, Erasmus University Rotterdam, P.O.Box 1738, 3000DR, The Netherlands
(Email: sun@ese.eur.nl)

Chen ZHOU
Economics and Research Division, De Nederlandsche Bank, P.O.Box 98, 1000AB, The Netherlands
Department of Economics, Erasmus University Rotterdam, P.O.Box 1738, 3000DR, The Netherlands
(Email: c.zhou@dnb.nl)

This version: 07 May 2013

Abstract
The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, designed to model volatility clustering, exhibits heavy-tailedness regardless of the distribution of on its innovation term. When applying the model to financial time series, the distribution of innovations plays an important role for risk measurement and option pricing. We investigate methods on diagnosing the distribution of GARCH innovations. For GARCH processes that are close to integrated-GARCH (IGARCH), we show that the method based on estimated innovations is not reliable, whereas an alternative approach based on analyzing the tail index of a GARCH series performs better. The alternative method leads to a statistical test on the distribution of GARCH innovations.

Keywords: GARCH(1,1), Extreme Value Theory, Hill Estimator, Dynamic Risk Management.
JEL Classification: C12, C58, G32

1Views expressed do not reflect official positions of De Nederlandsche Bank.
1 Introduction

Volatility of financial time series plays a key role in both risk management and option pricing. For asset return series, such as the returns of equity prices and foreign exchange rates, volatility clustering has been observed. That is, once a high volatility occurs, it persists for a while. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev [1986] is designed to capture the volatility clustering feature by modeling the dynamic of volatility. In each period, the return is modeled as a new innovation term with variance scaled up by the information of returns and volatilities in the past periods. When considering the volatility dynamic with only one lagged period, the GARCH(1,1) model has become a workhorse in both academic and practice due to its simplicity and intuitive interpretation.

When applying GARCH models in financial risk management, the distribution of GARCH innovations plays an important role. From the definition of GARCH model, it is obvious that the conditional distribution of future returns has the same shape as the distribution of the innovations. Therefore, an inappropriate model on the distribution of innovations may lead to either underestimation or overestimation of future risks. In addition, different distributions of GARCH innovations may also lead to different option pricing results. This paper examines an existing diagnosis framework on the distribution of GARCH innovations, and demonstrates its drawback when applying to financial time series. Further, we develop an alternative method, particularly towards applications to financial time series.

Our diagnosing method is build on the tail properties of GARCH models. A surprising consequence of the GARCH-type models is that they exhibit heavy-tailedness. Heavy tails in financial time series refer to the fact that the downside tail of the distribution decays in a power-law speed as oppose to the exponential speed as in that of normal distribution, see e.g. Mandelbrot [1963] and Jansen and De Vries [1991]. Following the result in Kesten [1973], the stationary solution of GARCH(1,1) process follows a heavy-tailed distribution, see e.g. Mikosch and Starica [2000], Davis and Mikosch [2009]. Although this result holds regardless of the distribution of innovations, as we will show, GARCH models with different innovation distributions lead to different levels of heavy-tailedness in the downside tail, hence different levels of tail risk in GARCH series.

The most often applied GARCH(1,1) model assumes that the innovation term follows a standard normal distribution. Recent literature in risk management shows that the conditional normality assumption does not perform well in evaluating 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.

The literature for the GARCH model in option pricing also documented the usefulness of deviating from the normal assumption on innovations. The GARCH model with normal distributed innovations has been applied to option pricing in, for instance, Duan [1995] and Heston and Nandi [2000]. Recent studies on GARCH option pricing models investigates non-normal innovations and find that GARCH models with non-normal innovations outperform the ones with normal innovations in pricing options, see, e.g. Menn and Rachev [2005] and Badescu and Kulperger [2008], and Barone-Adesi et al. [2008]. In addition, Duan [1999] argues that GARCH models with heavy-tailed and skewed innovations are able to capture the volatility smile.

The existing diagnostic method on the distribution of GARCH innovations is usually based on estimated innovations, see e.g. McNeil and Frey [2000], Mikosch and Starica [2000]. Such a diagnosis procedure can be divided into three steps as in McNeil and Frey [2000]: first, they get the Quasi-Maximum-Likelihood-Estimators (QMLE) (see e.g. Elie and Jeantheau [1995]) of GARCH coefficients and back out the innovations; second, they use model diagnosis tools such as the QQ plot to check whether the innovations fit a standard normal distribution; third, with evidence that the innovations follow a heavy-tailed or skewed distribution, they use specific distribution, such as the Generalized Pareto Distribution to fit the tail of the estimated innovations and consequently estimate the tail shape of the distribution of the innovations. We call such a procedure the backing-out method. This method has been followed by other studies, see e.g. Menn and Rachev [2005], Hang Chan et al. [2007] and Badescu and Kulperger [2008].

In this paper, we start by revisiting the backing-out method. We simulate GARCH series with normal innovation and apply the backing-out method to the simulated data. When the simulated GARCH process is close to an integrated-GARCH (IGARCH) model (for the definition of IGARCH, see Engle and Bollerslev [1986]), the backing-out method concludes that the innovations are heavy-tailed. This is against the initial normal distribution used in the simulation. Hence the backing-out method is not robust in diagnosing the distribution of GARCH innovations. When fitting the financial asset returns to a GARCH model, the fitted model is often close to an IGARCH model, see e.g. Mikosch and Starica [2003] and Berkes et al. [2005]. We call this phenomenon as Near-IGARCH. Thus, the backing-out method is particularly not robust when
applying to Near-IGARCH series, such as financial returns. Next, we provide theoretical reason to explain why this can occur. Lastly, we develop a diagnosis method based on analyzing the tail index of the GARCH process. Our method yields a formal test on the distribution of innovations. Taking normal and Student-t distributed innovations as examples, simulations show that our method is robust in the case of Near-IGARCH. 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 Near-IGARCH. With our formal test, we reject the normal innovation in most cases, while can not always reject the hypothesis on Student-t distributed innovations.

The rest of the paper is organized as follows. In Section 2, we show that the backing-out approach based on estimated innovations is not robust for diagnosing the distribution of GARCH 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. Simulation results are presented in Section 4. An empirical example in finance is in Section 5. Section 6 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:

\[
X_t = \varepsilon_t \sigma_t, \tag{1}
\]

\[
\sigma_t^2 = \lambda_0 + \lambda_1 X_{t-1}^2 + \lambda_2 \sigma_{t-1}^2, \tag{2}
\]

where \(\{\varepsilon_t\} \) are independent and identically distributed (i.i.d.) innovations with zero mean and unit variance, the parameters \(\lambda_0, \lambda_1, \lambda_2 \) are positive. Moreover, in order to have a stationary solution of the GARCH model, we assume \(\lambda_1 + \lambda_2 < 1 \).

The heavy-tailedness of the stationary solution of a GARCH model follows from the result of Kesten
[1973]. Consider a process \( \{Y_t\} \) satisfying a stochastic difference equation

\[
Y_t = Q_t Y_{t-1} + M_t,
\]

where \( \{(Q_t, M_t)\} \) are i.i.d. \( \mathbb{R}^2_+ \)-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 \( \kappa \), such that

\[
E(Q_t^\kappa) = 1, \quad E(Q_t^\kappa \log Q_t) < \infty, \quad 0 < E(M_t^\kappa) < \infty.
\]

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

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

where \( \kappa \) 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 \( \{\sigma_t^2\} \) as

\[
\sigma_t^2 = \lambda_0 + (\lambda_1 \epsilon_{t-1}^2 + \lambda_2)\sigma_{t-1}^2,
\]

which satisfies equation (3) with \( Q_t = \lambda_1 \epsilon_{t-1}^2 + \lambda_2 \) and \( M_t = \lambda_0 \). Denote \( \varphi \overset{d}{=} \epsilon_{t-1}^2 \). Suppose \( \kappa \) is the solution to the equation

\[
E[(\lambda_1 \varphi + \lambda_2)^\kappa] = 1.
\]

The stationary solution of \( \sigma_t^2 \) follows a heavy-tailed distribution with tail index \( \kappa \). Hence, \( \sigma_t \) follows a heavy-tailed distribution with tail index \( 2\kappa \). The relation (6) implies that \( E[\epsilon_t^{2\kappa}] < \infty \). Therefore, the tail of \( \sigma_t \) is heavier than that of \( \epsilon_t \): if \( \epsilon_t \) follows a thin-tailed distribution such as the normal distribution, \( \sigma_t \) follows a heavy-tailed distribution; if \( \epsilon_t \) follows a heavy-tailed distribution, then \( \sigma_t \) follows a distribution with a heavier tail.

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

\[
P(\{|X_t| > x\} = P(\{|\sigma_t \epsilon_t| > x\} \sim E[|\epsilon_t|^2]^{2\kappa} P(\sigma_t > x), \text{as} \ x \to \infty.
\]
Thus $|X_t|$ has a similar tail behavior as $\sigma_t$: the tail index of $|X_t|$ equals to $2\kappa$.

From the discussion, we observe that the general heavy-tailed feature of the GARCH model does not depend on the distribution of GARCH innovations. 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 the innovations: the solution to equation (6) differs for different distributions of $\varphi$. Hence different distributions of GARCH innovations may lead to different levels of heavy-tailedness of a GARCH series. This phenomenon motivates the study on diagnosing the distribution of GARCH innovation when fitting GARCH models to actual data.

### 2.2 The Backing-out Approach

As an example of the backing-out method, McNeil and Frey [2000] find that when modeling financial time series by a GARCH model, the innovations are heavy-tailed by making QQ plot on the estimated innovations against a normal distribution. We argue that since the innovations are backed out from an estimation procedure, its heavy-tailedness may potentially be imposed by the backing-out procedure. We first demonstrate this phenomenon by simulation.

Given the estimates of the GARCH coefficients $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2$, 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 McNeil and Frey [2000], we reproduce their QQ plot in Figure 1 (left panel). In addition, we generate a series of observations from a GARCH(1,1)-normal process with the model coefficients equivalent to the MLE obtained from the real data. Then we re-estimate the innovations for the simulated data by (7) & (8) and display the corresponding QQ plot in Figure 1 (right panel). We find that the estimated innovations from the generated data also violate the normality assumption, while exhibiting a heavy-tailed feature. 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 robust for testing the distribution 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

---

2The coefficients can be either estimated by QMLE (see Elie and Jeantheau [1995]) or MLE (see Bollerslev et al. [1986]).
Figure 1: QQ-Plots

![QQ-Plots](image)

**Note:** The QQ-plots show the estimated innovations against the sample quantiles from a normal distribution. If the plotted scatters fall on the straight line, the estimated innovations follow a normal distribution. The left panel demonstrates the estimated innovations from fitting the S&P 500 daily returns (from June 1985 to May 1989) to a GARCH(1,1) model by MLE. The right panel demonstrates the estimated innovations from fitting a simulated sample generated by a GARCH(1,1)-normal model, with the coefficients equivalent to the MLE obtained from actual data.

Combining equations (2) and (8), we have that

\[
\hat{\sigma}_t^2 - \sigma_t^2 = \lambda_2 (\hat{\sigma}_{t-1}^2 - \sigma_{t-1}^2) = \ldots = \lambda_2^t (\hat{\sigma}_0^2 - \sigma_0^2).
\]

Hence, \(\sigma_t^2\) 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 \(\sigma_t^2\) stems from that between \(\hat{\sigma}_0^2\) and \(\sigma_0^2\). Moreover, the difference between \(\hat{\sigma}_0^2\) and \(\sigma_0^2\) comes from the fact that \(\hat{\sigma}_0^2\) is simply some initial value chosen in estimation, while \(\sigma_0^2\) follows the stationary solution of the stochastic difference equation (5), i.e. a heavy-tailed distribution. If the GARCH process is Near-IGARCH, i.e. \(\lambda_1 + \lambda_2\) is close to 1, then the parameter \(\kappa\) 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 \(\sigma_0^2\) substantially, which implies that \(\hat{\sigma}_t^2\) underestimates \(\sigma_t^2\) with a finite \(t\). Following (7), \(\hat{\varepsilon}_t\) may demonstrate a heavier tail than \(\varepsilon_t\). This is the main intuition why given that \(\varepsilon_t\) follows a normal distribution, it is still possible to obtain heavy-tailedness in the distribution of \(\hat{\varepsilon}_t\).

Next, we formally show that based on finite observations generated from a GARCH model with normal innovations, the estimated innovations follow a heavy-tailed distribution in Proposition 1. The proof is postponed to the Appendix.

**Proposition 1** Consider a GARCH(1,1) model in (1) and (2) with normally distributed innovations \(\{\varepsilon_t\}\). Suppose \(\hat{\lambda}_0\), \(\hat{\lambda}_1\) and \(\hat{\lambda}_2\) perfectly estimate \(\lambda_0\), \(\lambda_1\), \(\lambda_2\), i.e. \(\hat{\lambda}_0 = \lambda_0\), \(\hat{\lambda}_1 = \lambda_1\), \(\hat{\lambda}_2 = \lambda_2\). The estimated innovations from (7) and (8) nevertheless follow a heavy-tailed distribution for any finite \(t\).
2.3 Testing the Distribution of GARCH Innovations

As discussed in Section 2.1, different distributions of GARCH innovations lead to different levels of heavy-tailedness of the GARCH series, measured by its tail index. This provides an alternative way to test which model on innovations fits the data.

We investigate the difference between normal and Student-t innovations as an example. The GARCH series with innovations following these two distributions exhibit different tail behavior. For given GARCH coefficients, we define the implied tail index as follows.

**Definition 2** Implied tail index for the GARCH(1,1)-normal model

For a GARCH(1,1)-normal model, the parameter \( \kappa \) solved from equation (6), with a standard normally distributed innovation \( \varepsilon_t \), is denoted by \( \kappa_n(\lambda_1, \lambda_2) \). Then the implied tail index for the GARCH series is \( \alpha_n(\lambda_1, \lambda_2) = 2\kappa_n(\lambda_1, \lambda_2) \).

**Definition 3** Implied tail index for the GARCH(1,1)-Student model

For a GARCH(1,1)-Student model, the parameter \( \kappa \) solved from equation (6), with a standardized Student-t distributed innovation \( \varepsilon_t \) with degree of freedom \( \nu \), is denoted by \( \kappa_s(\lambda_1, \lambda_2, \nu) \). The implied tail index for 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 \( \lambda_1 + \lambda_2 = 1 \).

With normal (Student-t) distributed innovations, the GARCH coefficients \( \lambda_1, \lambda_2 \) (and the degree of freedom \( \nu \)) determine the tail index according to \( \alpha_n(\lambda_1, \lambda_2) \) (\( \alpha_s(\lambda_1, \lambda_2, \nu) \)). Hence, we obtain the implied tail index \( \alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) \) (\( \alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}) \)) after estimating the GARCH coefficients \( \hat{\lambda}_1, \hat{\lambda}_2 \) (and \( \hat{\nu} \)).

The parameters \( \lambda_1, \lambda_2 \) and \( \nu \) can be estimated 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 \( \lambda_1, \lambda_2 \) and \( \nu \), 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 indices of the GARCH series. However, if the distribution of innovations is misspecified, the following lemma shows that even if the GARCH coefficients are consistently estimated, the estimate of the implied tail index is inconsistent.

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

\[
P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \to 1.
\]

The relation \( P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \to 1 \) indicates that \( \alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) \) overestimate the real tail index \( \alpha \) 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 \( X_1, \ldots, X_T \) be the observations from a heavy-tailed distribution as in equation (4). The Hill estimator is defined as

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

where \( k \) is the number of observations that exceed the threshold \( s \), satisfying \( k \to 0 \), \( \frac{k}{T} \to 0 \) as \( T \to \infty \). The Hill estimator is usually applied to i.i.d. sample. Nevertheless, Resnick and Starića [1998] show that the consistency of Hill estimator for time series satisfying stochastic difference equations in the form of equation (3) still holds. For the asymptotic normality of the Hill estimator applying to weakly dependent series, it follows from Hsing [1991], Drees [2000] and Carrasco and Chen [2002]. According to the result in Carrasco and Chen [2002], the GARCH(1,1) process satisfies the \( \beta \)-mixing conditions which are required in Hsing [1991] to guarantee the asymptotic normality of the Hill estimator. More specifically, the Hill estimator converges to the tail index \( \alpha \) with a speed of convergence \( \sqrt{k} \). The asymptotic limit is given as

\[
\sqrt{k} \left( \frac{1}{\hat{\alpha}_H} - \frac{1}{\alpha} \right) \overset{d}{\to} N(0, v^2),
\]

where the asymptotic variance \( v^2 \) equals to \( \frac{1+\chi+\omega-2\psi}{\alpha^2} \). The parameters \( \chi \), \( \omega \) and \( \psi \) 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:

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

Notice that \( \hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu} \) converges to the parameters \( \lambda_1, \lambda_2, \nu \) with a speed of convergence \( \sqrt{T} \). The speed of convergence is faster than that of the Hill estimator because \( k/T \to 0 \) as \( T \to \infty \). When comparing
the implied tail index based on estimated model parameters to the estimated tail index, the limit behavior of the Hill estimator dominates. Thus, we can formally distinguish the two models by comparing the two tail indices. The following theorem gives the asymptotic properties of the test statistics. It follows directly from the asymptotic normality of the Hill estimator.

**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 \( \sqrt{T} \). \( \hat{\alpha}_H \) is the Hill estimator using \( k \) observations in the tail region, such that \( k \to \infty, \frac{k}{T} \to 0 \) as \( T \to \infty \) and \( \sqrt{k}(\frac{\hat{\alpha}}{\hat{\alpha}_H} - 1) \overset{d}{\to} N(0, 1 + \chi + \omega - 2\psi) \).

1. If \( X_1, \ldots, X_T \) follow a GARCH(1,1)-normal process, then
   \( \sqrt{k}(\frac{\hat{\alpha}}{\hat{\alpha}_H} - 1) \overset{d}{\to} N(0, 1 + \chi + \omega - 2\psi), \) as \( T \to +\infty \).

2. If \( X_1, \ldots, X_T \) follow a GARCH(1,1)-Student process, then
   \( \sqrt{k}(\frac{\hat{\alpha}}{\hat{\alpha}_H} - 1) \overset{d}{\to} N(0, 1 + \chi + \omega - 2\psi), \) as \( T \to +\infty \).

Notice that the MLE satisfies the required speed of convergence, \( \sqrt{T} \), see Bollerslev et al. [1986]. Hence we can use MLE for estimating the model parameters which lead to the calculation of the implied tail index.

### 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. This can potentially distort our diagnosis method. In this section, we evaluate whether our method is robust when taking into account the bias problem. In the theoretical setup, the bias is neglected by assuming \( \sqrt{k}(\frac{\alpha}{\hat{\alpha}_H} - 1) \overset{d}{\to} N(0, 1 + \chi + \omega - 2\psi) \) as \( T \to \infty \). This can only be achieved with a high level of \( s \), which leads to a low level of \( k \) and thus a low level of estimation accuracy. To obtain a better accuracy in estimating the tail index and a corresponding strong power in the testing procedure, it is necessary to consider a lower level of \( s \) for which the asymptotic bias is present. We evaluate to which extent the bias may contaminate our testing procedure.

The following lemma shows how the approximation on the tail region of a distributed function influences the asymptotic bias and variance of the Hill estimator (see Goldie and Smith [1987]).

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

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

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

\[ E[\hat{\alpha}_H - \alpha] = \frac{B\alpha\beta}{(\alpha + \beta)} s^{-\beta}[1 + o(1)], \tag{11} \]

and asymptotic variance is

\[ \text{Var}[\hat{\alpha}_H] = \frac{\alpha^2 s^\alpha}{AT}[1 + o(1)]. \tag{12} \]

Combining the asymptotic bias and variance, and omitting the \( o(1) \) term, we obtain the Asymptotic Mean Squared Error (AMSE) of the estimator \( \hat{\alpha}_H \) as follows,

\[ \text{AMSE}(\hat{\alpha}_H) = \frac{B^2 \alpha^2 \beta^2}{(\alpha + \beta)^2} s^{-2\beta} + \frac{\alpha^2 s^\alpha}{AT}. \tag{13} \]

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 \( \text{GARCH}(1,1) \)-normal model, suppose the tail approximation of \( \sigma_t^2 \) satisfies equation (10), where its tail index \( \kappa \) is the solution to the equation (6). Then, we must have \( \beta = 1 \) and

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

**Remark 8** For the \( \text{GARCH}(1,1) \)-Student model, the second-order index \( \beta = 1 \) and \( B \) has the same form as in (14) if \( E[(\lambda_1 \varphi + \lambda_2)^{\kappa+1}] < \infty \).
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 $\lambda_2$, while in the right panel, we fix $\lambda_2 = 0.88$ and plot the bias against $\lambda_1$.

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)-normal model by simulation against various values of $\lambda_1$ and $\lambda_2$. We first fix $\lambda_1$ at 0.08 and vary $\lambda_2$ from 0.87 to 0.91, then fix $\lambda_2$ at 0.88 and vary $\lambda_1$ from 0.07 to 0.11, $\lambda_0$ is fixed at 0.5. The simulation algorithm is given as follows.

1. For any given $\lambda_1$ and $\lambda_2$, calculate the implied tail index $\alpha = 2\kappa$, where $\kappa$ is solved from equation (6).

2. Generate 2500 samples with sample size 5000 from a GARCH(1,1)-normal model with given coefficients $\lambda_0$, $\lambda_1$ and $\lambda_2$. Then, for each sample $i$, calculate the Hill estimator $\hat{\alpha}_i(k)$ of the downside tail for various $k$, where $i = 1, \ldots, 2500$, $k = 1, \ldots, 500$.

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 value $k^*$ corresponding to the minimum MSE,

5. Plot the bias $\frac{1}{2500} \sum_{i=1}^{2500} (\hat{\alpha}_i(k^*) - \alpha)$ against $\lambda_1$ or $\lambda_2$.

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.
Table 1: Tail Index Estimates of GARCH(1,1)-normal Series

<table>
<thead>
<tr>
<th>$\lambda_2$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\hat{\alpha}_s$</th>
<th>$\hat{\alpha}_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>11.88</td>
<td>12.02</td>
<td>11.55</td>
<td>6.15</td>
</tr>
<tr>
<td>0.88</td>
<td>10.62</td>
<td>11.34</td>
<td>10.88</td>
<td>5.94</td>
</tr>
<tr>
<td>0.89</td>
<td>9.12</td>
<td>8.82</td>
<td>8.51</td>
<td>5.30</td>
</tr>
<tr>
<td>0.90</td>
<td>7.28</td>
<td>7.42</td>
<td>7.17</td>
<td>4.86</td>
</tr>
<tr>
<td>0.91</td>
<td>5.00</td>
<td>5.21</td>
<td>5.07</td>
<td>4.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\hat{\alpha}_s$</th>
<th>$\hat{\alpha}_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>13.98</td>
<td>14.44</td>
<td>13.80</td>
<td>6.47</td>
</tr>
<tr>
<td>0.88</td>
<td>10.62</td>
<td>11.63</td>
<td>11.16</td>
<td>5.86</td>
</tr>
<tr>
<td>0.89</td>
<td>9.00</td>
<td>7.90</td>
<td>7.63</td>
<td>5.05</td>
</tr>
<tr>
<td>0.90</td>
<td>7.28</td>
<td>5.62</td>
<td>5.88</td>
<td>5.71</td>
</tr>
<tr>
<td>0.91</td>
<td>5.00</td>
<td>3.70</td>
<td>3.96</td>
<td>3.87</td>
</tr>
</tbody>
</table>

Note: This table presents the tail index estimation results on 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 $\lambda_2$ from 0.87 to 0.91, while in the right part, we fix $\lambda_2 = 0.88$ and vary $\lambda_2$ 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 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. The numbers reported are the average levels across 100 samples.

by the Hill estimation is particularly effective for the case that $\lambda_1 + \lambda_2$ is close to 1. In other words, our method works better when the GARCH process is Near-IGARCH. Compared to the backing-out method, our method is particularly robust for the Near-IGARCH case, hence is preferred for applications to financial time series.

4 Simulation

We use simulation to show the performance of our new diagnosis method on the distribution of GARCH innovations. We consider two distributions of the innovations, the normal and the Student-t distributions. 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 number of observations $k$ in the tail region.\(^3\)

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.

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 $\lambda_2$ from 0.87 to 0.91, then fix $\lambda_2 = 0.88$ with varying $\lambda_1$ from 0.07 to 0.11.

\(^3\)The 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 2: Tail Index Estimates of GARCH(1,1)-Student Series

<table>
<thead>
<tr>
<th>$\lambda_2$</th>
<th>$\alpha$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\hat{\alpha}_s$</th>
<th>$\hat{\alpha}_H$</th>
<th>$\lambda_1$</th>
<th>$\alpha$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\hat{\alpha}_s$</th>
<th>$\hat{\alpha}_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5.94</td>
<td>12.54</td>
<td>6.00</td>
<td>3.94</td>
<td>0.07</td>
<td>6.48</td>
<td>15.25</td>
<td>6.68</td>
<td>4.03</td>
</tr>
<tr>
<td>0.88</td>
<td>5.56</td>
<td>11.50</td>
<td>5.81</td>
<td>3.90</td>
<td>0.08</td>
<td>5.56</td>
<td>11.28</td>
<td>5.74</td>
<td>3.78</td>
</tr>
<tr>
<td>0.89</td>
<td>5.08</td>
<td>9.54</td>
<td>5.10</td>
<td>3.55</td>
<td>0.09</td>
<td>4.66</td>
<td>8.21</td>
<td>4.79</td>
<td>3.64</td>
</tr>
<tr>
<td>0.90</td>
<td>4.42</td>
<td>8.29</td>
<td>4.78</td>
<td>3.51</td>
<td>0.10</td>
<td>3.78</td>
<td>6.06</td>
<td>3.95</td>
<td>3.26</td>
</tr>
<tr>
<td>0.91</td>
<td>3.46</td>
<td>5.42</td>
<td>3.46</td>
<td>3.07</td>
<td>0.11</td>
<td>2.88</td>
<td>3.93</td>
<td>2.91</td>
<td>2.95</td>
</tr>
</tbody>
</table>

Note: This table presents the tail index estimation results on 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 $\lambda_2$ from 0.87 to 0.91, while in the right part, we fix $\lambda_2 = 0.88$ and vary $\lambda_2$ 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 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. The numbers reported are the average levels across 100 samples.

0.11. The constant $\lambda_0$ is fixed at 0.5. From the results, we observe that both $\hat{\alpha}_n$ and $\hat{\alpha}_s$ robustly estimate the tail index $\alpha$. However, the Hill estimate $\hat{\alpha}_H$ underestimates the tail index $\alpha$, i.e. overestimates the heavy-tailedness. The underestimation is severe when $\lambda_1 + \lambda_2$ is relatively low. This can be explained by the downward bias of the Hill estimator. 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 Near-IGARCH.

The simulation results of GARCH(1,1)-Student model are presented in Table 2. We use the same parameters $\lambda_0$, $\lambda_1$ and $\lambda_2$ 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 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 $\alpha$ when $\lambda_1 + \lambda_2$ is relatively low. Similar to the GARCH(1,1)-normal model, the Hill estimator $\hat{\alpha}_H$ performs well when the GARCH series is Near-IGARCH.

As a conclusion, comparing the implied tail index with the Hill estimate yields 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 Near-IGARCH.

5 An Empirical Example

In this section, we apply our method to diagnose the distribution of the GARCH innovations when fitting the GARCH(1,1) model to financial time series. The dataset consists of S&P 500 Composite Index and 12
Table 3: Tail Index Estimates of Real Data

<table>
<thead>
<tr>
<th></th>
<th>GARCH-normal</th>
<th>GARCH-Student</th>
<th>Tail Index</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\lambda}_1$</td>
<td>$\hat{\lambda}_2$</td>
<td>$\hat{\nu}$</td>
<td>$\hat{\alpha}_n$</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>0.0770</td>
<td>0.9174</td>
<td>0.0766</td>
<td>0.9225</td>
</tr>
<tr>
<td>Auto</td>
<td>0.0616</td>
<td>0.9290</td>
<td>0.0655</td>
<td>0.9284</td>
</tr>
<tr>
<td>BioTech</td>
<td>0.0608</td>
<td>0.9346</td>
<td>0.0536</td>
<td>0.9455</td>
</tr>
<tr>
<td>Media</td>
<td>0.0693</td>
<td>0.9257</td>
<td>0.0675</td>
<td>0.9285</td>
</tr>
<tr>
<td>Chemical</td>
<td>0.0818</td>
<td>0.9182</td>
<td>0.0747</td>
<td>0.9238</td>
</tr>
<tr>
<td>Pharm</td>
<td>0.0755</td>
<td>0.9152</td>
<td>0.0798</td>
<td>0.9152</td>
</tr>
<tr>
<td>Retail</td>
<td>0.0649</td>
<td>0.9298</td>
<td>0.0594</td>
<td>0.9373</td>
</tr>
<tr>
<td>SoftWare</td>
<td>0.0596</td>
<td>0.9356</td>
<td>0.0562</td>
<td>0.9438</td>
</tr>
<tr>
<td>Transport</td>
<td>0.0587</td>
<td>0.9365</td>
<td>0.0664</td>
<td>0.9260</td>
</tr>
<tr>
<td>Aero</td>
<td>0.0739</td>
<td>0.9197</td>
<td>0.0727</td>
<td>0.9198</td>
</tr>
<tr>
<td>Steel</td>
<td>0.0525</td>
<td>0.9418</td>
<td>0.0541</td>
<td>0.9433</td>
</tr>
<tr>
<td>Insurance</td>
<td>0.0925</td>
<td>0.9041</td>
<td>0.0891</td>
<td>0.9109</td>
</tr>
<tr>
<td>Bank</td>
<td>0.0870</td>
<td>0.9130</td>
<td>0.0863</td>
<td>0.9137</td>
</tr>
</tbody>
</table>

**Note:** This table presents the diagnosis results for S&P 500 index and 12 equity indices from US market. The dataset consists of daily returns from 01.01.1995 to 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. $\hat{p}_n$, $\hat{p}_s$ are the corresponding $p$-values based on the test statistics in Theorem 5 when testing whether the GARCH innovations follow a normal or Student-t distribution.

S&P equity sector indices. The time series of daily data runs from 1 January 1995 to 31 December 2010, with a sample 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 tests in Theorem 5, we calculate the corresponding $p$-values of the hypothesis tests in which either the normal innovation or the Student-t innovation is regarded as the null. 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 fitted GARCH models, no matter which distribution of innovations is used, are Near-IGARCH. This is exactly the case in which the backing-out 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 the two on 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.

Although we draw the same conclusion as previous studies that the GARCH innovations of the equity indices are more likely to follow the Student-t distribution rather than the normal distribution, the robustness
Note: The plots show the estimated daily VaR of S&P Auto Index in 2010 from the GARCH(1,1)-Student model against those from the GARCH(1,1)-normal model and the backing-out approach. The upper panel is for the VaR at the confidence level 99% while the lower panel is for the VaR at the confidence level 99.9%.

of our new method for the Near-IGARCH case provides more confidence on such a statement.

To demonstrate the economic significance on the difference of the two models, we take S&P Auto Index as an example to estimate its dynamic VaR by three different approaches: GARCH(1,1)-normal model, GARCH(1,1)-Student model and the backing-out approach. The backing-out approach estimates the GARCH(1,1) model by QMLE. We first back-out the innovation and model it as a Student-t distribution, then estimate the degree of freedom by MLE.\(^4\) For each approach, given the estimated GARCH coefficients, we can estimate the daily volatility as in (8), with taking the sample mean and variance as the initial values of \(X_0\) and \(\sigma_0^2\). Lastly, the VaR at the confidence level \(1 - p\) in day \(t\) can be calculated by

\[
\hat{\text{VaR}}_t^p = -\hat{\sigma}_t G^{-1}(p),
\]

where \(G^{-1}(p)\) is the \(p\)-quantile of the distribution of innovations.

Figure 3 plots the daily VaR of S&P Auto Index in 2010 estimated from the GARCH(1,1)-Student model against those from the other two approaches. The upper panel is for the VaR estimates at the confidence level 99% while the lower panel is for the VaR estimates at the confidence level 99.9%. We observe the VaR estimated from the GARCH-Student model are higher than those from the GARCH-normal model, but lower

\(^4\)In this example, \(\hat{\nu} = 4.1586\).
than those from the backing-out approach. This is consistent with our earlier discussion that when choosing
the normal innovation, it may lead to an underestimate of the potential tail risk, while using the backing-out
approach may overestimate the potential tail risk. The phenomenon is more severe when considering more
extreme tail events.

6 Conclusion and Discussion

This paper investigates diagnosis methods on the distribution of GARCH innovations. We first show that
the backing-out method based on estimated innovations is not robust. This is particularly the case when the
GARCH process is Near-IGARCH. We demonstrate, both by simulation and by a theoretical argument, that
the estimated innovations from a finite sample generated by a GARCH model with normal innovations may
follow a heavy-tailed distribution.

We provide an alternative approach on diagnosing the distribution of GARCH innovations by comparing
implied tail indices with the estimates 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 Near-IGARCH.
The simulation results support our proposed approach.

We apply our method to the returns of 13 stock indices. The estimated GARCH coefficients for all
indices in both models indicates the phenomenon of Near-IGARCH. Hence, the backing-out method is not
robust for these data, while our method can help diagnose the distribution of innovations. By applying our
method, we reconfirm that a GARCH model with Student-t innovation fits the data better than a GARCH
model with normal innovation.

One potential reason why the Student-t distribution fits better is that it exhibits heavy-tails while the
normal distribution does not. Acknowledging heavy-tails in the innovations leads to important implication
in dynamic risk management: when applying a GARCH model to evaluate conditional risk measures given
information on return level and volatility in the previous periods, 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 of the innovations based on their estimates. That may
overestimate the conditional risk measures. Lastly, to accurately evaluate the unconditional tail risk of a
GARCH process, it is better to consider the implied tail index instead of using the Hill estimator because of
the potential bias.

References


7 Appendix

7.1 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 \( \varepsilon_t \) follows a normal distribution and is independent from \( \hat{\sigma}_t \), it is only necessary to prove that \( \frac{\sigma_t}{\hat{\sigma}_t} \) follows a heavy-tailed distribution.

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

\[
A_t = \sum_{j=1}^{t} \lambda_0 \prod_{i=2}^{j} (\lambda_1 \varepsilon_{t-i+1}^2 + \lambda_2),
\]

\[
B_t = \prod_{i=1}^{t} (\lambda_1 \varepsilon_{t-i}^2 + \lambda_2).
\]

Hence, we have that

\[
P \left( \frac{\sigma_t^2}{\hat{\sigma}_t^2} > x \right) = P \left( A_t + B_t \sigma_0^2 > x (\lambda_2^2 (\hat{\sigma}_0^2 - \sigma_0^2) + A_t + B_t \sigma_0^2) \right)
\]

\[
= P \left( [B_t - x(B_t - \lambda_2^2)] \sigma_0^2 > (x - 1)A_t + x\lambda_2^2 \sigma_0^2 \right).
\]

We further derive a lower bound of this probability by conditional on \( B_t < \frac{x+1}{x} \lambda_2^2 \), which guarantees that \( B_t - x(B_t - \lambda_2^2) > \frac{\lambda_2^2}{x} \). We get that

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

\[
> P \left( \frac{\lambda_2^2}{x} \sigma_0^2 > (x - 1)A_t + x\lambda_2^2 \sigma_0^2, B_t < \frac{x + 1}{x} \lambda_2^2 \right)
\]

\[
> P \left( \frac{\lambda_2^2}{x} \sigma_0^2 > x(A_t + \lambda_2^2 \sigma_0^2), B_t < \frac{x + 1}{x} \lambda_2^2 \right)
\]

\[
= P \left( \sigma_0^2 > x^2 (\frac{A_t}{\lambda_2^2} + \sigma_0^2), B_t < \frac{x + 1}{x} \lambda_2^2 \right).
\]

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

$$P\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2} > x\right) = P\left(\sigma_0^2 > x^2 \left( \lambda_0 - \frac{\lambda_t - 1}{\lambda_t^2 - 1} - \frac{\lambda_0}{\lambda_0^2 - 1} B_t + \hat{\sigma}_0^2 \right), B_t < \frac{x+1}{x} \lambda_2^t \right)$$

$$= E_{B_t} 1_{B_t < \frac{x+1}{x} \lambda_2^t} P\left(\sigma_0^2 > x^2 \left( \lambda_0 - \frac{\lambda_t - 1}{\lambda_t^2 - 1} - \frac{\lambda_0}{\lambda_0^2 - 1} B_t + \hat{\sigma}_0^2 \right) | B_t \right)$$

$$\sim E_{B_t} 1_{B_t < \frac{x+1}{x} \lambda_2^t} \frac{C}{x^{2\kappa}} \left( \frac{\lambda_0}{\lambda_0^2 - 1} \frac{x+1}{x} \lambda_2^t + \hat{\sigma}_0^2 \right)^{-\kappa}, \text{ as } x \to +\infty.$$

The last step comes from the facts that $\sigma_0^2$ is independent from $B_t$ and follows a heavy-tailed distribution with tail index $\kappa$. Here $C$ is the tail scale of $\sigma_0^2$. Under the condition $B_t < \frac{x+1}{x} \lambda_2^t$, we have that

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

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 $\chi^2$ distributed random variable to simplify the calculation as follows:

$$P\left(B_t < \frac{x+1}{x} \lambda_2^t \right) = P\left(\log B_t - t \log \lambda_2 < \log \frac{x+1}{x} \right)$$

$$= P\left(\sum_{i=1}^{t} \lambda_1 \frac{e_i^2}{\lambda_2} < \log \frac{x+1}{x} \right)$$

$$> P\left(\sum_{i=1}^{t} \lambda_1 \frac{e_i^2}{\lambda_2} < \log \frac{x+1}{x} \right)$$

$$= \frac{2^{t/2} \left( \frac{\lambda_1}{2}\log \frac{x+1}{x} \right)^{t/2}}{\Gamma(t/2)}$$

$$= C_2 \left( \log \left( 1 + \frac{1}{x} \right) \right)^{t/2} \sim C_2 x^{-t/2}, \text{ as } x \to +\infty.$$
Therefore, we have that
\[
\liminf_{x \to +\infty} P \left( \frac{\sigma^2_t}{\sigma_t^2} > x \right) x^{(2\kappa + 1/2)} \geq C_1 C_2.
\]
Hence, \( \frac{\sigma^2_t}{\sigma_t^2} \) follows a heavy-tailed distribution, which implies that \( \frac{\sigma_t^2}{\sigma^2_t} \) follows a heavy-tailed distribution. This completes the proof of the proposition.

7.2 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 \( \nu \) finite. We start with stating the following lemma.5

Lemma 9 Let \( \psi = Z_\nu^2 \) and \( \phi = Z_\mu^2 \), where \( Z_\nu \) and \( Z_\mu \) follow standardized Student-t distribution with degree of freedoms \( \nu \) and \( \mu \) respectively. Then \( E(\psi) = E(\phi) = 1 \). Denote the cumulative distribution functions of \( \psi \) and \( \phi \) as \( F_\psi \) and \( F_\phi \), then \( D(x) = \int_0^t F(\psi \leq x)dx - \int_0^t F(\phi \leq x)dx \geq 0 \), \( \forall t > 0 \) and \( 2 < \nu < \mu < +\infty \).

Lemma 9 shows that \( \phi \) second-order stochastically dominates over \( \psi \). By taking \( \mu \to +\infty \), it leads to the following Corollary.

Corollary 10 Let \( \varphi \) be a random variable following the \( \chi^2(1) \) distribution. Then \( \varphi \) second-order stochastically dominates over \( \psi \) for any finite \( \nu \).

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 \( E[(\lambda_1 \varphi + \lambda_2)^{\kappa_1}] < E[(\lambda_1 \psi + \lambda_2)^{\kappa_1}] = 1 \). 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 \( \nu \) finite. From the consistency that \( \alpha_s(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\nu}) \xrightarrow{P} \alpha(\lambda_1, \lambda_2, \nu) \) as \( T \to +\infty \). It implies that \( P(\alpha_n(\hat{\lambda}_1, \hat{\lambda}_2) > \alpha) \to 1 \) as \( T \to +\infty \).

7.3 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 \( k \to \infty, \frac{k}{T} \to 0 \) as \( T \to +\infty \), we get that
\[
\sqrt{k}(\hat{\lambda}_1 - \lambda) \xrightarrow{P} 0 \quad \text{and} \quad \sqrt{k}(\hat{\lambda}_2 - \lambda) \xrightarrow{P} 0 \quad \text{as} \quad T \to +\infty.
\]

---

5The proof is available upon request.
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 \( \sqrt{k} \left( \kappa_n(\hat{\lambda}_1, \hat{\lambda}_2) - \kappa(\lambda_1, \lambda_2) \right) \xrightarrow{P} 0 \) as \( T \to +\infty \) 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 \( \lambda_1 \), 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
\[ \left| \frac{\partial \kappa}{\partial \lambda_1} \right| = \frac{E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1} \kappa \varphi]}{E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)]} = \frac{\kappa E[(\lambda_1 \varphi + \lambda_2)^\kappa] - \lambda_2 E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1}]}{E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)]}. \]

The conditions \( \lambda_1 > 0, \lambda_2 > 0 \) and \( \lambda_1 + \lambda_2 < 1 \) implies that \( \kappa > 1 \). From the proof of Lemma 7, we get 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
\[ \left| \frac{\partial \kappa}{\partial \lambda_1} \right| < \frac{\kappa(1+\lambda_2)}{\lambda_1 \lambda_2 D_0}. \]

Following a similar procedure, we take the partial derivative of (6) with respect to \( \lambda_2 \) to obtain that
\[ E[(\lambda_1 \varphi + \lambda_2)^\kappa] \left( \frac{\kappa}{\lambda_1 \varphi + \lambda_2} \varphi + \log(\lambda_1 \varphi + \lambda_2) \frac{\partial \kappa}{\partial \lambda_2} \right) = 0, \]
which implies that
\[ \left| \frac{\partial \kappa}{\partial \lambda_2} \right| = \frac{\kappa E[(\lambda_1 \varphi + \lambda_2)^{\kappa-1}]}{E[(\lambda_1 \varphi + \lambda_2)^\kappa \log(\lambda_1 \varphi + \lambda_2)]} < \frac{\kappa}{D_0}. \]
With the bounded partial derivatives, we conclude that

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

The asymptotic normality of the Hill Estimator follows from Hsing [1991]. The Hill estimator converges to the tail index $\alpha$ with speed of convergence $\sqrt{k}$. The asymptotic limit is given as

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

Therefore,

$$\sqrt{k} \left( \frac{\alpha_n}{\hat{\alpha}_H} - 1 \right) = \sqrt{k} \left[ \left( \alpha_n - \alpha \right) \left( \frac{\alpha}{\hat{\alpha}_H} - 1 \right) \right] = \sqrt{k} \left( \frac{\alpha_n - \alpha}{\hat{\alpha}_H} \right) + \sqrt{k} \left( \frac{\alpha}{\hat{\alpha}_H} - 1 \right).$$

With (15) and (16), we get that as $T \to \infty$,

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

**Proof of Statement 2:** the proof follows similar lines as that in the proof of Statement 1. We only need to check the boundedness of $\left| \frac{\partial f}{\partial \nu} \right|$. 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 $\sqrt{X / (\nu - 2)}$ follows a standardized $t(\nu)$ and $\sqrt{X / (\nu_0 - 2)}$ follows a standardized $t(\nu_0)$ distribution, both with variance 1. Recall equation (6), we have that,

$$E[\left( \lambda_1 \frac{X}{Y/(\nu - 2)} + \lambda_2 \right)^{\kappa(\lambda_1, \lambda_2, \nu)}] = 1,$$

$$E[\left( \lambda_1 \frac{X}{Y_0/(\nu_0 - 2)} + \lambda_2 \right)^{\kappa_0(\lambda_1, \lambda_2, \nu_0)}] = 1.$$
Hence,

\[
0 = E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] - E \left[ \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] \\
= \{E \left[ \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] - E \left[ \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] \} \\
+ \{E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] - E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] \} \\
+ \{E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] - E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] \} \\
=: I_1 + I_2 + I_3.
\]

We deal with the three parts separately. First, for \( I_1 \), by applying the Mean Value Theorem (MVT), there exists \( \tilde{\kappa} \) between \( \kappa_0 \) and \( \kappa \), such that

\[
I_1 = E \left[ \left( \frac{X}{Y_0} \right)^\kappa \log \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] (\kappa_{(\lambda_1, \lambda_2, \nu)} - \kappa_{(\lambda_1, \lambda_2, \nu)}) \\
=: J_1 (\kappa_{(\lambda_1, \lambda_2, \nu)} - \kappa_{(\lambda_1, \lambda_2, \nu)}) .
\]

Secondly, for \( I_2 \), by applying the MVT, there exists \( \xi \) between \( Y_0 \) and \( Y \), such that

\[
-I_2 = \kappa E \left[ \left( \frac{X}{\xi} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \left( \frac{X}{\xi^2} \right) \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] (Y - Y_0) \\
\leq \kappa E \left[ \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \left( \frac{X}{Y_0} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] E(Y - Y_0) \\
=: \kappa J_2 E(Y - Y_0).
\]

Thirdly, for \( I_3 \), by applying the MVT, there exists \( \tilde{\nu} \) between \( \nu_0 \) and \( \nu \) such that

\[
I_3 = \kappa E \left[ \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \left( \frac{X}{\tilde{\nu}} \right) \left( \frac{X}{Y} \right)^\kappa_{(\lambda_1, \lambda_2, \nu)} \right] (\nu - \nu_0) \\
=: J_3 (\nu - \nu_0).
\]

We can bound the terms \( J_1 \) and \( J_3 \) from below by positive numbers \( D_1 \) and \( D_3 \) and we can also bound...
$J_2$ by $J_2 < D_2$ for some finite positive number $D_2$. Combining 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.

### 7.4 Proof of Lemma 7

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

$$P\{\sigma_t^2 > x\} = P\{\lambda_0 + (\lambda_1 \varphi_{t-1} + \lambda_2) \sigma_{t-1}^2 > x\}$$

$$= E\left[ P\{\sigma_t^2 > \frac{x - \lambda_0}{\lambda_1 \varphi_{t-1} + \lambda_2} | \varphi_{t-1}\} \right]$$

$$= \int_0^\infty \tilde{F}\left( \frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right) f(\varphi) d\varphi.$$

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

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

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

$$+ o \left( \frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2} \right)^{-\beta}. $$

With the Tailor expansion that

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

---

\[6\] The proof on the existence of such lower and upper bounds are available upon request.
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

\[
\int_0^\infty \bar{F}\left(\frac{x - \lambda_0}{\lambda_1 \varphi + \lambda_2}\right) f(\varphi) d\varphi = A x^{-\kappa} \left\{ E[(\lambda_1 \varphi + \lambda_2)\kappa] + B E[(\lambda_1 \varphi + \lambda_2)^{\kappa+\beta}] x^{-\beta} + \kappa \lambda_0 E[(\lambda_1 \varphi + \lambda_2)\kappa+1] x^{-1} + \max\{o(x^{-\beta-1}), o(x^{-2})\}\right\}.
\]

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 \(\beta\) must be 1. Furthermore, by considering the second-order tail scale on both sides, we get that \(B\) 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 \(g''(s) = E[(\lambda_1 \varphi + \lambda_2)^s \log^2(\lambda_1 \varphi + \lambda_2)] > 0\), thus \(g(s)\) is a convex function, with \(g(0) = g(\kappa) = 1\). From Cauchy Mean Value Theory, there exists \(\bar{s} \in [0, \kappa]\), such that \(g'(\bar{s}) = 0\). Because \(g'(s)\) is increasing, we get that \(g'(s) > 0\) for \(s > \bar{s}\). Therefore, \(g(s)\) is an increasing function for \(s \geq \bar{s}\), which implies that \(g(\kappa + 1) > 1\). This leads to \(B < 0\).