Indirect estimation of GARCH models with alpha-stable innovations

Parrini, Alessandro

Universita' di Firenze

18 April 2012

Università degli Studi di Firenze
Dipartimento di Statistica "G. Parenti"

Master's Degree in Statistical Sciences
Thesis in Financial Econometrics

Indirect Estimation of GARCH Models with α-stable Innovations

Alessandro Parrini

Supervisor: Prof. Giorgio Calzolari
Co-supervisor: Dr. Roxana Halbleib

Academic Year 2010/2011

Contents

Acknowledgments

I am grateful to Prof. Giorgio Calzolari for introducing me to this idea and to always point me in the right direction.
Special thanks go to Dr. Roxana Halbleib for her valuable encouragments and advices, and for the practical assistance she gave me during my thesis research at the University of Konstanz.
Dr. Marco J. Lombardi, at the European Central Bank, gave some precious suggestions in the early stage of this work.
I also benefited from the presentation of a preliminary version of this work at the internal seminar of the Department of Economics, Universität Konstanz.

Ringrazio i miei compagni di viaggio, in particolare Gabriele e Marco, senza i quali, quasi certamente, oggi non sarei qua.

Ringrazio la mia famiglia tutta, per essermi sempre stata vicino e per avermi supportato (e sopportato) in questi anni. Un pensiero particolare va a mia mamma e a Monika, per tutto l'amore (ricambiato, a dispetto della apparenza) che mi hanno saputo dare.

Abbraccio i miei amici di sempre, pochi ma boni. Ed è inutile farne i nomi, perchè sono coloro che sanno di poter contare su di me, come io ho sempre contato su di loro.

Zum Schluss, möchte ich mich bei Allen bedanken, die ich in Konstanz getroffen habe, und die dazu beigetragen haben, dass mein Aufenthalt zu einem wunderbar Erlebnis wurde.

Preface

Modeling financial time series is a complex problem. This complexity is not only due to the variety of the series in use (stocks, exchange rates, interest rates, etc.), or to the importance of the frequency of the observations (second, minute, hour, day, etc.). It is mainly due to the existence of statistical regularities (stylized facts) which are common to a large number of financial time series, but are difficult to reproduce artificially using stochastic models. Most of these stylized facts were put forward in a seminal paper by Mandelbrot (1963)[29]; since then, they have been documented, and completed, by many empirical studies.

Let P_t denote the price of an asset at time t and let $r_t = \ln(P_t/P_{t-1})$ be the corresponding (log-)return. It is well-established, in the financial literature, that returns series generally display small linear autocorrelations, making it close to a *white noise* and reducing the possibility to correctly forecast their expected value (and thus, the profit opportunities). On the other hand, squared returns (or absolute returns) are generally strongly autocorrelated and they tend to appear in clusters, meaning that highly risky time subperiods are typically followed by highly risky ones and vice versa. These phenomena suggest that forecasting squared returns, i.e. - assuming absence of autocorrelation - forecasting the variance of returns (*volatility*) is easier than forecasting returns themselves.

The most extensively studied models of time-varying (conditional) volatility are the *autoregressive conditional heteroskedastic* (ARCH) models, introduced by Engle (1982)[11]. Since the introduction of the *generalized* ARCH (GARCH) models (Bollerslev, 1986 [5]), such a framework has become extremely popular among both academics and practitioners, being simple enough to be used in practice, but also rich in theoretical problems (some of them unsolved). In fact, GARCH models led to a fundamental change to the approaches used in finance, through an efficient modeling and forecasting of the volatility of financial asset returns[1].

The GARCH class of models assumes that returns are mean zero, serially uncorrelated but not serially independent, with nonconstant variances conditionally on the past. Namely, $r_t \equiv \epsilon_t = z_t \sqrt{\sigma_t^2}$, where the innovations z_t are standard normal distributed and the conditional variances σ_t^2 depend on past information, $\sigma_t^2 = \sigma_t^2(\mathcal{I}_{t-1})$. An important assumption of GARCH-type models is that the conditional distribution of the process has finite second moment, imposing limits on the heaviness of the tails of its unconditional distribution. Given that a wide range of financial data exhibits remarkably fat tails, this assumption represents a major shortcoming of GARCH models in handling financial time series. This observation has led to the use of the Student's t in place of the normal as conditional distribution (for instance, Fiorentini et al., 2003 [13]). The Student-t distribution allows for heavier-than-normal tails, but, in contrast with the normal, lacks the desirable *stability* property. Stability is desirable because stable distributions have domains of attraction, and thus provide very good approximations for large class of distributions. It is therefore difficult to find theoretical reasons for which the innovations should be Student-t distributed. Moreover, some empirical studies (e.g. Yang & Brorsen, 1993 [44]) indicate that the tail behavior of GARCH models remains too short even with Student-t innovations.

For what claimed so far, the family of α-stable distributions, which includes the normal as a special case, seems a natural candidate for the conditional distribution of GARCH-type models[2]. The use of models based on α-stable distributions is encouraged by the *generalized central limit theorem* (Gnedenko and Kolmogorov, 1954), according to which α-stable distributions

[1] In 2003, the Nobel Prize for Economics was jointly awarded to Robert F. Engle and Clive W.J. Granger "for methods of analyzing economic time series with time-varying volatility (ARCH)".

[2] GARCH models with (symmetric) stable innovations have been first proposed by McCulloch (1985)[30].

are the limiting law of standardized sums of independent random variables in a wider range of cases than the normal[3]. Mandelbrot (1963) first proposed to use the other members of the stable class rather than the normal for fitting data in which extreme values are frequent. In fact, this class of distributions has four parameters, two of which deal with, respectively, tail-thickness and asymmetry. Hence, it manages to accommodate heavy-tailed financial series - producing more reliable measure of risk - and, in addition, it is able to capture skewness in distribution, another characteristic feature of financial time series unaccounted for by GARCH models.
In the light of these considerations, it may sound strange that α-stable distributions have not enjoyed a better fortune in applied fields. This is probably due to the associated estimation difficulties, since, in most cases, the density function of α-stable laws cannot be expressed in closed form. Such estimation difficulties have somehow dampened also the academic interest in α-stable models. A notable exception was an interesting analysis of the relation between GARCH models and α-stable distributions proposed by de Vries (1991)[10] and Ghose & Kroner(1993) [19], which also highlight the main source of difficulties that will be encountered in this work.

To estimate the parameters of the proposed GARCH models with α-stable innovations, we use the indirect inference methods introduced by Gouriéroux et al. (1993)[21] and Gallant & Tauchen (1996)[16]. These methods can be applied in situations in which the likelihood function cannot be expressed in closed form or it is difficult to compute, while it is simple to produce simulated observations from the model of interest. Since pseudo-random numbers from α-stable distributions can be readily generated - by means of the algorithm by Chambers et al. (1976) [9] - the indirect inference approach could prove useful to overcome the estimation difficulties arising from stable models.
Indirect inference involves the use of an auxiliary model (easier to handle than the model of interest), whose parameters are recovered through the maximization of its pseudo-likelihood function. In this work, we use as auxiliary model

[3] Asset returns are commonly thought of as the result of the aggregation of the asset allocation decisions of market partecipants. Therefore, the resulting distributions should arise, in the limit, from a central limit theorem.

a GARCH with skewed Student's t innovations. The skew-t distribution[4] appears as a good candidate for our purpose since it has four parameters, with a sort of one-to-one correspondence with those of the α-stable distribution.

This work is structured as follows. In the first chapter, the traditional ARCH and GARCH models are presented, with a particular focus on the weaknesses of (Gaussian) GARCH processes. The second chapter introduces to α-stable family of distributions and its main properties; the algorithm to simulate α-stable pseudo-random numbers is then reported. Chapter 3 presents the indirect inference methods and their asymptotic properties, whose proofs are sketched in Appendix A. The first three chapters are autonomously readable; their content is then linked and exploited in chapter 4. Here, after briefly introducing the skew-t distribution, we describe in detail the model we propose to estimate and the chosen auxiliary model. Then, we address the estimation difficulties encountered and how they have been overcome. Simulation studies - conducted using the GAUSS matrix programming language - are displayed and commented; finally, the proposed models are used to estimate the IBM weekly returns series, as an illustration of how they perform on real data. A part of the GAUSS code used for the simulation studies can be found in Appendix B, to see how stable GARCH models have been simulated.

[4]Some alternative versions are available in the literature; we use the one by Azzalini & Capitanio (2003)[2].

Chapter 1

GARCH Processes

Volatility is by far the most used measure of risk in financial applications. Consequently, volatility measuring, modeling and forecasting is of central importance in financial econometrics. Typically computed as the standard deviation of the underlying asset return, the concept of volatility has a long history in finance and has become a key ingredient in many theoretical frameworks such as risk management, portfolio management, CAPM, option pricing etc. The practical implementation of these frameworks requires estimate of the unobserved volatility. In fact, a particular feature of stock volatility is that it is not directly observable and this makes it difficult also to evaluate the forecasting performance of volatility models.

Initially, the volatility was considered to be constant in time and was computed as the sample standard deviation over a given time window (*historical volatility*). Its accuracy clearly depends on the historical window size. Since it gives the same weight to all past and recent observations, the historical volatility might not be a relevant measure of today's risk if computed on very large windows. Additionally, it becomes very noisy when computed on a short window. However, it is widely accepted that the volatility is time varying, i.e. it evolves over time in a continuous manner. Moreover, there are some other features commonly seen in asset returns that play an important role in the development of dynamic volatility models. First, as noted by Mandelbrot in his study on speculative prices (1963) [29], there exists *volatility clustering*, that is "large changes tend to be followed by large changes, of either sign, and small

changes tend to be followed by small changes". Second, volatility does not diverge to infinity, or statistically stated, volatility is often stationary. Third, volatility seems to react differently to a big price increase or a big price drop, referred to as the *leverage effect*. These empirical properties make it easier to forecast conditional variance of asset returns than returns themselves.
The most extensively studied models of time-varying conditional volatility are the autoregressive conditional heteroskedastic (ARCH) models, first introduced by Engle (1982)[11] and then modified by Bollerslev (1986)[5], who proposed the *generalized* autoregressive conditional heteroskedastic (GARCH) models.

1.1 General setting

Let us observe that any time series y_t can be decomposed into a predictable part and an unpredictable part:

$$y_t = \mathbb{E}(y_t|\mathcal{I}_{t-1}) + \epsilon_t \tag{1.1}$$

where $\mathcal{I}_{t-1}$ denotes the information set consisting of all the relevant information up to and including time $t-1$. The process $\{\epsilon_t\}$ is a *weak white noise*: $\mathbb{E}(\epsilon_t|\mathcal{I}_{t-1}) = 0$, $\text{cov}(\epsilon_t, \epsilon_{t+h}) = 0 \;\forall\, h \neq 0$. For instance, a reasonable model for the log-price process is of the form $\ln P_t = c + \ln P_{t-1} + \epsilon_t$, so that the corresponding log-return process is given by

$$r_t := \ln P_t - \ln P_{t-1} = c + \epsilon_t$$

with $\mathbb{E}(r_t|\mathcal{I}_{t-1}) = c$ (the notation ":=" is used to mean "defined as"). In this case the predictable part of the return process is simply a constant (the so called *drift* parameter).
In the following we are interested in focusing on the conditional variance of the unpredictable part. For the conditional mean any meaningful specification (e.g. an ARMA(p, q) model) can be introduced. However, it is worthy to recall that serial dependence of a stock return series is generally weak, if it exists at all.

1.2 ARCH Models

In an ARMA framework the error term ϵ_t is assumed to be both conditionally and unconditionally homoskedastic, i.e.

$$\mathrm{V}(\epsilon_t) = \mathrm{V}(\epsilon_t|\mathcal{I}_{t-1}) \equiv \sigma_\epsilon^2 \quad \forall\, t.$$

To generalize this implausible setting, the ARCH-type models relax the conditionally homoskedastic assumption and assume the following decomposition for the unpredictable part:

$$\epsilon_t = z_t\sqrt{\sigma_t^2} \tag{1.2}$$

where z_t, conditional upon the information available at $t-1$, is an independent identically distributed (iid) error term with zero expectation and unit variance. The assumption of unit variance can be introduced without any loss of generality. The function σ_t^2 is assumed to depend on past information: $\sigma_t^2 = \sigma_t^2(\mathcal{I}_{t-1})$. The conditional variance becomes

$$\mathrm{V}(\epsilon_t|\mathcal{I}_{t-1}) = \sigma_t^2\,\mathrm{V}(z_t|\mathcal{I}_{t-1}) = \sigma_t^2\,\mathrm{V}(z_t) = \sigma_t^2 = \sigma_t^2(\mathcal{I}_{t-1}) \tag{1.3}$$

Therefore, $\mathrm{V}(\epsilon_t|\mathcal{I}_{t-1})$ depends on past values of the conditioning variables, i.e. ϵ_t is conditionally heteroskedastic. An important property of the class of ARCH models is that the volatility is a $\mathcal{I}_{t-1}$-measurable function or, simply stated, it is deterministic given the information set in $\mathcal{I}_{t-1}$.
For the unconditional variance we have

$$\sigma_\epsilon^2 \equiv \mathrm{V}(\epsilon_t) = \mathbb{E}[\mathrm{V}(\epsilon_t|\mathcal{I}_{t-1})] + \mathrm{V}[\mathbb{E}(\epsilon_t|\mathcal{I}_{t-1})] = \mathbb{E}(\sigma_t^2).$$

It follows that the unconditional variance is constant as long as $\sigma_t^2(\mathcal{I}_{t-1})$ is mean-stationary.
The basic idea of ARCH models is to describe the dependence of ϵ_t with a simple quadratic function of its lagged values. Specifically, an ARCH(1) model assumes that

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 \tag{1.4}$$

where, in order to guarantee the positivity of the conditional variance, $\omega > 0$ and $\alpha_1 \geq 0$. The distribution of the iid innovations is often assumed to be standard normal or standard Student-t.

The name ARCH results from the fact that the conditional variance is autoregressive in the error term:

$$\begin{aligned} \mathrm{V}(\epsilon_t|\mathcal{I}_{t-1}) &= \mathbb{E}(\epsilon_t^2|\mathcal{I}_{t-1}) - \mathbb{E}(\epsilon_t|\mathcal{I}_{t-1})^2 = \mathbb{E}(z_t^2\sigma_t^2|\mathcal{I}_{t-1}) \\ &= \sigma_t^2(\mathcal{I}_{t-1}) = \omega + \alpha_1\epsilon_{t-1}^2 \end{aligned}$$

From the structure of the model, it is seen that a large past squared shock ϵ_{t-1}^2 implies a large conditional variance σ_t^2 for the innovation ϵ_t. Consequently, ϵ_t tends to assume a large value in modulus[1]. Moreover, the influence of positive and negative shocks is equal, i.e. the model is symmetric.

1.2.1 Properties of the ARCH Process

The proposition below shows the main properties of the ARCH(1) model. Not all the proofs are reported (interested readers can refer, for instance, to Tsay 2010)[41].

Proposition 1.1 (Properties of the ARCH(1) process).

(i) $\epsilon_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \nu_t \quad \text{with} \quad \nu_t := \epsilon_t^2 - \sigma_t^2 = \sigma_t^2(z_t^2 - 1)$

(ii) ϵ_t^2 is covariance-stationary if $|\alpha_1| < 1$

(iii) $\mathbb{E}(\epsilon_t) = 0, \quad \mathrm{V}(\epsilon_t) \equiv \sigma_\epsilon^2 = \frac{\omega}{1-\alpha_1}$

(iv) $\epsilon_t^2 = \sigma_\epsilon^2 + \alpha_1(\epsilon_{t-1}^2 - \sigma_\epsilon^2) + \nu_t \;\Leftrightarrow\; \sigma_t^2 = \sigma_\epsilon^2 + \alpha_1(\epsilon_{t-1}^2 - \sigma_\epsilon^2)$

(v) $\mathbb{E}(\epsilon_t^4) = \mathbb{E}(z_t^4)\,\mathbb{E}[(\sigma_t^2)^2] \geq \mathbb{E}(z_t^4)\,\mathbb{E}[(\sigma_t^2)]^2 = \mathbb{E}(z_t^4)\,\mathbb{E}[(\epsilon_t^2)]^2$

(vi) $K(\epsilon_t) := \frac{\mathbb{E}(\epsilon_t^4)}{\mathbb{E}(\epsilon_t^2)^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$ for $z_t \overset{iid}{\sim} \mathrm{N}(0,1)$ and $\alpha_1^2 \in [0, \frac{1}{3})$.

From property (i) we see that the ARCH(1) process can be written as an AR(1) process in the squared innovations, where ν_t is a *martingale difference sequence*, i.e. $\mathbb{E}(\nu_t) = 0$ and $\mathrm{cov}(\nu_t, \nu_{t-h}) = 0 \;\forall\, h > 0$. Hence, the dynamics

[1] Note that the word *tend* is used because a large variance does not necessarily produce a large realization. It only says that the probability of obtaining a large variate is greater than that of a smaller variance.

of the second moment of the series y_t follow directly from the properties of the AR(1) model, which lead to property (ii) that states that the squared innovations are covariance-stationary whenever $|\alpha_1| < 1$. It is clear that the process $\{\epsilon_t\}$ is not an independent identically distributed process. However, its unconditional mean remains zero because $\mathbb{E}(\epsilon_t) = \mathbb{E}[\mathbb{E}(\epsilon_t|\mathcal{I}_{t-1})] = \mathbb{E}[\sigma_t \mathbb{E}(z_t)] = 0$ and its unconditional variance follows from $\mathrm{V}(\epsilon_t) = \mathbb{E}(\epsilon_t^2) = \omega + \alpha_1 \mathbb{E}(\epsilon_{t-1}^2)$ and $\mathbb{E}(\epsilon_t^2) = \mathbb{E}(\epsilon_{t-1}^2) = \sigma_\epsilon^2$ given the stationarity of ϵ_t^2. Property (iv) depicts how the ARCH(1) model is able to capture the volatility clustering: when $0 \leq \alpha_1 < 1$, if ϵ_{t-1}^2 is larger (smaller) than its unconditional mean σ_ϵ^2, ϵ_t^2 is expected to be larger (smaller) than σ_ϵ^2 as well. In other words, high volatility is more likely to be followed by high volatility and vice versa. Besides capturing the volatility clustering effect, the ARCH(1) approach also depicts the fat tails typical of financial returns. In fact, property (v) shows[2] that the kurtosis of ϵ_t always exceeds the kurtosis of z_t, and property (vi) illustrates that, if z_t is standard normal distributed, the unconditional kurtosis of ϵ_t is finite as long as $\alpha_1^2 \in [0, \frac{1}{3})$ and larger than that of a normal distribution. Thus, the ARCH(1) model manages to produce a larger number of extremes than expected from an iid sequence of normal random variables, in agreement with the empirical features of most financial time series.

The ARCH(1) process can naturally be extended to an ARCH(p) process to allow for a richer dynamic structure:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 \tag{1.5}$$

Sufficient conditions to ensure the positivity of σ_t^2 are $\omega > 0$ and $\alpha_i \geq 0 \; \forall \, i = 1, \ldots, p$, while $\sum_{i=1}^{p} \alpha_i < 1$ implies that the unconditional variance of ϵ_t is finite.

Proposition 1.2 (Properties of the ARCH(p) process).

(i) $\epsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_p \epsilon_{t-p}^2 + \nu_t$

(ii) ϵ_t^2 is covariance-stationary if all the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0$ lie outside the unit circle

[2] It follows from Jensen's inequality, which states that, for any convenx function ϕ, $\mathbb{E}[\phi(X)] \geq \phi[\mathbb{E}(X)]$.

(iii) $\mathbb{E}(\epsilon_t) = 0, \quad \mathrm{V}(\epsilon_t) \equiv \sigma_\epsilon^2 = \frac{\omega}{1-\sum_{i=1}^{p} \alpha_i}$

(iv) $\sigma_t^2 = \sigma_\epsilon^2 + \alpha_1(\epsilon_{t-1}^2 - \sigma_\epsilon^2) + \cdots + \alpha_p(\epsilon_{t-p}^2 - \sigma_\epsilon^2)$

(v) $K(\epsilon_t) > 3$ for $z_t \overset{iid}{\sim} \mathrm{N}(0,1)$ and certain parameter constraints to ensure the existence of the fourth moment of ϵ_t.

1.3 GARCH Models

Although the ARCH model is simple, it often requires a lot of lags to adequately describe the volatility process of an asset return. One thus needs to estimate a high number of parameters and to impose complicated parametric conditions in order to guarantee the stationarity and the positivity of σ_t^2.
Bollerslev (1986)[5] proposed a useful and more parsimonious extension known as generalized ARCH (GARCH) model. It provides more flexible dependence patterns and, even in its simplest form, it has proven surprisingly successful in predicting conditional variances.
For a (log-)return series y_t, let $\epsilon_t = y_t - \mu_t$ be the innovation at time t. The (Gaussian) GARCH(1,1) process is then given by

$$\epsilon_t|\mathcal{I}_{t-1} \overset{iid}{\sim} \mathrm{N}(0, \sigma_t^2), \quad \sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{1.6}$$

with $\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ and $(\alpha_1 + \beta_1) < 1$.
Clearly, equation in (1.6) reduces to an ARCH(1) if $\beta_1 = 0$. The α_1 and β_1 parameters are referred to as ARCH and GARCH parameters, respectively. In the GARCH specification the conditional variance is a weighted average of the constant (long-run) variance, lagged variances and lagged squared shocks which represent the new information not captured in the lagged variances. This corresponds to a sort of adaptive learning mechanism and can be thought of as Bayesian updating. The model allows the data to determine how fast the variance adapts to new information and how fast it reverts to its long run mean. Similar to the ARCH approach, the impact of positive and negative shocks is equal.

1.3.1 Properties of the GARCH Process

Proposition 1.3 (Properties of the GARCH(1,1) process).

(i) $\epsilon_t^2 = \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t, \quad \nu_t := \epsilon_t^2 - \sigma_t^2 = \sigma_t^2(z_t^2 - 1)$

(ii) ϵ_t^2 is covariance-stationary if $|\alpha_1 + \beta_1| < 1$

(iii) $\mathbb{E}(\epsilon_t) = 0, \;\; \mathrm{V}(\epsilon_t) \equiv \sigma_\epsilon^2 = \frac{\omega}{1-\alpha_1-\beta_1}$

(iv) $\sigma_t^2 = \sigma_\epsilon^2 + \alpha_1(\epsilon_{t-1}^2 - \sigma_\epsilon^2) + \beta_1(\sigma_{t-1}^2 - \sigma_\epsilon^2)$

(v) $K(\epsilon_t) = 3\frac{(1-(\alpha_1-\beta_1)^2)}{1-2\alpha_1^2-(\alpha_1-\beta_1)^2} > 3$ for $z_t \overset{iid}{\sim} \mathrm{N}(0,1)$ and $2\alpha_1^2 + (\alpha_1 - \beta_1)^2 < 1$

(vi) $\sigma_t^2 = \omega \sum_{i=1}^{\infty} \beta_1^{i-1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \epsilon_{t-i}^2, \quad$ if $\beta_1 < 1$

(vii) Assuming stationarity and if $2\alpha_1^2 + (\alpha_1 - \beta_1)^2 < 1$, the autocorrelations of ϵ_t^2 are given by:

$$\mathrm{corr}(\epsilon_t^2, \epsilon_{t-1}^2) = \alpha_1 + \frac{\alpha_1^2\beta_1}{1-2\alpha_1\beta_1-\beta_1^2},$$
$$\mathrm{corr}(\epsilon_t^2, \epsilon_{t-h}^2) = (\alpha_1 + \beta_1)^{h-1}\,\mathrm{corr}(\epsilon_t^2, \epsilon_{t-1}^2) \quad \text{for } h = 2, 3, ...$$

Otherwise, if $2\alpha_1^2 + (\alpha_1 - \beta_1)^2 \geq 1$, one can derive the following approximations:

$$\mathrm{corr}(\epsilon_t^2, \epsilon_{t-1}^2) \approx \alpha_1 + \frac{\beta_1}{3},$$
$$\mathrm{corr}(\epsilon_t^2, \epsilon_{t-h}^2) \approx (\alpha_1 + \beta_1)^{h-1}\,\mathrm{corr}(\epsilon_t^2, \epsilon_{t-1}^2) \quad \text{for } h = 2, 3, ...$$

(viii) $\mathrm{cov}(\epsilon_t, \sigma_{t+1}^2 | \mathcal{I}_{t-1}) = \alpha_1(\omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2)^{3/2}\,\mathbb{E}(z_t^3)$.

From property (i) we see how a GARCH model can be regarded as an application of the ARMA idea to the squared series ϵ_t^2. Such a formulation is convenient to derive both conditions for stationarity, stated in property (ii), and the unconditional mean of ϵ_t^2 (i.e. the unconditional variance of ϵ_t), in property (iii). However, formulation (1.6) is easier to work with and it is the one commonly used in practice. From property (iv) one can see that a large ϵ_{t-1}^2 or σ_{t-1}^2 gives rise to a large σ_t^2, generating again the well-known volatility cluster effect. Property (v) confirms that the tails of GARCH models are thicker than those of the Gaussian distribution; moreover, given stationarity, the GARCH process reveals larger excess of kurtosis than the ARCH, i.e., adding an autoregressive term σ_t^2 in the specification of the ARCH process, the tails become heavier and deviate more from the normal distribution[3]. The ability of the GARCH model to capture more complex and large dependence patterns in conditional volatility is illustrated in property (vi). It is shown that, for $\beta_1 < 1$, any GARCH(1,1) may be seen as an ARCH(∞) process with the parameters being restricted to be functions of α_1 and β_1.

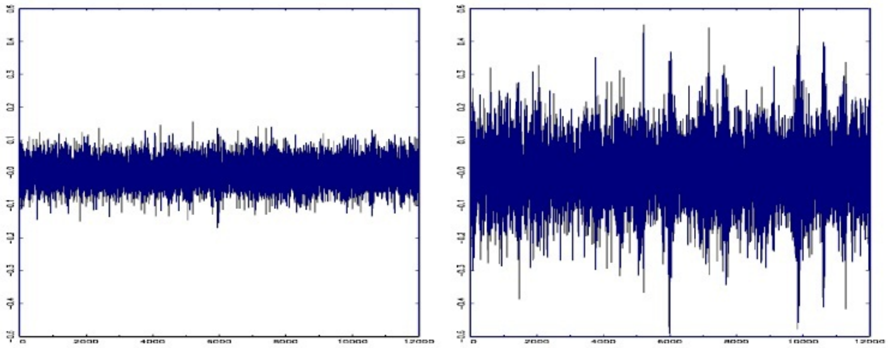


Figure 1.1: Line graph of time series with GARCH(1,1) conditional variance characterized by $\alpha_1 = 0.01, \beta_1 = 0.8$ (left panel) and by $\alpha_1 = 0.08, \beta_1 = 0.9$ (right panel).

Figure 1.1 plots time series with a GARCH(1,1) conditional variance for different values of $(\alpha_1 + \beta_1)$. We deduce that a larger $(\alpha_1 + \beta_1)$ describes series with more volatility clustering and more and larger extreme values. Also, conform to property (vii), a larger value for $(\alpha_1 + \beta_1)$ depicts a slower decay of the autocorrelation function of the squared series than for a small value of $(\alpha_1 + \beta_1)$, which is referred to as *persistence* of the GARCH[4] (see Figure 1.2). Finally, property (viii) shows that as long as z_t has a symmetric distribution, GARCH models cannot capture the leverage effect, since $\mathrm{cov}(\epsilon_t, \sigma_{t+1}^2 | \mathcal{I}_{t-1}) = 0$.

[3]However, from the kurtosis function given in property (v) it is easy to see that if $\alpha_1 = 0$, then $K(\epsilon_t) = 3$, meaning that the corresponding model does not have heavy tails.

[4]In empirical applications on daily returns series, it is generally observed $0 < \alpha_1 < 0.1$ and $0.75 < \beta_1 < 0.9$ (Gallo & Pacini, 2002 [17]).

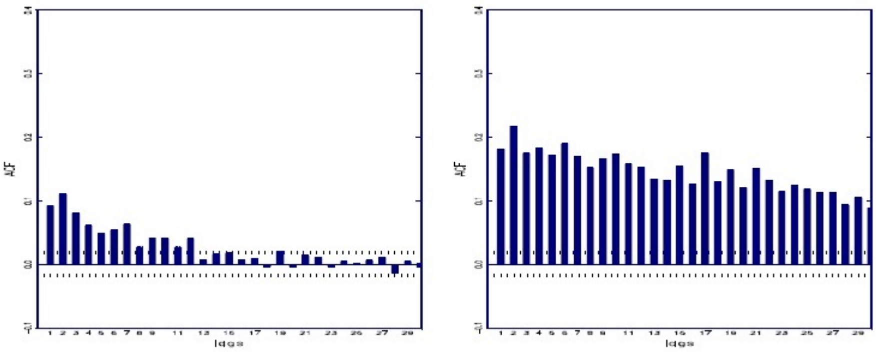


Figure 1.2: Autocorrelation function of squared time series with GARCH(1,1) conditional variance characterized by $\alpha_1 = 0.01, \beta_1 = 0.8$ (left panel) and by $\alpha_1 = 0.08, \beta_1 = 0.9$ (right panel).

A natural extension of the GARCH(1,1) model is the GARCH(p, q), where the conditional variance is defined as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1.7}$$

with, as usual, $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0$ and $\sum_{i=1}^{m}(\alpha_i + \beta_i) < 1$. Here it is understood that $m = \max(p, q)$ and $\alpha_i = 0$ for $i > p, \beta_i = 0$ for $i > q$. Equation (1.7) reduces to a pure ARCH(p) if $q = 0$.

Proposition 1.4 (Properties of the GARCH(p, q) process).

(i) $\epsilon_t^2 = \omega + \sum_{i=1}^{m}(\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-j} + \nu_t$

(ii) ϵ_t^2 is covariance-stationary if all the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m - \beta_1 z - \beta_2 z^2 - \cdots - \beta_m z^m = 0$ lie outside the unit circle

(iii) $\mathbb{E}(\epsilon_t) = 0, \quad \mathrm{V}(\epsilon_t) = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}$

(iv) $\sigma_t^2 = \sigma_\epsilon^2 + \sum_{i=1}^{p} \alpha_i(\epsilon_{t-i}^2 - \sigma_\epsilon^2) + \sum_{j=1}^{q} \beta_j(\sigma_{t-j}^2 - \sigma_\epsilon^2)$

(v) $K(\epsilon_t) > 3$ for $z_t \overset{iid}{\sim} \mathrm{N}(0, 1)$ and some parameter constraints to ensure the existence of the fourth moment of ϵ_t

(vi) $\sigma_t^2 = \frac{\omega}{(1 - \sum_{j=1}^{q} \beta_j)} + \sum_{i=1}^{\infty} \delta_i \epsilon_{t-i}^2$, where δ_i are functions of α_i and β_j.

1.3.2 Drawbacks of (G)ARCH models

The advantages of the (G)ARCH class of models include the properties discussed in the previous subsection. However, it has also some weaknesses:

- The (G)ARCH models describe the conditional variance as a function of lagged squared shocks and lagged variances. While this structure is valuable for forecasting, it does not provide any new insight for understanding the source of variations of financial markets. It is noteworthy that, if the true causes were included in the model (e.g. macroeconomic announcements, other market volatility, company specific announcements or other exogenous variables), then the lags would generally not be needed.

- It is a well-known feature of financial markets that past shocks have a high persistence on conditional volatility. However, even in the GARCH specification, that has a higher persistence than the ARCH one, the impact of large past shocks decays very fast. In fact, the GARCH representation provides an exponential declining of the autocorrelation for ϵ_t^2, albeit the empirical evidence suggests that it declines at a hyperbolic rate. The appropriate high persistence can sometimes be achieved via highly parameterization (at the cost of a higher estimation effort) or by means of other specifications such as the *Fractional Integrated* GARCH.

- To study the tail behavior of the innovations ϵ_t we have to ensure the finiteness of the fourth moment. The very restrictive conditions needed, in practice, limit the ability of (Gaussian) (G)ARCH processes to capture the excess of kurtosis of financial time series.

- In the (G)ARCH models, positive and negative shocks have the same effect on conditional volatility since it depends on the squared of the previous shocks. In practice, we observe that a falling stock price gives rise to a greater uncertainty and hence to a greater volatility. On the contrary, the reaction to positive shocks is generally lower. To circumvent this difficulty nonlinear GARCH models have been developed. In the next section the *Threshold* GARCH specification will be briefly illustrated.

1.4 The Threshold GARCH (TGARCH) Model

The models described so far have ignored the information on the direction of returns; only magnitude matters. However, as stated before, there is very convincing evidence that the direction does affect volatility. To take into account this phenomenon, the so called threshold GARCH model was proposed independently by Glosten, Jagannathan & Runkle in 1993 and Zakoian in 1994. The general TGARCH(p, q) assumes the form

$$\sigma_t^2 = \omega + \sum_{i=1}^{p}(\alpha_i \epsilon_{t-i}^2 + \gamma_i \epsilon_{t-i}^2 \mathbb{I}_{[\epsilon_{t-i}<0]}) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1.8}$$

where $\mathbb{I}_{[\cdot]}$ is the indicator function, α_i, γ_i and β_j are nonnegative parameters satisfying conditions similar to those of classic GARCH models. By introducing an interaction term of the lagged squared shocks with a dummy for the sign of the shock, the TGARCH specification manages to account for the leverage effect. In particular, if $\gamma_i > 0$, then the impact of a negative shock in $t-i$ on the conditional variance in t is larger if compared to the impact of a positive shock. Clearly, the slope from a positive to a negative shock is not smooth, but discrete.

1.5 Estimation, Forecasting and Model Diagnostic

The simplest way to estimate Gaussian GARCH models is to use maximum likelihood by substituting $sigma_t^2$ for σ^2 in the normal likelihood and then maximizing with respect to the parameters. For any set of parameters $\omega, \alpha_i, \beta_j$ and initial values for the variances and the error terms of the first m observations, it is easy to calculate the estimated variance for the $m+1$ observation, by using the updating formula (1.5) or (1.7). The starting values for the variances and for the squared lagged error terms are typically set equal to the unconditional (sample) variance of the time series. The maximum likelihood method provides a systematic way to adjust the parameters to give the best

fit. Unfortunately, the likelihood function of GARCH models is not globally concave, and so, its maximization is often problematic, relying crucially on the choice of starting values of the parameters to be estimated.
As we saw deriving the properties of (G)ARCH models, even if z_t is normal distributed, the unconditional distribution of ϵ_t is non-Gaussian, with heavier-than-normal tails. Nevertheless, there is a fair amount of evidence that the conditional distribution of ϵ_t is often non-Gaussian as well. However, the same maximum likelihood approach can be used with distribution different from the normal, such as a standardized Student-t distribution (as proposed by Bollerslev, 1987 [6]; see also Fiorentini et al., 2003 [13]) or a skew-t distribution to account also for the asymmetry of asset returns.

1.5.1 Forecast of GARCH Models

Consider the GARCH(1,1) in (1.6) and assume that the forecast origin is t. For the 1-step-ahead forecast we have

$$\sigma^2_{t+1|t} = \omega + \alpha_1 \epsilon^2_t + \beta_1 \sigma^2_t$$

where ϵ_t and σ^2_t are known at time t. The 2-step-ahead forecast is given by

$$\sigma^2_{t+2|t} = \omega + \alpha_1 \epsilon^2_{t+1|t} + \beta_1 \sigma^2_{t+1|t} = \omega + (\alpha_1 + \beta_1)\sigma^2_{t+1|t}$$

where we use the fact that $\epsilon^2_{t+1|t} = z^2_{t+1|t} \cdot \sigma^2_{t+1|t} = \sigma^2_{t+1|t}$, because

$$\mathrm{V}(z_{t+1}|\mathcal{I}_t) = \mathrm{V}(z_{t+1}) = \mathbb{E}(z^2_{t+1}) = z^2_{t+1|t} = 1$$

In general, we have

$$\sigma^2_{t+h|t} = \omega + (\alpha_1 + \beta_1)\sigma^2_{t+h-1|t} \tag{1.9}$$

and, by repeated substitutions in (1.9), we obtain that the h-step-ahead forecast for the conditional variance can be written as

$$\sigma^2_{t+h|t} = \omega \sum_{i=0}^{h-2} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1} \sigma^2_{t+1} \tag{1.10}$$

which allows the h-step-ahead forecast to be computed directly from σ^2_{t+1}. Note that σ^2_{t+1} is contained in the information set $\mathcal{I}_t$, as it can be computed

from observations $y_t, y_{t-1}, ...$ (given the knowledge of the parameters of the model).
Provided that the GARCH(1,1) is stationary (i.e. $|\alpha_1 + \beta_1| < 1$), and using $\sum_{i=0}^{n} r^i = (1 - r^{n+1})/(1 - r)$ for all r with $|r| < 1$, equation (1.10) becomes

$$\sigma^2_{t+h|t} = \frac{\omega[1 - (\alpha_1 + \beta_1)^{h-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{h-1}\sigma^2_{t+1} \tag{1.11}$$

which can be written as

$$\sigma^2_{t+h|t} = \sigma^2_\epsilon + (\alpha_1 + \beta_1)^{h-1}(\sigma^2_{t+1} - \sigma^2_\epsilon) \tag{1.12}$$

with $\sigma^2_\epsilon = \frac{\omega}{1-\alpha_1-\beta_1}$. Therefore,

$$\sigma^2_{t+h|t} \longrightarrow \sigma^2_\epsilon, \quad \text{as} \quad h \longrightarrow \infty.$$

Consequently, the multistep-ahead forecast of a GARCH(1,1) converges to the unconditional variance of ϵ_t when the forecast horizon increases to infinity (and σ^2_ϵ exists).

In the TGARCH case, the presence of the asymmetry parameter slightly changes the forecast. Consider a TGARCH(1,1) and assume that ϵ_t has a conditional distribution symmetric around zero. The 1-step-ahead forecast is given by

$$\sigma^2_{t+1|t} = \omega + \alpha_1\epsilon^2_t + \beta_1\sigma^2_t + \gamma_1\epsilon^2_t \mathbb{I}_{[\epsilon_t<0]}$$

where ϵ_t, σ^2_t and $\mathbb{I}_{[\epsilon_t<0]}$ are known at time t.
From the 2-step-ahead forecast, the sign of the innovation is not known. However, given the symmetry of the distribution of ϵ_t, we have

$$\begin{aligned}
\sigma^2_t &= \omega + \alpha_1\epsilon^2_{t+1|t} + \beta_1\sigma^2_{t+1|t} + \gamma_1\mathbb{E}(\epsilon^2_{t+1}\mathbb{I}_{[\epsilon_{t+1}<0]}|\mathcal{I}_t) \\
&= \omega + \alpha_1\sigma^2_{t+1|t} + \beta_1\sigma^2_{t+1|t} + \gamma_1\mathbb{E}(\epsilon^2_{t+1} \cdot 1|\mathcal{I}_t)\mathbb{P}(\epsilon_{t+1} < 0|\mathcal{I}_t) \\
&\quad + \gamma_1\mathbb{E}(\epsilon^2_{t+1} \cdot 0|\mathcal{I}_t)\mathbb{P}(\epsilon_{t+1} \geq 0|\mathcal{I}_t) \\
&= \omega + \alpha_1\sigma^2_{t+1|t} + \beta_1\sigma^2_{t+1|t} + \frac{\gamma_1}{2}\mathbb{E}(\epsilon^2_{t+1}|\mathcal{I}_t) \\
&= \omega + (\alpha_1 + \beta_1 + \frac{\gamma_1}{2})\sigma^2_{t+1|t}
\end{aligned} \tag{1.13}$$

Since the volatility of an asset return is not directly observable, comparing the forecasting performance of different volatility models is a challenge to econometricians. In the literature, some researchers use out-of-sample forecasts and compare the volatility forecast $\sigma^2_{t+h|t}$ with the shock ϵ^2_{t+h} to assess the forecasting performance of a model. This approach often finds a low correlation coefficient between $\sigma^2_{t+h|t}$ and ϵ^2_{t+h}, that is, low R^2. However, such a finding is not surprising because ϵ^2_{t+h} alone is not an adequate measure of the volatility at time index $t+h$. Consider the 1-step-ahead forecast: from a statistical point of view, $\mathbb{E}(\epsilon^2_{t+1}|\mathcal{I}_t) = \sigma^2_{t+1|t}$ so that ϵ^2_{t+1} is a consistent estimate of $\sigma^2_{t+1|t}$. Yet, it is not an accurate estimate of $\sigma^2_{t+1|t}$ because a single observation of a random variable cannot provide an accurate estimate of its variance. As a proxy for the latent volatility one can alternatively use the historical volatility, simply computed as the standard deviation of the returns in the given time period[5]. The existent literature shows that it is much less noisy and more precise than the estimator based on only one daily observation, and it provides more robust forecast results regardless of the loss function considered in the comparison.

1.5.2 Model building and diagnostic

The process to build a volatility model for an asset return series consists of four steps:

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, build an econometric model for the returns series to remove any linear dependence.

2. Use the residuals of the mean equation (it is often enough to remove the sample mean from the data, if significantly different from zero) and test for conditionally heteroskedasticity of ϵ^2_t.

3. Specify a volatility model and find the most appropriate specification checking the significance of the ARCH and GARCH effects.

4. Check the fitted model and refine it, if necessary.

[5]Some references about realized volatility can prove useful to investigate this problem; see, for instance, Halbleib (Chiriac) & Voev (2011)[22].

Let $\hat{\epsilon}_t = y_t - \hat{\mu}_t$ be the estimated residuals of the mean equation. The squared series $\hat{\epsilon}_t^2$ is typically used to test for conditional heteroskedasticity, which is also known as the *ARCH effect*. The test one can use is the Lagrange multiplier (LM) test of Engle (1982). It consists of regressing the squared series $\hat{\epsilon}_t^2$ on a constant and m of its own lagged values:

$$\hat{\epsilon}_t^2 = \zeta + \alpha_1 \hat{\epsilon}_{t-1}^2 + \alpha_2 \hat{\epsilon}_{t-2}^2 + \cdots + \alpha_m \hat{\epsilon}_{t-m}^2 + e_t \tag{1.14}$$

where e_t denotes the error term and $t = m+1, ..., T$. The null hypothesis is the absence of ARCH effects, namely $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. The sample size T times the R^2 from the auxiliary regression (1.14) converges in distribution to a χ^2 variable with m degrees of freedom when the null hypothesis is true and the underlying process is standard normal.
Alternatively, one can use the usual F statistic to test for the joint significance of the m coefficients in (1.14):

$$\hat{F} = \frac{R^2/m}{(1-R^2)/(T-m)} \overset{H_0}{\sim} F(m, T-m).$$

Note that to employ the LM test, it is not necessary to specify any GARCH model. Only if some coefficient is found significant, one will proceed to specify an appropriate volatility model.

The same test can be applied on the squared standardized residuals to check whether the fitted model is correctly specified or there is some remaining dynamic structure which is still unaccounted for. The squared standardized residuals are constructed as

$$\hat{z}_t^2 = \frac{\hat{\epsilon}_t^2}{\hat{\sigma}_t^2}$$

where $\hat{\sigma}_t^2$ is the estimated conditional variance from the chosen GARCH model. Hence, if one has estimated a GARCH(1,1), say, (s)he can run the auxiliary regression (1.14) with $\hat{z}_t^2$ in place of $\hat{\epsilon}_t^2$ and check if there are any residual ARCH effects not captured by the GARCH(1,1).
An alternative approach one can employ is to investigate directly the autocorrelation of $\hat{z}_t^2$ by using the Ljung-Box statistics, defined as

$$Q(m) = T(T+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l}$$

where $\hat{\rho}_l$ is the lag-sample autocorrelation. Here, the null hypothesis is H_0 : $\rho_1 = \cdots = \rho_m = 0$ against $H_1 : \rho_i = 0$ for some $i \in \{1, ..., m\}$. When the null hypothesis is true, $Q(m)$ is asymptotically distributed as a χ^2_m. Under the correct GARCH specification, the squared standardized residuals should not exhibit any autocorrelation. The Ljung-Box test is typically used with 15 lagged autocorrelations.

Finally, if we have assumed that z_t is standard normal, we can use the distributional test to check how well $\hat{z}_t$ conforms to the normality assumption. One possibility is to use the Jarque and Bera (JB) test, defined as

$$JB = \frac{\hat{S}^2(x)}{6/T} + \frac{[\hat{K}(x) - 3]^2}{24/T}$$

where $\hat{S}(x)$ and $\hat{K}(x)$ are the sample skewness and the sample kurtosis. Under the normality assumption, $\hat{S}(x)$ and $\hat{K}(x) - 3$ are distributed asymptotically as a normal with zero mean and variance $6/T$ and $24/T$ respectively, so that $JB \overset{H_0}{\sim} \chi^2_2$.

Chapter 2

Stable Distributions

The stable family of distributions constitutes a generalization of the Gaussian distribution that has intriguing theoretical and practical properties, allowing for asymmetry and heavy tails. This class of distributions was characterized by Paul Lévy in his study of independent identically distributed terms[1].
From a theoretical point of view, the use of models based on stable distributions is justified by the generalized version of the central limit theorem[2], in which the condition of finite variance is replaced by a much less restricting one concerning a regular behavior of the tails. It turns out that stable distributions are the only possible limiting laws for normalized sums of iid random variables[3].
There are also empirical reasons for modeling with stable distributions: many large data sets exhibit skewness and heavy tails and therefore they are poorly described by a Gaussian model. Since stable distributions have four parameters, two of which deal with, respectively, asymmetry and heavy-taildness, they are more adequate to model a wide range of phenomena possessing these empirical features. Examples of such data sets may be found in fields as diverse as economics, finance (see, for instance, Mandelbrot, 1963 [29]), natural sciences and engineering.
The lack of closed formulas for density and distribution functions (except for

[1] *Calcul des probabilités*, 1925.
[2] Due to Gnedenko and Kolmogorov, 1954.
[3] An excellent reference for this theory is Feller (1966)[14].

few particular cases) has been a major drawback to the use of stable distributions in applied fields. Most of these difficulties have been overcome by reliable computer programs[4] which can now compute stable pdfs, cdfs and quantiles. These programs include the algorithm proposed by Chambers et al. (1976)[9], thanks to which stable pseudo-random numbers can be straightforwardly simulated.

In the following, the main characteristics and properties of α-stable distributions will be described. No proof is reported; interested readers can find them, for instance, in Zolotarev (1986)[45], Samorodnitsky & Taqqu (1994)[38] and Nolan (2003)[33].

2.1 Definitions of stable

Definition 2.1 (**Stability, Samorodnitsky & Taqqu**). *A random variable X is said to have a* stable distribution *if and only if for any positive numbers c_1 and c_2 there exist a positive number c and a real number d such that*

$$cX + d \stackrel{d}{=} c_1 X_1 + c_2 X_2 \tag{2.1}$$

where the X_1 and X_2 are independent and have the same distribution of X. If $d = 0$, X is said to be strictly stable.

The symbol $\stackrel{d}{=}$ here means equality in distribution. The term *stable* is used because the shape of the distribution is preserved (up to scale and shift) under sums of the type (2.1). An equivalent definition of stability that can be easily derived from (2.1) is the following:

Definition 2.2 (**Stability**). *A random variable X is said to have a* stable distribution *if and only if for any natural number $n \geq 2$ there exist a positive number C_n and a real number D_n such that*

$$X \stackrel{d}{=} \frac{X_1 + X_2 + \cdots + X_n}{C_n} - D_n \tag{2.2}$$

where the X_i's are independent copies of X. If $D_n = 0\ \forall n$, X is said to be strictly stable.

[4]For example the program STABLE by Nolan, available at academic2.american.edu/ jpnolan/stable/stable.html.

Example 2.1. The normal distribution is stable. *Let us consider a random variable* $X \sim \mathrm{N}(\mu, \sigma^2)$*. The sum of n independent copies of X is* $\mathrm{N}(n\mu, n\sigma^2)$ *distributed, so setting* $C_n = \sqrt{n}$ *and* $D_n = (n-1)\mu$ *one can obtain* $X \stackrel{d}{=} [X_1 + X_2 + \cdots + X_n]/C_n - D_n$.

2.2 Characteristic function

The most concrete way to describe all possible stable distributions is by means of their characteristic function (cf), whose expression will be derived in the following theorem. It is the case to note that, since the theorem works in both directions, it also provides an alternative way of defining stable distributions.

Theorem 2.1 (**Lévy-Khintchine**)**.** *The characteristic function of a stable random variable*[5] $X \sim \mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$ *is of the form*

$$\phi_1(t) := \mathbb{E}(\mathrm{e}^{\mathrm{i}tX}) = \begin{cases} \exp\left\{\mathrm{i}\delta_1 t - \gamma^\alpha |t|^\alpha \left[1 - \mathrm{i}\beta \operatorname{sgn}(t) \tan\frac{\pi\alpha}{2}\right]\right\} & \alpha \neq 1 \\ \exp\left\{\mathrm{i}\delta_1 t - \gamma |t| \left[1 - \mathrm{i}\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln|t|\right]\right\} & \alpha = 1 \end{cases} \tag{2.3}$$

where $\theta = (\alpha, \beta, \gamma, \delta_1) \in \Theta \subseteq\ ]0, 2] \times [-1, 1] \times \mathbb{R}^+ \times \mathbb{R}$; $\operatorname{sgn}(t) = t/|t|$ *for* $t \neq 0$ *(and 0 for t=0). Conversely, if a random variable X has characteristic function of the form* (2.3)*, it is stable.*

Remark 2.1. *Note that, when* $\alpha = 1$, $\phi_1(t)$ *contains the term* $\ln|t|$ *and therefore it is not continuous with respect to the parameters, having discontinuities at all points of the form* $\alpha = 1, \beta \neq 0$*. This is a source of problems for what concerns estimation and inferential purposes*[6].

[5] Subscripts will be used to distinguish between different parameterizations that will be presented later.

[6] The main practical disadvantage of $\mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$ is that the location of the mode is unbounded in any neighborhood of $\alpha = 1, \beta \neq 0$.

Alternative parameterization

An alternative way[7] to write the cf that overcomes the problem of discontinuity is the following:

$$\phi_0(t) = \begin{cases} \exp\left\{\mathrm{i}\delta_0 t - \gamma^\alpha |t|^\alpha \left[1 + \mathrm{i}\beta \operatorname{sgn}(t) \tan \frac{\pi\alpha}{2} \left(|\gamma t|^{1-\alpha} - 1\right)\right]\right\} & \alpha \neq 1 \\ \exp\left\{\mathrm{i}\delta_0 t - \gamma |t| \left[1 + \mathrm{i}\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln(\gamma |t|)\right]\right\} & \alpha = 1 \end{cases} \tag{2.4}$$

In this case, the distribution will be denoted by $\mathrm{S}_0(\alpha, \beta, \gamma, \delta_0)$. Expression (2.4) is quite more cumbersome, and the analytic properties, as it will be shown below, have less intuitive meaning. Despite that, this formulation is much more useful for what concerns statistical applications and, unless otherwise stated, we will refer to it in what follows.

The correspondence between δ_0 in S_0 and δ_1 in S_1 is given by:

$$\delta_0 = \begin{cases} \delta_1 + \beta\gamma \tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \delta_1 + \beta \frac{2}{\pi} \gamma \ln \gamma & \text{if } \alpha = 1 \end{cases} \tag{2.5}$$

On the basis of the above relationship, a $\mathrm{S}_0(\alpha, \beta, 1, 0)$ corresponds to a $\mathrm{S}_1(\alpha, \beta, 1, -\beta \tan \frac{\pi\alpha}{2})$, provided that $\alpha \neq 1$.

2.3 Meaning and properties of the parameters

The characteristic functions presented in the previous subsection show that a general stable distribution depends on four parameters: an *index of stability* or *characteristic exponent* $\alpha \in \,]0, 2]$, an asymmetry parameter $\beta \in [-1, 1]$, a scale parameter $\gamma \in \mathbb{R}^+$ and a location parameter $\delta \in \mathbb{R}$.
The notation $X \sim \mathrm{S}_k(\alpha, \beta, \gamma, \delta)$ will be shorthand for $X \sim \mathrm{S}_0(\alpha, \beta, \gamma, \delta_0)$ and $X \sim \mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$ simultaneously. In what follows, the properties of α-stable distributions will be described analyzing the exact meaning of each parameter. Recall that the difference between parameterization 0 and 1 lies only in the parameter δ, so that the properties that concern the other parameters hold for both cases.

[7]Introduced by Zolotarev (1986)[45].

Property 2.1 (Reflection). *Let* $X_1 \sim \mathrm{S}_k(\alpha, \beta, 1, 0)$ *and* $X_2 \sim \mathrm{S}_k(\alpha, -\beta, 1, 0)$*; it then follows that* $X_2 \stackrel{d}{=} -X_1$*, i.e.* $f_2(x) = f_1(-x)$ *and* $F_2(x) = 1 - F_1(x)$*.*

Hence, when $\beta = 0$, the distribution is symmetric; on the other hand, when $\beta > 0$ the distribution turns out to be rightward skewed, i.e. $P(X > x) > P(X < -x)$ for large $|x|$. By the reflection property, the behavior of the $\beta < 0$ cases are reflections of the $\beta > 0$ ones, with left tail being thicker. The case $\beta = +1$ corresponds to a perfect positive skewness: the distribution has density zero on the negative semi-axis and positive values on the positive one. Conversely, when $\beta = -1$ the distribution is totally skewed to the left.
Note also that when $\beta = 0$, the immaginary term (the asymmetry factor) disappears, thus the two parameterization (2.3) and (2.4) coincide.

By the following result, we will identify α as the tail thickness parameter: as it decreases, tails tend to get thicker.

Property 2.2 (Tail behavior). *Let* $X \sim \mathrm{S}_0(\alpha, \beta, \gamma, \delta_0)$, $\alpha < 2$ *and* $-1 < \beta \leq 1$*. Then:*

$$\lim_{x\to\infty} \mathbb{P}(X > x) = \gamma^\alpha \frac{\Gamma(\alpha)}{\pi} \sin \tfrac{\pi\alpha}{2} (1 + \beta) x^{-\alpha} \tag{2.6}$$

$$\lim_{x\to\infty} f(x; \alpha, \beta) = \alpha\gamma^\alpha \frac{\Gamma(\alpha)}{\pi} \sin \tfrac{\pi\alpha}{2} (1 + \beta) x^{-(\alpha+1)} \tag{2.7}$$

Similar results for the left tail behavior follow straightforwardly from the reflection property.
From the above property, we can observe that:

1. in the limit, the tails behave as a power (Pareto) law[8]; when $\beta = \pm 1$ the left (right) tail decays faster than any power;

2. according to (2.6), as α increases the tails get thinner;

3. the density of the right tail is greater than the one of the left tail as $\beta > 0$, which is consistent with Property 2.1.

[8]Pareto distributions are a class of distributions with upper tail probabilities given exactly by the right hand side of (2.6); the term *stable Paretian laws* is sometimes used to distinguish between the fast decay of the Gaussian law ($\alpha = 2$) and the $\alpha < 2$ cases.

The point at which the tail approximation becomes useful is a complicated issue and depends on both the parameterization and the parameters.

The shape of a α-stable distribution is determined by α and β, which are indeed considered shape parameters. Figure 2.1[9] shows the shape of $\mathrm{S}_0(\alpha, \beta, 1, 0)$ random variables for various choices of α and β.
The γ and δ parameters determine respectively scale and location, according to the following property:

Property 2.3 (**Standardization**). *Let* $Z \sim \mathrm{S}_0(\alpha, \beta, 1, 0)$*; then:*

$$X = \begin{cases} \gamma(Z - \beta \tan \frac{\pi\alpha}{2}) + \delta_0 & \text{if } \alpha \neq 1 \\ \gamma Z + \delta_0 & \text{if } \alpha = 1 \end{cases} \tag{2.8}$$

has $S_0(\alpha, \beta, \gamma, \delta_0)$ *distribution.*

If, on the other hand, $Z \sim S_1(\alpha, \beta, 1, 0)$*, then:*

$$X = \begin{cases} \gamma Z + \delta_1 & \text{if } \alpha \neq 1 \\ \gamma(Z + \beta \frac{2}{\pi} \ln \gamma) + \delta_1 & \text{if } \alpha = 1 \end{cases} \tag{2.9}$$

has $\mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$ *distribution.*

Z is thus a sort of standardized version of X. In the sequel, a standardized α-stable distribution $\mathrm{S}_k(\alpha, \beta, 1, 0)$ will be denoted by $\mathrm{S}_k(\alpha, \beta)$.

Remark 2.2. *The characteristic function of a standardized* α*-stable distribution, symmetric around zero, reduces to*

$$\phi_k(t) = \mathrm{e}^{-|t|^\alpha}$$

for both $k = 0, 1$.

[9] From Nolan (2003)[33].

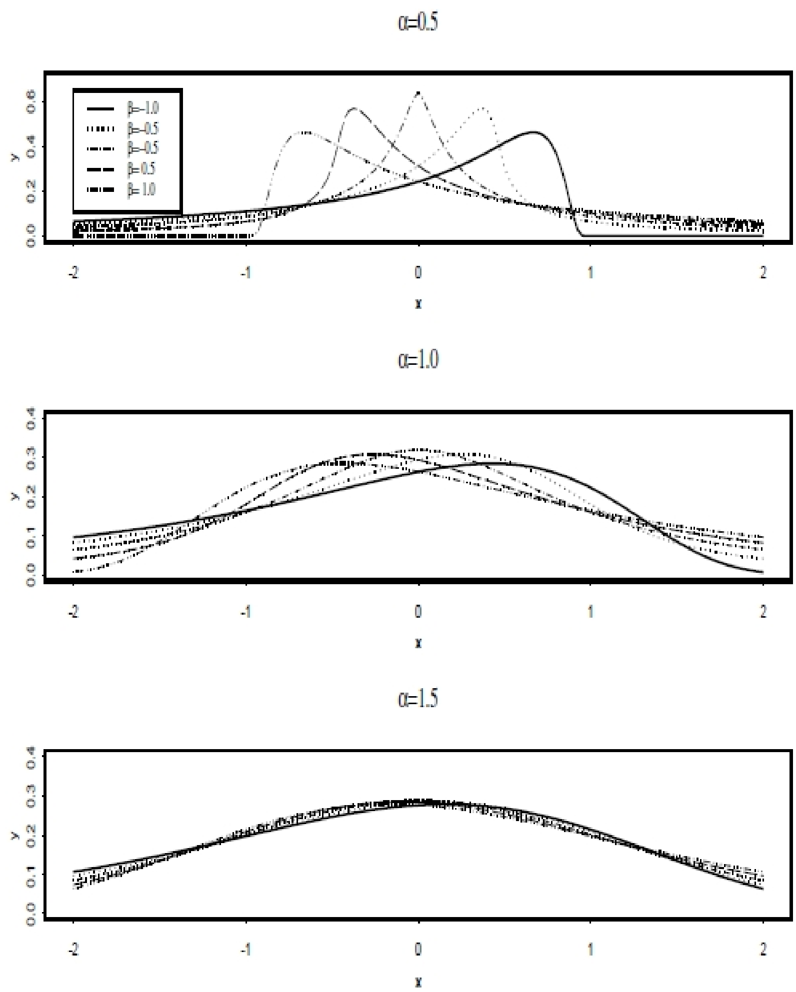


Figure 2.1: $S_0(\alpha, \beta)$ densities for different α and β.

2.4 Moments and moment properties

From the previous section it is clear that the four parameters of α-stable distributions are closely related to location, scale, asymmetry and tail thickness: one may thus argue that there is a close relationship between them and the theoretical moments. Unfortunately, one consequence of heavy tails is that not all the moments exist[10].
Let us introduce the so called *fractional absolute moments*:

$$\mathbb{E}(|X|^r) = \int_{-\infty}^{\infty} |x|^r f(x)dx \tag{2.10}$$

where r is a real number. It may be easily shown that fractional moments of order greater than α do not exist when $0 < \alpha < 2$.

Property 2.4 (**Moments**). *Let* $X \sim \mathrm{S}_k(\alpha, \beta, \gamma, \delta)$. *Then for* $0 < \alpha < 2$, $\mathbb{E}(|X|^r) < \infty$ *if and only if* $0 < r < \alpha$.

It then follows that, except for the Gaussian case ($\alpha = 2$), the variance and the higher moments never exist, while the mean does only when $\alpha > 1$. There is, in fact, a close relationship between the mean and the location parameter, as the following property shows.

Property 2.5 (**Mean**). *Let* $X \sim \mathrm{S}_0(\alpha, \beta, \gamma, \delta_0)$ *with* $\alpha > 1$. *Then*

$$\mathbb{E}(X) = \delta_0 - \beta\gamma \tan \tfrac{\pi\alpha}{2} \tag{2.11}$$

If, on the other hand, $X \sim \mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$ *with* $\alpha > 1$. *Then*

$$\mathbb{E}(X) = \delta_1 \tag{2.12}$$

We thus observe that, under parameterization 1, the location parameter coincides with the mean. On the other hand, consider what happens to the mean of $X \sim \mathrm{S}_0(\alpha, \beta)$, as $\alpha \to 1^+$. When $\beta = 0$, the distribution is symmetric and the mean is always 0; when $\beta > 0$, the mean tends to $+\infty$ because while both tails are getting heavier, the right tail is heavier than the left[11]. By reflection,

[10] This is not an issue restricted to stable laws: any distribution with power law decay will not have certain moments.

[11] In fact, $\lim_{\alpha \to 1^+} \tan \frac{\pi\alpha}{2} = -\infty$.

when $\beta < 0$ the mean tends to $-\infty$. Finally, if $\alpha = 1$, the tails are too heavy for the integral (2.10) to converge.
Recall that this is a population moment: in contrast, sample moments of all orders will exist. One can always compute, for instance, the variance of a sample. The problem is that it is not informative about stable laws because the sample variance does not converge to a well-defined population moment (unless $\alpha = 2$).

2.5 Particular cases

As claimed above, α-stable density functions admit closed form only in very few special cases: it can be easily shown, by handling the characteristic function, that the Gaussian, the Cauchy and the Lévy distributions are particular cases of the α-stable distribution.

Example 2.2 (Gaussian distribution). *When $\alpha = 2$, the stable distribution coincides with a normal with mean δ and variance $2\gamma^2$. Since $\tan\frac{\pi\alpha}{2} = 0$, the cf is real and hence the distribution is always symmetric, no matter what the value of β which becomes unidentified*[12]*; one can thus write $\mathrm{S}_k(2, 0, \gamma, \delta) = \mathrm{N}(\delta, 2\gamma^2),\ \forall\ k = 0, 1$.*

Example 2.3 (Cauchy distribution). *When $\alpha = 1$ and $\beta = 0$, the stable distribution coincides with a Cauchy distribution with position δ and scale γ: $\mathrm{S}_k(1, 0, \gamma, \delta) = \mathrm{Cauchy}(\delta, \gamma),\ \forall\ k = 0, 1$.*

Example 2.4 (Lévy distribution). *When $\alpha = 1/2$ and $\beta = \pm 1$, the stable distribution coincides with a Lévy distribution with location δ and scale γ: $\mathrm{S}_k(1/2, \pm 1, \gamma, \delta) = \text{Lévy}(\delta, \gamma),\ \forall\ k = 0, 1$.*

Figure 2.2 shows a plot of these three densities. Both normal and Cauchy distribution are symmetric and bell-shaped; the main qualitative distinction between them is that the Cauchy density has much heavier tails ($\alpha = 1$). Table (2.1) gives a numerical idea of this tail heaviness: for example, in a sample of

[12] In general, as α approaches 2, all stable distributions get closer and closer to be symmetric and β becomes less meaningful in applications (and harder to estimate accurately).

data from a normal and a Cauchy, there will be, on average, more than 100 times as many values above 3 in the Cauchy case than in the normal case. In contrast, the Lévy distribution is totally skewed to the right ($\beta = 1$) and it is even more leptokurtic than the Cauchy ($\alpha = 1/2$).

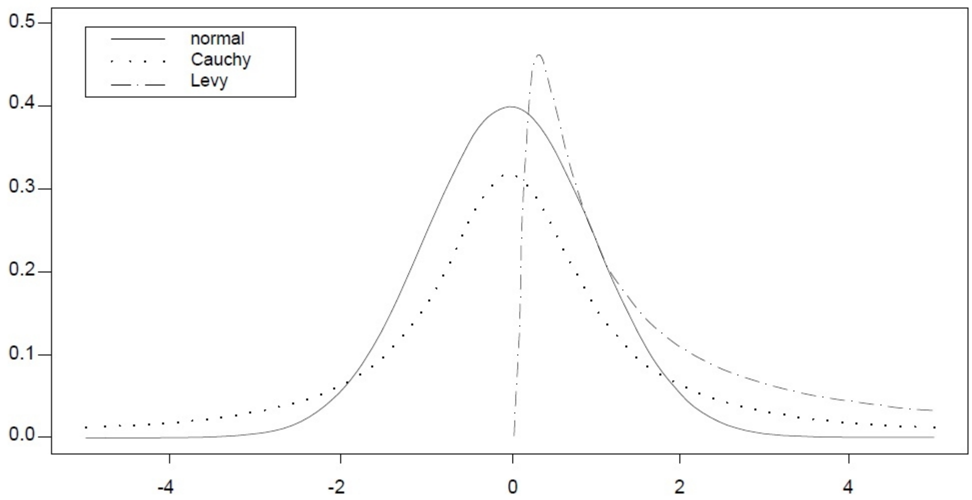


Figure 2.2: Standardized normal, Cauchy and Lévy densities.

Unfortunately, there are no more known cases in which the pdf takes a closed form. This may seem to doom the use of stable models in practice, but recall that there is no closed formula for the normal cdf as well. Furthermore, now computer programs to compute quantities of interest for α-stable distributions are available, so it is possible to use them in practical problems.

2.6 Analytic properties

Even if there are no explicit formulas for general stable densities, a few very important analytic properties concerning the probability density function have been derived.

Property 2.6 (Continuity). *Each stable distribution has continuous and infinitely differentiable probability density function.*

$P(X > c)$			
c	Std normal	Cauchy	Lévy
0	0.5000	0.5000	1.0000
1	0.1587	0.2500	0.6827
2	0.0228	0.1476	0.5205
3	0.001347	0.1024	0.4363
4	0.00003167	0.0780	0.3829
5	0.0000002866	0.0628	0.3453

Table 2.1: Comparison of tail probabilities for standard normal, Cauchy, Lévy distributions.

Property 2.7 (**Support**). *The support of stable distributions is the real line when* $|\beta| \neq 1$ *or* $\alpha \geq 1$*; otherwise it depends on the parameterization choice:*

$$\text{support } f_0(x;\alpha,\beta,\gamma,\delta_0) = \begin{cases} [\delta_0 - \gamma \tan\frac{\pi\alpha}{2}, +\infty) & \alpha < 1, \beta = 1 \\ (-\infty, \delta_0 + \gamma \tan\frac{\pi\alpha}{2}] & \alpha < 1, \beta = -1 \\ (-\infty, +\infty) & \textit{otherwise} \end{cases} \tag{2.13}$$

$$\text{support } f_1(x;\alpha,\beta,\gamma,\delta_1) = \begin{cases} [\delta_1, +\infty) & \alpha < 1, \beta = 1 \\ (-\infty, \delta_1] & \alpha < 1, \beta = -1 \\ (-\infty, +\infty) & \textit{otherwise} \end{cases} \tag{2.14}$$

The notation $f_k(\cdot)$ has been used to indicate the pdf of $\mathrm{S}_k(\alpha,\beta,\gamma,\delta)$.

Property 2.8 (**Mode**). *Stable distributions are unimodal. For symmetric stable distributions with* $1 < \alpha \leq 2$*, the mode coincides with the mean* (2.11) *or* (2.12)*; in the other cases it takes no closed form and needs to be numerically computed.*

Linear transformations and combinations

Let us introduce a useful result concerning linear transformations of stable distributions.

Property 2.9 (Linear transformations). *If $X \sim \mathrm{S}_0(\alpha, \beta, \gamma, \delta_0)$, then for any $a \neq 0, b \in \mathbb{R}$,*

$$aX + b \sim \mathrm{S}_0(\alpha, \beta \operatorname{sgn}(a), |a|\gamma, a\delta_0 + b) \tag{2.15}$$

If instead $X \sim \mathrm{S}_1(\alpha, \beta, \gamma, \delta_1)$, then

$$aX + b \sim \begin{cases} \mathrm{S}_1(\alpha, \beta \operatorname{sgn}(a), |a|\gamma, a\delta_1 + b) & \text{if } \alpha \neq 1 \\ \mathrm{S}_1(\alpha, \beta \operatorname{sgn}(a), |a|\gamma, a\delta_1 + b - \beta\gamma\frac{2}{\pi} a \ln|a|) & \text{if } \alpha = 1 \end{cases} \tag{2.16}$$

The above property shows that γ and δ are standard scale and location parameters in the $k = 0$ parameterization, but not in the $k = 1$ parameterization when $\alpha = 1$.

As stated before, a basic property of stable laws is that sums of α-stable random variables are α-stable. The exact parameters are given below. In these results, it is essential that the summands all have the same α, otherwise the sum would not be stable. When the summands are dependent the precise statement is more complicated, but the sum is still stable.

Property 2.10 (Linear combinations). *Let $X_1 \sim \mathrm{S}_0(\alpha, \beta_1, \gamma_1, \delta_1)$ and $X_2 \sim \mathrm{S}_0(\alpha, \beta_2, \gamma_2, \delta_2)$ and let $X_1 \perp X_2$. Then, $X_1 + X_2 \sim \mathrm{S}_0(\alpha, \beta, \gamma, \delta)$ with*

$$\beta = \frac{\beta_1\gamma_1^\alpha + \beta_2\gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}, \; \gamma^\alpha = \gamma_1^\alpha + \gamma_2^\alpha,$$

$$\delta = \begin{cases} \delta_1 + \delta_2 + \tan\frac{\pi\alpha}{2}(\beta\gamma - \beta_1\gamma_1 - \beta_2\gamma_2) & \alpha \neq 1 \\ \delta_1 + \delta_2 + \frac{2}{\pi}(\beta\gamma\ln\gamma - \beta_1\gamma_1\ln\gamma_1 - \beta_2\gamma_2\ln\gamma_2) & \alpha = 1 \end{cases}$$

If, on the other hand, $X_1 \sim S_1(\alpha, \beta_1, \gamma_1, \delta_1)$ and $X_2 \sim \mathrm{S}_1(\alpha, \beta_2, \gamma_2, \delta_2), X_1 \perp X_2$, then $X_1 + X_2 \sim \mathrm{S}_1(\alpha, \beta, \gamma, \delta)$ with

$$\beta = \frac{\beta_1\gamma_1^\alpha + \beta_2\gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}, \;\; \gamma^\alpha = \gamma_1^\alpha + \gamma_2^\alpha, \;\; \delta = \delta_1 + \delta_2$$

Note that $\gamma^\alpha = \gamma_1^\alpha + \gamma_2^\alpha$ is a generalized version of the additive rule for the variances of independent random variables: $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

By induction, one can get formulas for sums of n independent stable variables: for $X_j \sim S_k(\alpha, \beta_j, \gamma_j, \delta_j)$, $j = 1, 2, ..., n$ and arbitrary $w_1, ..., w_n$, the sum

$$w_1 X_1 + w_2 X_2 + ... + w_n X_n \sim S_k(\alpha, \beta, \gamma, \delta) \tag{2.17}$$

where

$$\beta = \frac{\sum_{j=1}^n \beta_j \operatorname{sgn}(w_j) |w_j \gamma_j|^\alpha}{\gamma^\alpha}, \quad \gamma^\alpha = \sum_{j=1}^n |w_j \gamma_j|^\alpha,$$

$$\delta = \begin{cases} \sum_j w_j \delta_j + \tan \frac{\pi\alpha}{2} (\beta\gamma - \sum_j \beta_j w_j \gamma_j) & k = 0, \ \alpha \neq 1 \\ \sum_j w_j \delta_j + \frac{2}{\pi} (\beta\gamma \ln \gamma - \sum_j \beta_j w_j \gamma_j \ln |w_j \gamma_j|) & k = 0, \ \alpha = 1 \\ \sum_j w_j \delta_j & k = 1, \ \alpha \neq 1 \\ \sum_j w_j \delta_j - \frac{2}{\pi} \sum_j \beta_j w_j \gamma_j \ln |w_j| & k = 1, \ \alpha = 1 \end{cases}$$

This is a generalization of (2.1): it allows different skewness, scales and locations in the terms. Note that if $\beta_j = 0$ for all j, then $\beta = 0$ and $\delta = \sum_j w_j \delta_j$. An important case is the *scaling property* for stable random variables: when the terms are independent and identically distributed, say $X_j \sim S_k(\alpha, \beta, \gamma, \delta)$, then

$$X_1 + ... + X_n \sim S_k(\alpha, \beta, n^{1/\alpha}\gamma, \delta_n) \tag{2.18}$$

where

$$\delta_n = \begin{cases} n\delta + \gamma\beta \tan \frac{\pi\alpha}{2} (n^{1/\alpha} - n) & k = 0, \ \alpha \neq 1 \\ n\delta + \gamma\beta \frac{2}{\pi} n \ln n & k = 0, \ \alpha = 1 \\ n\delta & k = 1 \end{cases}$$

It turns out that the shape of the sum of n iid terms is the same as the original shape. Note that no other distribution has this property.

2.7 Simulation

For the three special cases introduced above, there are simple ways to generate stable random variables. For the normal case, if U_1, U_2 denote independent Uniform$(0, 1)$ random variables, then

$$\begin{aligned} X_1 &= \mu + \sigma \sqrt{-2 \ln U_1} \cos 2\pi U_2 \\ X_2 &= \mu + \sigma \sqrt{-2 \ln U_1} \sin 2\pi U_2 \end{aligned}$$

give two independent $\mathrm{N}(\mu, \sigma^2)$ random variables[13].

For the Cauchy case, denoting with U a Uniform(0,1),

$$X = \gamma \tan\left(\pi(U - \tfrac{1}{2})\right) + \delta$$

is Cauchy(γ, δ).

For the Lévy case,

$$X = \gamma \frac{1}{Z^2} + \delta$$

is Lévy(γ, δ) as long as $Z \sim \mathrm{N}(0,1)$.

In the general case, the following result gives a method for simulating any stable random variate.

Theorem 2.2 (**Chambers, Mallows & Stuck**). *Let V and W be independent with V uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$, W exponentially distributed with mean 1, and $0 < \alpha \leq 2$.*
(a) The symmetric random variable

$$Z = \begin{cases} \frac{\sin \alpha V}{(\cos V)^{1/\alpha}} \left[\frac{\cos((\alpha-1)V)}{W}\right]^{(1-\alpha)/\alpha} & \text{if } \alpha \neq 1 \\ \tan W & \text{if } \alpha = 1 \end{cases} \tag{2.19}$$

has a $S_0(\alpha, 0) = S_1(\alpha, 0)$ distribution.
(b) In the nonsymmetric case, for any $-1 \leq \beta \leq 1$, define $\zeta = \arctan(\beta \tan \frac{\pi\alpha}{2})/\alpha$, then

$$Z = \begin{cases} \frac{\sin \alpha(\zeta + V)}{(\cos \alpha\zeta \cos V)^{1/\alpha}} \left[\frac{\cos(\alpha\zeta + (\alpha-1)V)}{W}\right]^{(1-\alpha)/\alpha} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi}\left[(\frac{\pi}{2} + \beta V)\tan V - \beta \ln\left(\frac{\frac{\pi}{2} W \cos V}{\frac{\pi}{2} + \beta V}\right)\right] & \text{if } \alpha = 1 \end{cases} \tag{2.20}$$

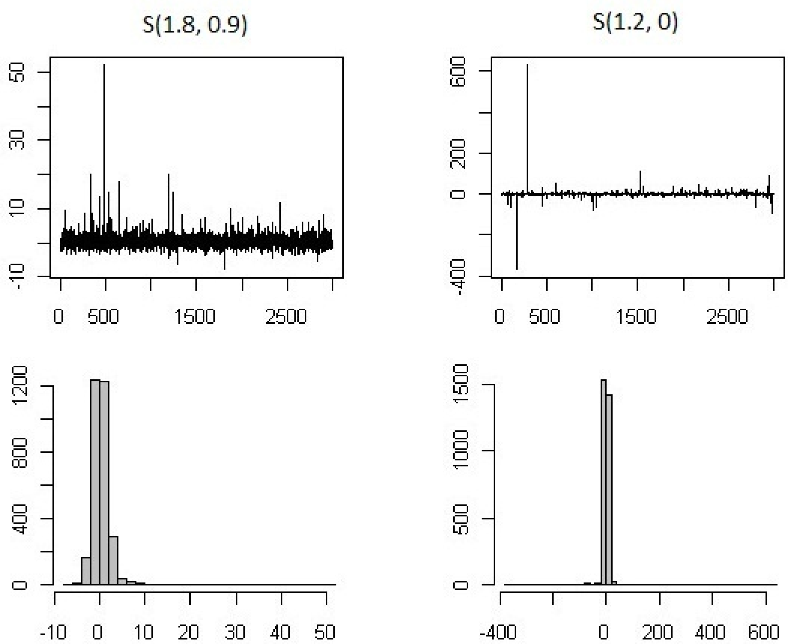
has a $S_0(\alpha, \beta)$ distribution[14].

It is easy to get V and W from two independent Uniform(0,1) random variables U_1 and U_2: set $V = \pi(U_1 - \frac{1}{2})$ and $W = -\ln U_2$.
Pseudo-random numbers for the general case containing also the position and the scale parameters δ and γ may be straightforwardly obtained using

[13] This is known as the Box-Muller algorithm.

[14] Despite Weron (1996)[42] reports a slightly different formula for $\alpha = 1$ in (2.20), the correct version is the one given in Chambers et al. (1976)[9]; for details, see Weron (1996)[43].

the standardization Property 2.3. Similarly, pseudo-random numbers with $S_1(\alpha, \beta, \gamma, \delta_1)$ distribution can be obtained exploiting (2.5).
Figure 2.3 gives the pattern, the histogram and the summary statistics of two different simulated random vectors[15].



Observations	3000
Mean	0,2975
Minimum	-7,7980
Maximum	51,990
Median	0,1091
Std. Dev.	2,0173
Skewness	7,19379
Kurtosis (excess)	153,5295
Jarque-Bera (p-value)	0,00000

Observations	3000
Mean	0,1283
Minimum	-367,60
Maximum	629,70
Median	-0,0582
Std. Dev.	14,5442
Skewness	21,6930
Kurtosis (excess)	1307,085
Jarque-Bera (p-value)	0,00000

Figure 2.3: Pattern, histogram and summary statistics for a $S_0(1.8, 0.9)$ and a $S_0(1.2, 0)$ random vectors.

[15] These results have been obtained by means of the statistical package R.

Chapter 3

Indirect Inference

Indirect inference[1] (Gouriéroux, Monfort & Renault, 1993 [21]) is an inferential approach based on simulation which is suitable for many situation in which the estimation of the model of interest is too difficult to be performed directly. Econometric models often lead to complex formulations for the conditional distributions of the endogenous variables; these formulations may even be such that it is impossible to efficiently estimate the parameters of interest because of the analytical intractability of the likelihood function. In such cases one can replace the model of interest with an approximated one which is easier to handle (*auxiliary model*). The first requirement is that it be possible to simulate data from the initial model.
The idea of indirect inference is to "calibrate" the parameters of the model of interest so that the parameters of the auxiliary model, estimated using either the observed and the simulated data, turn out to be as close as possible.

3.1 Framework and notation

Let $y_t,\ t = 1, ..., T$ be the endogenous variables, i.e. the variables whose values have to be explained by the econometric model. Let us denote by $z_t,\ t = 1, ..., T$ a set of exogenous variables, in the sense that we are interested in the conditional distribution of $\mathbf{y_t}$ given $\mathbf{z_t}$ and initial conditions $\mathbf{y_0}$:

[1] This chapter was mainly inspired by Gouriéroux & Monfort (1996)[20].

$f_0(y_1, ..., y_T | z_1, ..., z_T, \mathbf{y_0})$. This probability density function may be decomposed into

$$\begin{aligned} & f_0(y_1, ..., y_T | z_1, ..., z_T, \mathbf{y_0}) \\ & = \prod_{t=1}^{T} f_0(y_t | z_1, ..., z_T, \mathbf{y_0}, y_1, ..., y_{t-1}) \\ & = \prod_{t=1}^{T} f_0(y_t | z_1, ..., z_T, \mathbf{y_{t-1}}), \text{ with } \mathbf{y_t} = (\mathbf{y_0}, y_1, ..., y_t). \end{aligned}$$

Assuming that there is no *Granger causality*[2] of z_t on y_t, one can write

$$f_0(y_1, ..., y_T | z_1, ..., z_T, \mathbf{y_{t-1}}) = f_0(y_1, ..., y_T | z_1, ..., z_t, \mathbf{y_{t-1}}) \;\; \forall \, t.$$

In other words, the z_t are strongly exogenous variables[3]. Under this condition,

$$f_0(y_1, ..., y_T | z_1, ..., z_T, \mathbf{y_0}) = \prod_{t=1}^{T} f_0(y_t | x_t) \tag{3.1}$$

where $x_t = (\mathbf{z_t}, \mathbf{y_{t-1}})$.
Note that we have implicitly assumed that the conditional pdf $f_0(y_t|x_t)$ does not depend on t: more precisely we assume that the process $\{y_t, z_t\}$ is strongly stationary.
In order to make inference about $f_0(y_t|x_t)$ we introduce a conditional parametric model (M). This model is a family of conditional distributions indexed by a p-dimensional parameter θ:

$$M = \{f(y_t|x_t; \theta), \theta \in \Theta\}$$

where $\Theta \subset \mathbb{R}^p$. This model is assumed to be well-specified, that is $f_0(y_t|x_t)$ belongs to M, and identifiable, i.e. there exists a unique (unknown) value θ_0 such that $f_0(y_t|x_t) = f(y_t|x_t; \theta_0)$; θ_0 is the "true" value of the parameter.
This set of general assumption is denoted by **(A1)**.

[2] z_t fails to "Granger cause" y_t if in a regression of y_t on lagged y_t's and lagged z_t's, the coefficients of the latter are not significantly different from zero. Simply stated, the term "Granger causality" means "precedence" (see, for instance, Maddala, 2001 [28]).

[3] Note that models in which non strongly-exogenous variables appear are not simulable (Gouriéroux et al., 1993, Section 2.1).

3.2 The Principle

3.2.1 The econometric model

We consider an econometric model defined by the following reduced form:

$$y_t = f(\mathbf{y_{t-1}}, \mathbf{z_t}, \epsilon_t; \theta) \tag{3.2}$$

where f is a known function and ϵ_t is a white noise whose distribution is known. Thus, it is possible to generate independent random draws[4] of ϵ_t and to obtain artificial values $y_1^s, ..., y_T^s$ conditional on a given value of the parameter θ, an observed path of the exogenous variables $\mathbf{z_t}$, and on initial values. Obviously, for the lagged endogenous variables, one can use either observed (*conditional simulations*) or simulated values (*path simulations*).

3.2.2 The auxiliary model

The estimation of the econometric model (3.2) may be so complex and discouraging that econometricians replace it with an approximation, easier to handle, like

$$y_t = f^a(\mathbf{y_{t-1}}, \mathbf{z_t}, \eta_t; \beta) \tag{3.3}$$

where f^a has a convenient analytical expression, η_t are random terms, and $\beta \in B \subset \mathbb{R}^q$ is assumed to be easily estimable. Its estimation may be, for instance, based on an approximation of the exact likelihood, or on an exact likelihood of an approximated model. Since this model is misspecified, a simple estimator of β that uses the observed data, given by

$$\hat{\beta} = \arg\max_{\beta} \sum_{t=1}^{T} \ln f^a(y_t | \mathbf{y_{t-1}}, \mathbf{z_t}; \beta) \tag{3.4}$$

is generally an inconsistent estimator of θ: the idea of indirect inference is to exploit simulations performed under the original model to correct for the asymptotic bias of $\hat{\beta}$.

[4] Recall that the term "random number generation" is an oxymoron: these generators use deterministic devices to produce chains of numbers that mimic the properties of a realization from the target distribution. Therefore, a more accurate term is *pseudo-random numbers*.

3.2.3 Indirect estimation

The **first step** consists of computing the pseudo-maximum likelihood estimate (PML) of β, denoted as $\hat{\beta}$, using the observed endogenous y_t.
In the **second step**, one simulates a set of S vectors of size T from the initial model on the basis of a tentative value for the true vector of parameters, $\tilde{\theta}$, say[5]. It will be used as a starting point for the iterative calibration procedure. Then one estimates the parameter β of the auxiliary model from the pseudo-random series $y_t^s(\tilde{\theta})$:

$$\hat{\beta}(\tilde{\theta}) = \arg\max_{\beta} \sum_{s=1}^{S} \sum_{t=1}^{T} \ln f^a(y_t^s(\tilde{\theta}) | \mathbf{y_{t-1}^s}(\tilde{\theta}), \mathbf{z_t}; \beta) \tag{3.5}$$

Finally (**third step**), an indirect inference estimator of θ is defined by choosing a value of $\hat{\theta}$ for which $\hat{\beta}$ and $\hat{\beta}(\tilde{\theta})$ are as close as possible:

$$\hat{\theta}(\Omega) = \arg\min_{\theta} [\hat{\beta} - \hat{\beta}(\theta)]' \; \Omega \; [\hat{\beta} - \hat{\beta}(\theta)] \tag{3.6}$$

where Ω is a symmetric nonnegative definite matrix, which defines the metric. The estimation step is performed with a numerical algorithm, which computes $\hat{\theta}(\Omega)$ as:

$$\hat{\theta}(\Omega) = \lim_{p \to \infty} \tilde{\theta}^{(p)}$$

where

$$\tilde{\theta}^{(p+1)} = h\big(\tilde{\theta}^{(p)}, \hat{\beta}(\tilde{\theta}^{(p)})\big)$$

and $h(\cdot)$ is the updating function of the algorithm. In practical terms, the two vectors of parameters $\hat{\beta}$ and $\hat{\beta}(\tilde{\theta})$ are compared. If they are "very close", the procedure has come to its end, otherwise the tentative values $\tilde{\theta}$ are modified (calibrated) and the procedure starts again from the second step. The iterations continue until the quadric form (3.6) is minimized.
A very important point is that the pseudo-random errors ϵ_t, generated and plugged into equation (3.2) in order to obtain the $y_t^s(\tilde{\theta})$, must not be re-generated. The values of the series $y_t^s(\tilde{\theta})$ change across iterations only as an effect of changing $\tilde{\theta}$.

[5] As initial value of $\tilde{\theta}$ one can use $\tilde{\theta}^{(0)} = \hat{\beta}$.

3.2.4 Estimation based on the score

An alternative approach, introduced by Gallant & Tauchen (1996)[16], considers directly the score of the auxiliary model:

$$\sum_{t=1}^{T} \frac{\partial \ln f^a}{\partial \beta}(y_t|\mathbf{y_{t-1}}, \mathbf{z_t}; \beta) \tag{3.7}$$

which is clearly equal to zero for the PML estimator of β. For the sake of simplicity, $\sum_{t=1}^{T} \ln f^a(y_t|\mathbf{y_{t-1}}, \mathbf{z_t}; \beta)$ will be denoted hereafter as $\mathcal{L}^a(y_t; \beta)$.
The idea is to choose θ such that the score, computed on the simulated observations, results as close as possible to zero. Namely,

$$\check{\theta}(\Sigma) = \arg\min_{\theta} \left\{ \sum_{s=1}^{S} \frac{\partial \mathcal{L}^a}{\partial \beta} \left[y_t^s(\theta); \hat{\beta} \right] \right\}' \Sigma \left\{ \sum_{s=1}^{S} \frac{\partial \mathcal{L}^a}{\partial \beta} \left[y_t^s(\theta); \hat{\beta} \right] \right\} \tag{3.8}$$

where Σ is a symmetric nonnegative definite matrix. As usual, the estimate is obtained minimizing (3.8) by means of a numerical algorithm.
Provided that a closed form for the gradient of the auxiliary model is available, this approach has an important computational advantage: it allows one to avoid the numerical optimization for the estimation of $\hat{\beta}(\tilde{\theta})$, for different tentatives of the parameter of interest.

3.3 Properties of the I.I. estimators

3.3.1 The dimension of the auxiliary parameter

First, one should note that the dimension of the auxiliary parameter β must be greater than or equal to the dimension of the parameter of interest θ, in order to get a unique solution $\hat{\theta}$ (or $\check{\theta}$). It is a kind of order identifiability condition.
Second, when the problem is just identified, i.e. the dimension of the parameter vectors agree, the estimator enjoys three nice properties.

Proposition 3.1 (Identification). *If* $\dim \beta = \dim \theta$ *and* T *is sufficiently large:*

1. $\hat{\theta}(\Omega) = \hat{\theta}$

2. $\check{\theta}(\Sigma) = \check{\theta}$

3. $\hat{\theta} = \check{\theta}$

In other words, when the number of parameters of the econometric model and of the auxiliary one is the same ($p = q$), the results are independent of the choice of the matrices that define the metrics. Minimization of the quadratic forms (3.6) and (3.8) is in fact clearly obtained when $\hat{\beta}(\tilde{\theta}) = \hat{\beta}$ (for a complete proof of the proposition, see Appendix A). On the contrary, when $q > p$ it is necessary to choose a metric to measure the distance between $\hat{\beta}$ and $\hat{\beta}(\tilde{\theta})$.
Furthermore, in the just identified case, the two different approaches yield identical results, so that we can choose the one that suits the best for the practical problem to be analyzed.

3.3.2 The binding function

In order to assess the asymptotic properties of indirect inference estimators, it is worth introducing a concept that will be very useful.
Let us consider the asymptotic behavior of the log-likelihood function of the auxiliary model:

$$\lim_{t\to\infty} \frac{1}{T}\mathcal{L}^a(y_t;\beta) = \mathbb{E}_\theta\left[\mathcal{L}^a(y_t;\beta)\right]$$

The solution of the optimization problem in this asymptotic setting is then:

$$b(\theta) := \arg\max_{\beta\in B} \mathbb{E}_\theta\left[\mathcal{L}^a(y_t;\beta)\right] \tag{3.9}$$

It thus turns out that, $\forall\, \theta \in \Theta$, $\hat{\beta}(\theta)$ is a consistent estimator of $b(\theta)$.
The function $b : \Theta \to B$, $b(\cdot) = \arg\max_{\beta\in B} \mathbb{E}_{(\cdot)}\left[\mathcal{L}^a(y_t;\beta)\right]$, called *binding function*, maps the parameter (sub-)space of the model of interest onto the parameter space of the auxiliary model.
The indirect inference estimator of θ is based on the pseudo-true value of β, i.e. the value of the binding function evaluated at the true value of the parameter of interest: $b(\theta_0)$. Indeed, under the assumption that the observed data are generated by the econometric model (whose parameter's true value is θ_0), $\hat{\beta}$, estimated in the observed data, converges to the pseudo-true value $b(\theta_0)$.
To sum up, the indirect inference based on auxiliary PML estimator, consist in:

- determining $\hat{\beta}$, a direct consistent estimator of $b(\theta_0)$
- determining $\hat{\beta}(\tilde{\theta})$, a direct consistent estimator of the function $b(\cdot)$ (when $T \to \infty, S$ fixed)
- solving approximately $b(\theta_0) = b(\theta)$ to get an estimator of θ_0

In the finite sample, the calibration procedure aims at solving the system of equation

$$\hat{\beta}(\tilde{\theta}) = \hat{\beta} \tag{3.10}$$

These equations are only implicitly defined and usually cannot be expressed in closed form. It is usually possible to solve the system (3.10) only in the just identified case, where the number of unknowns ($\tilde{\theta}$) equals the number of equations ($\dim \beta$).
Note that the maximization of $\mathbb{E}_\theta\big[\mathcal{L}^a(y_t;\beta)\big]$ with respect to β is equivalent to the minimization of the Kullback-Leibler information criterion:

$$KLIC := \mathbb{E}_\theta \frac{f(y_t|\mathbf{y_{t-1}}, \mathbf{z_t};\theta)}{f^a(y_t|\mathbf{y_{t-1}}, \mathbf{z_t};\beta)}$$

that underlines the importance of the proximity of the auxiliary model to the econometric one. $f^a(y_t|\mathbf{y_{t-1}}, \mathbf{z_t}; b(\theta))$ corresponds to the conditional distribution of the approximated model that is the closest to $f(y_t|\mathbf{y_{t-1}}, \mathbf{z_t};\theta)$.

3.3.3 Asymptotic Properties

The asymptotic properties of the I.I. estimators are given below for a general criterion function (like, for instance, the log-likelihood)

$$\psi_T(y_t^s(\theta);\beta)$$

Let us add some regularity conditions that will be needed in proving such properties (proofs are reported in Appendix A).

(A2) The criterion function $\psi_T(y_t^s(\theta);\beta)$ tends almost surely, as $T \to \infty$, to a deterministic limit function $\psi_\infty(\theta,\beta)$

(A3) ψ_T and ψ_∞ are differentiable with respect to β, and ψ_∞ has a unique maximum (w.r.t. β): $b(\theta) = \arg\max_\beta \psi_\infty(\theta,\beta)$

(A4) The only solution of the asymptotic first order conditions is $b(\theta)$: $\frac{\partial \psi_\infty}{\partial \beta}(\theta, \beta) = 0 \Rightarrow \beta = b(\theta)$

(A5) The binding function is injective and its first derivative with respect to θ is of full column rank.

Proposition 3.2 (Consistency). *Under condition (A1)-(A5), the indirect inference estimator $\hat{\theta}(\Omega)$ is consistent for S fixed and $T \to \infty$.*

In order to prove the asymptotic normality, it is necessary to add three more conditions about the behavior of the auxiliary model's criterion function:

(A6) The Hessian matrix of the criterion function converges to a non-stochastic limit:

$$\mathrm{plim}_{T\to\infty} - \frac{\partial^2 \psi_T}{\partial \beta \partial \beta'}[y_t^s(\theta); b(\theta)] = -\frac{\partial^2 \psi_\infty}{\partial \beta \partial \beta'}[\theta_0; b(\theta_0)] := J_0$$

(A7) The gradient of the criterion function converges, in distribution, to a normal law:

$$\sqrt{T}\frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0); b(\theta_0)] \xrightarrow[T\to\infty]{d} \mathrm{N}(0, I_0)$$

$$I_0 := \mathrm{plim}_{T\to\infty} \ \mathrm{V}\Big\{\sqrt{T}\frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta); b(\theta)]\Big\}$$

(A8) The asymptotic covariance between the gradients of two units, s_1 and s_2, of the simulated sample is costant:

$$\lim_{T\to\infty} \mathrm{cov}_0\Big\{\sqrt{T}\frac{\partial \psi_T}{\partial \beta}[y_t^{s_1}(\theta_0); b(\theta_0)], \sqrt{T}\frac{\partial \psi_T}{\partial \beta}[y_t^{s_2}(\theta_0); b(\theta_0)]\Big\}$$
$$:= K_0, \ \forall\, s_1 \neq s_2.$$

Proposition 3.3 (Asymptotic normality). *Under assumptions (A1)-(A8) and the usual regularity conditions, the indirect inference estimator $\hat{\theta}(\Omega)$ is asymptotically normal, for S fixed and $T \to \infty$:*

$$\sqrt{T}(\hat{\theta}(\Omega) - \theta_0) \xrightarrow{d} \mathrm{N}\big(0, W(S, \Omega)\big)$$

$$\begin{aligned} W(S,\Omega) = \Big(1 + \frac{1}{S}\Big) \Big[\frac{\partial b'}{\partial \theta}(\theta_0)\,\Omega\,\frac{\partial b}{\partial \theta'}(\theta_0)\Big]^{-1} \frac{\partial b'}{\partial \theta}(\theta_0) \\ \times\,\Omega\, J_0^{-1}(I_0 - K_0) J_0^{-1}\,\Omega\,\frac{\partial b}{\partial \theta'}(\theta_0)\Big[\frac{\partial b'}{\partial \theta}(\theta_0)\,\Omega\,\frac{\partial b}{\partial \theta'}(\theta_0)\Big]^{-1}. \end{aligned}$$

The indirect inference estimator $\hat{\theta}(\Omega)$ forms a class of estimators indexed by the matrix Ω. In fact, the asymptotic variance-covariance matrix depends on the metric Ω and, as usual, there is an optimal choice of this matrix.

Proposition 3.4 (**Optimality**). *The optimal choice of the Ω matrix, assuming that $(I_0 - K_0)$ is invertible, is*

$$\Omega^* = J_0\,(I_0 - K_0)^{-1}\,J_0\,,$$

i.e. $W_S^* := W(S, \Omega^*) = \min_\Omega W(S, \Omega)$, *and*

$$W_S^* = \Big(1 + \frac{1}{S}\Big)\Big[\frac{\partial b'}{\partial \theta}(\theta_0)\,J_0\,(I_0 - K_0)^{-1}\,J_0\,\frac{\partial b}{\partial \theta'}(\theta_0)\Big]^{-1}.$$

The optimal estimator thus obtained is denoted by $\hat{\theta}^*$.
Note that in the exact identified case ($\dim\beta = \dim\theta$) the estimator and, therefore, its asymptotic precision, are independent of Ω. Hence, being the Jacobian $\frac{\partial b}{\partial \theta'}(\theta_0)$ invertible (from (A5)), the variance-covariance matrix reduces to

$$W(S, \Omega) = \Big(1 + \frac{1}{S}\Big)\Big[\frac{\partial b'}{\partial \theta}(\theta_0)\,J_0\,(I_0 - K_0)^{-1}\,J_0\,\frac{\partial b}{\partial \theta'}(\theta_0)\Big]^{-1} \tag{3.11}$$

Therefore, we have $W(S, \Omega) = W_S^*\ \forall\,\Omega$.

Similar results may be derived for the estimator (3.8) based on the score. They are direct consequence of the following proposition.

Proposition 3.5 (**Asymptotic equivalence of estimators**). *The estimators $\check{\theta}(\Sigma)$ and $\hat{\theta}(J_0\Sigma J_0)$ are asymptotically equivalent:*

$$\sqrt{T}\Big[\check{\theta}(\Sigma) - \hat{\theta}(J_0\Sigma J_0)\Big] \approx 0.$$

It is straightforward to derive the optimal choice of Σ for estimators based on the score: since $\Omega = J_0\Sigma J_0$, it has to be $\Sigma^* = (I_0 - K_0)^{-1}$.

For what concerns the efficiency of the I.I. estimators, expression (3.11) clearly puts in evidence the components that contribute to the precision. The term in square brackets depends on the auxiliary model adopted and on the estimation method used for the auxiliary parameter: to reduce this term one should obviously estimate β in the most efficient way and should have the Jacobian as close as possible to the identity (i.e. to choose an auxiliary model as close as possible to the econometric model). The component in round brackets summarizes the effect of simulations, which appears in a multiplicative factor, since

$$W(S,\Omega) = \Big(1 + \frac{1}{S}\Big) W(\infty, \Omega).$$

This term can be made arbitrarily close to one, at the cost of a large computational effort.
Note[6] finally that if there were no exogenous variables, the term $(I_0 - K_0)$ would become I_0, and the variance-covariance matrix would be generally greater. In other words, the accuracy of the indirect estimators is improved by the observation of the z_t's (which have not to be simulated).

3.3.4 Estimation of W_S^*

The expressions of the asymptotic variance-covariance matrix of indirect inference estimators contains the derivative of the binding function at the true value of the parameter. This expression cannot directly be computed because explicitating and differentiating the binding function is in general a very difficult task. Luckily, it is possible to consistently estimate this quantity without determining the binding function and its derivative. Indeed, $b(\theta)$ is the solution of:

$$b(\theta) = \arg\max_{\beta} \operatorname{plim} \psi_T(y_t^s(\theta); \beta) = \arg\max_{\beta} \psi_\infty(\theta; \beta)$$

It therefore satisfies the first order conditions:

$$\frac{\partial \psi_\infty}{\partial \beta}[\theta, b(\theta)] = 0, \;\; \forall\, \theta \in \Theta.$$

Deriving this relation with respect to θ gives:

$$\frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}[\theta, b(\theta)] + \frac{\partial^2 \psi_\infty}{\partial \beta \partial \beta'}[\theta, b(\theta)] \frac{\partial b}{\partial \theta'}(\theta) = 0$$

[6] Gouriéroux et al. (1993)[21], Section 3.

Then, in $\theta = \theta_0$,

$$\begin{aligned}\frac{\partial b}{\partial \theta'}(\theta_0) &= \left\{\frac{\partial^2 \psi_\infty}{\partial \beta \partial \beta'}[\theta_0, b(\theta_0)]\right\}^{-1} \frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}[\theta_0, b(\theta_0)] \\ &= J_0^{-1} \frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}[\theta_0, b(\theta_0)]\end{aligned} \tag{3.12}$$

We can thus derive an alternative expression of the asymptotic variance-covariance matrix of the optimal I.I. estimator $\hat{\theta}^*$ which may be directly computed from the criterion function:

$$W_S^* = \left(1 + \frac{1}{S}\right)\left(\frac{\partial^2 \psi_\infty}{\partial \theta \partial \beta'}(I_0 - K_0)^{-1} \frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}\right)^{-1} \tag{3.13}$$

As far as an estimation of W_S^* is concerned, a consistent estimator of ψ_∞ is needed. Such an estimator can be obtained by a numerical derivation of $\frac{\partial \psi_T}{\partial \beta'}[y_t^s(\theta); \hat{\beta}]$ with respect to θ, evaluated at $\hat{\theta}^*$. For the derivation of a consistent estimator of $(I_0 - K_0)$, see Gouriéroux et al. (1993), Appendix 2.

3.4 A simple example

To fix ideas, it might be useful to consider a simple example[7] involving a nonlinear data generating process (dgp) like

$$y_t = \exp\{\mathbf{z_t}'\theta\} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2) \tag{3.14}$$

Let the auxiliary model be

$$y_t = \mathbf{z_t}'\beta + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{3.15}$$

Note that:

$$\frac{\partial \mathbb{E}[y_t|\mathbf{z_t}]}{\partial \mathbf{z_t}} = \beta \quad \text{(under the auxiliary model)}$$

$$\frac{\partial \ln \mathbb{E}[y_t|\mathbf{z_t}]}{\partial \mathbf{z_t}} = \frac{\partial \mathbb{E}[y_t|\mathbf{z_t}]}{\partial \mathbf{z_t}} \cdot \frac{1}{\mathbb{E}[y_t|\mathbf{z_t}]} = \theta \quad \text{(under the model of interest)}$$

One can thus deduce that the binding function is

$$\beta = \theta\, \mathbb{E}[y_t|\mathbf{z_t}], \text{ or}$$

[7] From Cameron & Trivedi (2005)[8], Section 12.6.

$$\theta = \left(\mathbb{E}[y_t|\mathbf{z_t}]\right)^{-1}\beta \tag{3.16}$$

Note that $\dim(\theta) = \dim(\beta)$.

A *naive* estimation of the auxiliary parameter, drawn on the observed data $y_t, \mathbf{z_t}, t = 1, ..., T$, can be easily obtained (e.g. by least squares). Let us denote such an (inconsistent) estimate by $\hat{\beta}$. Now, given a T-dimensional pseudo-random draw, denoted by $\epsilon^{(0)}$, and chosen $\tilde{\theta}^{(0)} = \hat{\beta}$, say, it is easy to generate $y_t^{(1)}$, $t = 1, ..., T$ using

$$y_t^{(1)} = \exp\{\mathbf{z_t}'\tilde{\theta}^{(0)}\} + \epsilon_t^{(0)}$$

and obtain a revised estimator $\hat{\beta}^{(1)} = (\sum_t \mathbf{z_t}\mathbf{z_t}')^{-1}\sum_t \mathbf{z_t}y_t^{(1)}$, which in turn is used to derive $\theta^{(p)}$ from (3.16) and to generate a new set of pseudo-random observation from (3.14). The entire simulation cycle is repeated, holding $\epsilon^{(0)}$ fixed, until $[\hat{\beta} - \hat{\beta}(\tilde{\theta}^{(p)})]'[\hat{\beta} - \hat{\beta}(\tilde{\theta}^{(p)})]$ is minimized, i.e. until the calibration procedure has corrected for the bias of the naive estimator. The resulting estimate of θ is the indirect inference estimate.

Chapter 4

Indirect Estimation of Stable GARCH Processes

Several studies have highlighted the fact that heavy-tailedness of asset returns can be the consequence of conditional heteroskedasticity. ARCH and GARCH models have thus become very popular, given their ability to account for volatility clustering and, implicitly, heavy tails. However, as outlined in chapter 1, these models encounter some difficulties in handling financial time series, as they respond equally to positive and negative shocks; in addition, some empirical studies (for instance, Yang & Brorsen, 1993 [44]) indicate that the tail behavior of GARCH models remains too short even with Student-t error terms[1]. To overcome these weaknesses we apply GARCH models with α-stable innovations[2]. Since simulated values from α-stable distributions can be straightforwardly obtained (see section 2.7), the indirect inference approach (described in chapter 3) is particularly suited to the situation at hand. Here we provide a description of how to implement such a method by using a GARCH with skewed Student's t innovations as auxiliary model. This distribution has four parameters which have a clear and interpretable matching with those of

[1] Furthermore, the Student-t distribution lacks the stability-under-addition property. Stability is desirable because stable distributions, having domains of attraction, provide a very good approximation for large classes of distributions.

[2] GARCH models with symmetric stable innovations have been first proposed by McCulloch (1985)[30].

the α-stable distribution. Among the many proposals of skew-t density functions appeared in the literature, we have adopted the one recently introduced by Azzalini & Capitanio (2003)[2], which is briefly reviewed in the following section[3]. In section 4.2 the models implemented are presented and the simulations results are shown in section 4.3. Finally, the proposed models are used to estimate the IBM weekly returns series, to see how they perform on real data.

4.1 The skew-t distribution

To be better informed about the four stable parameters $(\alpha, \beta, \gamma, \delta)$, it is intuitively to go through a quasi-likelihood function which entails similar parameters with similar interpretations. Therefore, the family of skew-Student's t distributions introduced by Azzalini & Capitanio (2003)[2] seems to be a natural choice.

The idea follows from an extension of the skew-normal distribution (Azzalini, 1985 [1]), in which the symmetry of the density is perturbated by means of the distribution function evaluated at a certain point. More formally, the univariate skew-normal density function is defined as:

$$f(x; \tilde{\beta}, \mu, \sigma) = 2\,\phi(z)\,\Phi(\tilde{\beta} z) \tag{4.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the density and the distribution function of the standard normal distribution and $z = \frac{x-\mu}{\sigma}$. The parameter[4] $\tilde{\beta} \in \mathbb{R}$ plays the role of shape parameter dealing with the degree of skewness; when $\tilde{\beta} = 0$ we recover the regular normal density and we write $\mathrm{SN}(\mu, \sigma, 0) = \mathrm{N}(\mu, \sigma)$. Among the many formal properties shared with the normal class, a noteworthy fact is that if $X \sim \mathrm{SN}(\mu, \sigma, \tilde{\beta})$, then $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1$.

The usual construction of the t distribution is by means of the ratio of a normal variate and an appropriate transformation of a chi-square. Hence, replacing

[3] A widely used alternative, adopted for instance in Garcia et al. (2011)[18] is the version introduced by Fernàndez & Steel in 1998.

[4] In the original paper, $\tilde{\beta}$ is denoted by α; this different notation is adopted here to avoid confusion with the stable distribution's tail parameter.

the normal variate above by a skew one, leads to an asymmetric variant of the t distribution, whose density is given by

$$
\begin{aligned}
f(x;\nu,\tilde{\beta},\sigma,\mu) &= \frac{2}{\sigma}\, f_t(z;\nu)\, F_t\Big(\tilde{\beta} z\sqrt{\frac{\nu+1}{z^2+\nu}};\nu+1\Big) \\
&= \frac{2}{\sigma}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}\Big[1+\frac{z^2}{\nu}\Big]^{-\frac{\nu+1}{2}} F_t\Big(\tilde{\beta} z\sqrt{\frac{\nu+1}{z^2+\nu}};\nu+1\Big)
\end{aligned}
\tag{4.2}
$$

where z is defined as before, $f_t(\cdot)$ and $F_t(\cdot)$ denote density and distribution function of a Student-t variable with ν degrees of freedom. Distribution (4.2) is called skew-t and we write $X \sim \mathrm{St}(\nu, \tilde{\beta}, \sigma, \mu)$. Figure 4.1 shows the pdf of a $\mathrm{SN}(0, 1, 8)$ (left panel) and of a $\mathrm{St}(2, 3.5, 1, 0)$ (right panel).

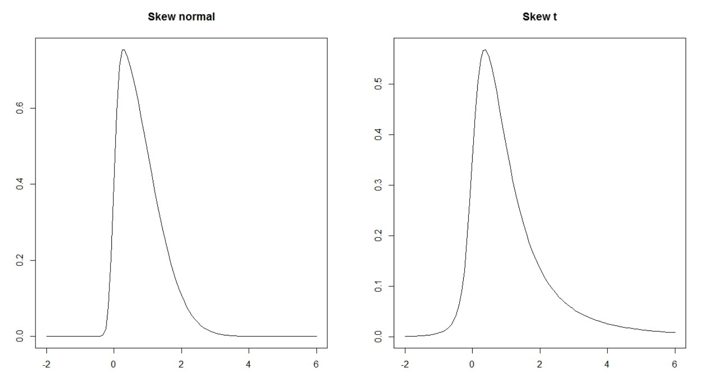


Figure 4.1: Probability density function of a skew-normal with $\mu = 0, \sigma = 1, \tilde{\beta} = 8$ (left) and a skew-t with $\nu = 2, \tilde{\beta} = 3.5, \sigma = 1, \mu = 0$ (right).

The four parameters of the skew-t distribution all have a clear interpretation: $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ model location and dispersion, respectively; the additional parameter $\tilde{\beta} \in \mathbb{R}$ influences the asymmetry; $\nu \in \mathbb{R}^+$ captures the thickness of the tails[5]. In an indirect inference framework, one can thus expect the skew-t auxiliary parameters to be very informative about the stable ones. In fact, Garcia et al. (2011)[18] prove four analytical results that show the correspondence between these two set of auxiliary and structural parameters, as summarized by Table 4.1.

[5]The first four moments of a skew-t distribution with ν degrees of freedom are defined only for ν larger than the corresponding order of the moment. Note the similarity with the moments of α-stable distributions (Property 2.4).

Characteristic	Structural	Auxiliary
Tail thickness	α	ν
Asymmetry	β	$\tilde{\beta}$
Scale	γ	σ
Location	δ	μ

Table 4.1: Relation between structural and auxiliary parameters.

For skew-t-based models, maximum likelihood (ML) is a feasible estimator. The log-likelihood function for a skew-t sample of n observation is:

$$\begin{aligned} \ln \mathcal{L}(\nu, \tilde{\beta}, \sigma, \mu | x) &= n\left[\ln \tfrac{2}{\sigma} + \ln \Gamma(\tfrac{\nu+1}{2}) - \ln \Gamma(\tfrac{\nu}{2}) - \tfrac{1}{2}\ln(\pi\nu)\right] \\ &+ \sum_{i=1}^{n} \ln F_t\Big(\tilde{\beta} z_i \sqrt{\frac{\nu+1}{z_i^2+\nu}}; \nu+1\Big) - \frac{\nu+1}{2}\sum_{i=1}^{n} \ln\big(1+\frac{z_i^2}{\nu}\big) \end{aligned} \tag{4.3}$$

The log-likelihood of the auxiliary models presented in the following section have been computed exploiting the *skew* GAUSS library[6] implemented by Roncalli & Lagache (2004)[37].

4.2 Structural and auxiliary models

The model of interest we wish to estimate to describe the volatility of an asset return is a GARCH(1,1) with α-stable innovations. Let r_t be the return series. It is well-known that dependence in the second moment of the returns' density function is much stronger than dependence in the first moment; thus, we assume r_t to be serially uncorrelated, but not serially independent, namely $r_t = c + \epsilon_t$, where

$$\epsilon_t = z_t \sqrt{\sigma_t^2}, \quad z_t | \mathcal{I}_{t-1} \overset{iid}{\sim} \mathrm{S}_0(\alpha, \beta, 1, 0) \tag{4.4}$$

i.e. z_t, given the information set in $t-1$, is a sequence of iid α-stable error terms with location zero and unit scale parameter. The conditional variance

[6] Downloadable at www.thierry-roncalli.com.

is

$$\sigma_t^2 \equiv V[r_t|\mathcal{I}_{t-1}] = V[\epsilon_t|\mathcal{I}_{t-1}] = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{4.5}$$

and, exploiting the linear transformation property of α-stable distributions (Property 2.9), one can write

$$\epsilon_t|\mathcal{I}_{t-1} \overset{iid}{\sim} \mathrm{S}_0(\alpha, \beta, \sigma_t, 0) \tag{4.6}$$

As a natural auxiliary model one can entertain the skew-t analog of the "true" model, i.e. a GARCH(1,1) model with innovations $z_t^a|\mathcal{I}_{t-1} \overset{iid}{\sim} \mathrm{St}(\nu, \tilde{\beta}, 1, 0)$.

4.2.1 Estimation

Dealing with the indirect estimation of the model, as given in the previous subsection, we have encountered several difficulties. The main problem has arisen from the heavier-than-normal tails that both GARCH models and stable noise capture. In fact, de Vries (1991)[10] shows that the stable and GARCH-like processes are observationally equivalent from the view point of the unconditional distribution. In particular, both models share the fact that the unconditional distribution has fat tails (Ghose and Kroner, 1995 [19]).
From the tail behavior property of stable distributions (Property 2.2), one can see how the asymmetry parameter affects the tails: if the α-stable innovations are skewed (i.e. $\beta \neq 0$), the heavy-taildness increases considerably.
As an illustration of how these considerations affect the indirect estimation of the proposed model, Table 4.2 shows a simulated return series under a GARCH(1,1) dgp with skewed stable innovations. In this example α is set to 1.99 and β to -0.1, but even with a tail parameter "close to the normal" the simulation can rapidly explode; this is mainly due to the high kurtosis of the skewed stable sampled shocks and the GARCH term in (4.5). In fact, from the kurtosis of a (Gaussian) GARCH model (Proposition 1.3), it is easy to verify that $\partial K[\epsilon_t]/\partial\beta_1 > \partial K[\epsilon_t]/\partial\alpha_1 > 0$; in other words, β_1 plays the main role in determining the tail behavior of GARCH models. Therefore, a naive way to decrease the heaviness of the tails could be to constraint β_1 to zero, i.e. to reduce the GARCH(1,1) to an ARCH(1). By means of this solution one can manage to make the procedure converge, but as ARCH models are seldom suitable for empirical applications, simulations results concerning this

t	$r_t = z_t\sqrt{\sigma_t^2}$
1	1.1496
2	0.0334
3	-2.9761
$\vdots$	$\vdots$
50	-1211.8399
51	9707.3895
$\vdots$	$\vdots$
500	6.2405e+028
501	2.54367e+028
502	-7.4799e+029

Table 4.2: Simulation of a return series under model (4.4) with $\omega = 0.1, \alpha_1 = 0.05, \beta_1 = 0.8, \alpha = 1.99, \beta = -0.1, \gamma = 1, \delta = 0$.

model will not be reported.
Alternatively, the idea one can pursue is to constraint the asymmetry parameter of the α-stable distribution to zero. This leads to a GARCH(1,1) model with symmetric (standard) stable shocks, with the tail parameter to be estimated[7]. We thus have four parameters in the model of interest and five in the auxiliary model (over-identified approach); numerical results concerning this model are displayed in section 4.3. However, we remark that, to make sure that the simulated dgps do not explode, some constraints on the parameters are still needed. In particular, we choose to bound α to $1.98 \leq \alpha \leq 2$ [8] and β_1 to $0 \leq \beta_1 \leq 0.82$. In the simulation experiment the true values of β_1 is set to 0.75, which is low if compared with most empirical findings. Yet, as noted in de Vries (1991), a justification for the relative low β_1 one can find in stable GARCH-type models may be that stable models are intrinsically “robust” against outliers.
Although this approach performs quite well, this structural model is still

[7]Note that also in Panorska et al. (1995)[35], to show the existence and uniqueness of strictly stationary solutions for a stable GARCH process, β is restricted to zero.

[8]To give an idea of the tail behavior of α-stable densities, simulating 10000 samples of size 2000 with $\alpha = 1.98$, one obtains, on average, a kurtosis about equal to 22.

not able to capture the leverage effect since we have constrained the innovations. One can thus think to replace the GARCH(1,1) structural model with a TGARCH(1,1), using the same skew-t GARCH(1,1) as auxiliary model. This enable one to employ a just-identified procedure, with a true model which is more flexible and suitable to financial applications. Simulation results, presented in the following section, seem to be promising.

4.3 Simulation results

The results have been obtained by means of the Efficient Method of Moments (EMM) of Gallant & Tauchen (1996)[16] outlined in section 1.3, based on a numerical computation of the auxiliary model's score.
The first simulations carried out concern the estimation of the two models presented before with random samples of two different size, namely $T = 1000$ and $T = 3000$; the experiments are based on a set of $R = 500$ replications with $S = 2$ simulations.
Recall that in both the GARCH(1,1) and the TGARCH(1,1), whose results are shown in Table 4.3 and 4.4 respectively, the shocks are held symmetric ($\beta = 0$); also, the auxiliary model is a skew-t GARCH(1,1) for both the true models, so that in the former case the quadratic form (3.8) has been weighted by means of Σ^* (that is the inverse of the auxiliary model's outer product matrix), in the latter case the procedure is exactly identified and no weighting matrix is needed. The quadratic forms have been minimized by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) *quasi-Newton* algorithm, while for the estimation of the auxiliary model the *cml* GAUSS routine ("constrained maximum likelihood") has been used. In fact, as noted by Garcia, et al. (2011)[18], the relation between α and ν is exponential, in the sense that as $\alpha \to 2$ we get closer to the Gaussian distribution and therefore $\hat{\nu} \to \infty$. Having set the true value of α to 1.985, we need to constraint $\hat{\nu}$ as $\hat{\nu} \leq 120$, say; indeed, without bounds, if a random sample with very thin tails is drawn, the estimate of ν is attracted towards $+\infty$ giving rise to problems in the EMM step[9].

[9] In Fiorentini et al. (2003)[13] the reciprocal of the Student's t degrees of freedom is used as parameter; thus its constraints become more manageable.

Table 4.3: Monte Carlo means and standard errors (in parentheses) for the parameters of the true model GARCH(1,1) with shocks $z_t \overset{iid}{\sim} \mathrm{S}_k(1.985, 0, 1, 0)$, for $T = 1000$ and $T = 3000$ $(S = 2, R = 500)$.

	GARCH(1,1) $\beta = 0$			
	$\omega = 0.1$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\alpha = 1.985$
$T = 1000$	0.0848	0.0686	0.7327	1.9894
	(0.0430)	(0.0228)	(0.0712)	(0.0065)
$T = 3000$	0.0804	0.0615	0.7595	1.9898
	(0.0283)	(0.0.0145)	(0.0336)	(0.0054)

Table 4.4: Monte Carlo means and standard errors (in parentheses) for the parameters of the true model TGARCH(1,1) with shocks $z_t \overset{iid}{\sim} \mathrm{S}_k(1.985, 0, 1, 0)$, for $T = 1000$ and $T = 3000$ $(S = 2, R = 500)$.

	TGARCH(1,1) $\beta = 0$				
	$\omega = 0.02$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\gamma_1 = 0.05$	$\alpha = 1.985$
$T = 1000$	0.0285	0.0454	0.7040	0.0549	1.9899
	(0.0199)	(0.0237)	(0.0994)	(0.0247)	(0.0058)
$T = 3000$	0.0206	0.0559	0.7458	0.0460	1.9895
	(0.0122)	(0.0185)	(0.0466)	(0.0198)	(0.0050)

For these first simulation experiments the starting values supplied to the optimization routine have been set to the true values (the effect of the choice of $\tilde{\theta}^{(0)}$ will be examined in what follows). The results suggest that the estimators converge asymptotically to the true values.

The next simulation studies have been conducted to explore the effect of the number of simulations S on the performance of the estimators. These experiments are based on a set of 100 replications[10] with $T = 2000$. As one could expect, increasing the number of simulations the estimated standard errors

[10] Unfortunately it has not been possible to employ a higher number of replications due to the computational slowness and the short time available.

decrease considerably. The Monte Carlo means do not seem to improve, but this could be due to the low number of replications utilized.

Table 4.5: Monte Carlo means and standard errors for the parameters of the true model GARCH(1,1) for various S ($T = 2000, R = 100$).

	GARCH(1,1) $\beta = 0$			
	$\omega = 0.1$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\alpha = 1.985$
$S = 1$	0.0821	0.0642	0.7471	1.9914
	(0.0479)	(0.0244)	(0.0654)	(0.0065)
$S = 2$	0.0771	0.0681	0.7503	1.9911
	(0.0470)	(0.0198)	(0.0510)	(0.0063)
$S = 5$	0.0672	0.0676	0.7673	1.9897
	(0.0324)	(0.0170)	(0.0302)	(0.0061)

Table 4.6: Monte Carlo means and standard errors for the parameters of the true model TGARCH(1,1) for various S ($T = 2000, R = 100$).

	TGARCH(1,1) $\beta = 0$				
	$\omega = 0.02$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\gamma_1 = 0.05$	$\alpha = 1.985$
$S = 1$	0.0292	0.0456	0.7067	0.0574	1.9899
	(0.0202)	(0.0257)	(0.0867)	(0.0264)	(0.0061)
$S = 2$	0.0225	0.0463	0.7331	0.0542	1.9912
	(0.0144)	(0.0236)	(0.0631)	(0.0236)	(0.0057)
$S = 5$	0.0236	0.0497	0.7327	0.0524	1.9889
	(0.0143)	(0.0206)	(0.0580)	(0.0231)	(0.0051)

The last simulation experiments aim at assessing how the starting values supplied to the optimization algorithm affect the estimates. It is known, although seldom discussed in literature[11], that estimation in a GARCH framework often proved troublesome and highly sensitive to initial values. Furthermore, in

[11] A good reference on this issue is Belsley & Kontoghiorghes (2009)[3], Section 2.2.

finite samples, different initializations for $\hat{\epsilon}_0$ and $\hat{\sigma}_0$, providing different conditional likelihood, lead naturally to different parameter estimates[12]. Readers can glance at Appendix B to see how GARCH dgps have been simulated. The starting values $\tilde{\theta}^{(0)}$ have thus been set to a value "not too far" from the true one. Results, displayed in Tab 4.7 and 4.8, suggest that as long as one manages to achieve the convergence, the initial guesses do not affect the consistency of the estimators. This finding, however, deserves a more detailed simulation study.

Table 4.7: Monte Carlo means and standard errors for the parameters of the true model GARCH(1,1) for starting values different from the true ones ($T = 3000, S = 2, R = 100$).

	GARCH(1,1) $\beta = 0$			
	$\omega = 0.1$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\alpha = 1.985$
$\tilde{\theta}^{(0)} \neq \theta_0$	0.0815	0.0673	0.7502	1.9875
	(0.0252)	(0.0141)	(0.0320)	(0.0049)

Table 4.8: Monte Carlo means and standard errors for the parameters of the true model TGARCH(1,1) for starting values different from the true ones ($T = 3000, S = 2, R = 100$).

	TGARCH(1,1) $\beta = 0$				
	$\omega = 0.02$	$\alpha_1 = 0.05$	$\beta_1 = 0.75$	$\gamma_1 = 0.05$	$\alpha = 1.985$
$\tilde{\theta}^{(0)} \neq \theta_0$	0.0184	0.0583	0.7507	0.0456	1.9885
	(0.0110)	(0.0175)	(0.0437)	(0.0188)	(0.0049)

Finally, to give an idea of the estimation of the auxiliary model's parameters,

[12]Inspite of the GARCH benchmark developed by Fiorentini, Calzolari & Panattoni in 1996, McCullough and Renfro (1998)[31] analyze seven widely used econometric packages and four of which are found to provide different answers to the same non linear estimation problem; in most cases this is because the developer does not indicate which conditional likelihood is being maximized.

some results from the PML step are reported in Table 4.9. The first line shows the parameter estimates of a GARCH(1,1) with skew-t innovations, when the data are generated under a stable GARCH(1,1). In the second line, the true data generating process is a stable TGARCH(1,1). The true parameters are set to the usual values, namely $\omega = 0.1, \alpha_1 = 0.05, \beta_1 = 0.75$ for the GARCH, and $\omega = 0.02, \alpha_1 = 0.05, \beta_1 = 0.75, \gamma_1 = 0.05$ for the TGARCH, with $z_t \overset{iid}{\sim} \mathrm{S}_k(1.985, 0)$. Here, the sample size is $T = 2000$ with $R = 500$ Monte Carlo replications.

Table 4.9: Monte Carlo means and standard errors for the parameters of the auxiliary model when the dgps is a stable GARCH(1,1) and a stable TGARCH(1,1) ($T = 2000, R = 500$).

	Skew-t GARCH(1,1)				
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\hat{\beta}}$	$\hat{\nu}$
Stable-GARCH dgp	0.2064	0.0959	0.7335	-0.0004	47.3543
	(0.086)	(0.0219)	(0.0787)	(0.0283)	(39.631)
Stable-TGARCH dgp	0.0876	0.1006	0.7358	-0.0017	39.0235
	(0.0272)	(0.0220)	(0.0602)	(0.02797)	(35.295)

4.4 An empirical application

In this section we apply the two models described in the previous section on real data. The set of data considered is the historical series of the IBM stock's weekly prices in Figure 4.2; this data set is composed of 1973 observations, from 2nd January 1973 to 25th October 2010[13]. Figure 4.3 presents the returns computed as $\ln(P_t/P_{t-1}) \times 100$ (left panel), and the squared returns (right panel).

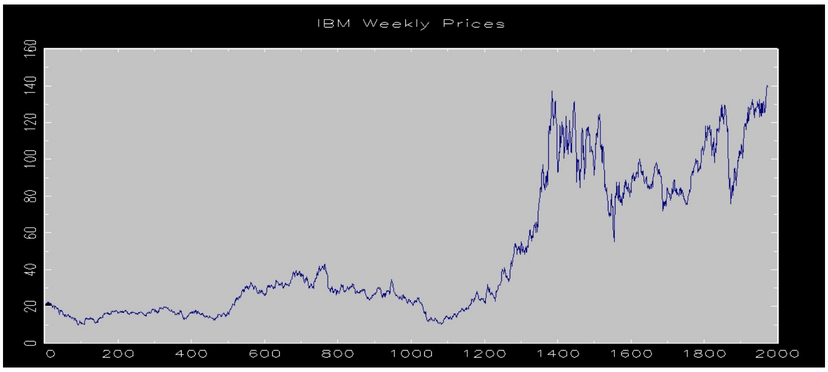


Figure 4.2: IBM weekly prices, 2 January 1973 - 25 October 2010.

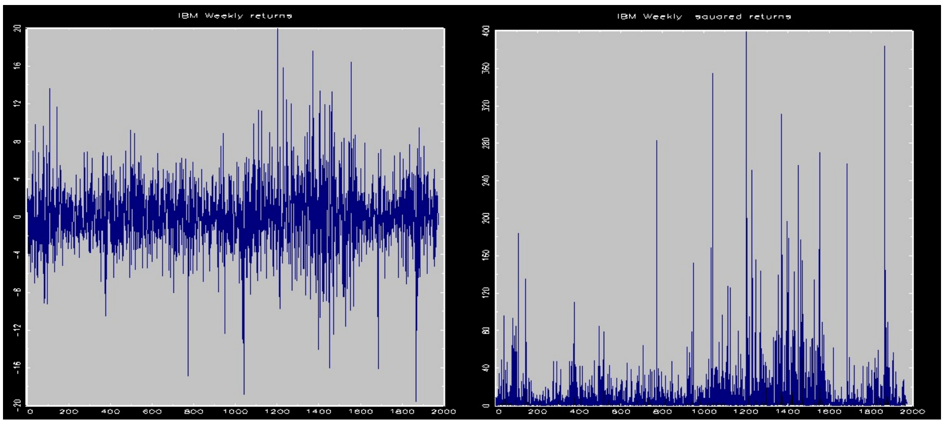


Figure 4.3: IBM weekly returns (left) and squared returns (right).

[13]Free stock quotes are available, for instance, at finance.yahoo.com.

Summary statistics of the series outlined in Table 4.10 depict a clear departure from normality, albeit the skewness is nearly zero. While Ljung-Box test on returns suggests that no ARMA term is needed, the squared returns series shows at least a lag-1 autocorrelation. However, if one estimate a classic Gaussian GARCH(1,1), (s)he finds standardized residuals far from normality (see Figure 4.4).

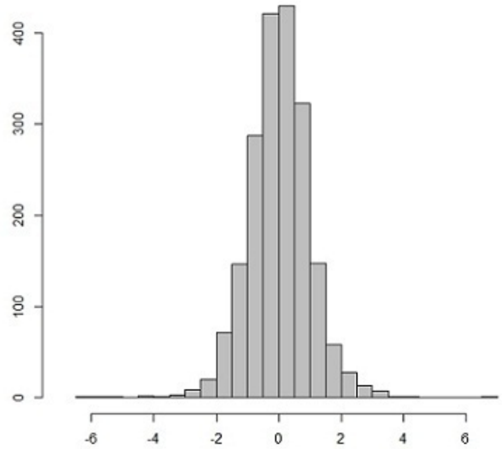
Estimation results for both the Gaussian and the symmetric stable GARCH(1,1) models are reported in Table 4.11. As one could expect, the persistence of the stable model is smaller than the Gaussian one. This is a direct consequence of the tail parameter of the stable distribution, which plays an important role in capturing the heavy tailed features of the noise.

Table 4.10: Descriptive analysis of IBM weekly returns data set.

Mean	0.0972
Minimum	-19.5759
Maximum	19.9765
Std. dev.	3.5808
Skewness	-0.0010
Kurtosis excess	3.2321
JB (p-value)	0.0000
LB(1) r_t (p-value)	0.1759
LB(1) r_t^2 (p-value)	0.0000

Table 4.11: Estimates and standard errors of a Gaussian and a symmetric stable GARCH(1,1) model for the IBM weekly returns.

	Stable				**Gaussian**		
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$
Mean	0.1420	0.0203	0.9183	1.9904	0.2567	0.0600	0.9213
Sd	0.0424	0.0122	0.0150	0.0060	0.0730	0.0109	0.0134



Observations	1973
Mean	0.0197
Minimum	-6.089
Maximum	6.54
Median	0.0363
Std. Dev.	0.9985
Skewness	-0.0871
Kurtosis (excess)	3.139
Jarque-Bera (p-value)	0.0000

Figure 4.4: Estimated standardized residuals of a Gaussian GARCH(1,1) model for the IBM return series.

Finally we try to estimate the TGARCH(1,1) with symmetric α-stable shocks. Results are displayed in Table 4.12. The decrease in the ARCH term is due to the asymmetric effect of negative and positive innovations. In fact, the estimated impact of the negative shocks is $\hat{\alpha}+\hat{\gamma} = 0.0779$ while the impact of positive shocks is estimated to be $\hat{\alpha} = 0.0182$.

	Stable TGARCH				
	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\alpha}$
Mean	0.1385	0.0182	0.9183	0.0597	1.9934
Sd	0.0488	0.0124	0.0167	0.0245	0.0041

Table 4.12: Estimates and standard errors of a symmetric stable TGARCH(1,1) model for the IBM weekly returns.

4.5 Conclusions

The goal of this work was to estimate GARCH-type models with α-stable shocks by means of the indirect inference approach. The main motivation of such an effort was to introduce a time-varying volatility model capable to account for the excess of kurtosis and the skewness typical of financial data sets, features not captured by traditional GARCH models.

The characteristics and the analytical properties of α-stable distributions are especially appreciated in the financial field: the possibility of accommodating for asymmetry and heavy-tails allows appropriate risk measurement, and the presence of a central limit theorem constitutes a theoretical basis which leads to preferring the α-stable family over other heavy-tailed alternatives. Since α-stable models do not have a closed-form likelihood function, but it is easy to simulate α-stable pseudo-random numbers, indirect inference constitutes the ideal estimation method for this framework. Indeed, such an inferencial approach is particularly suited to situations where the model of interest is too difficult to estimate, but relatively easy to simulate. One can thus replace the model of interest with an auxiliary one, estimate its parameters using either the observed and the simulated data, and then calibrate the parameters of the original model minimizing the distance between these two sets of estimates.

As auxiliary model we used a GARCH(1,1) with skew-t distributed innovations. The chosen model of interest was, at first, a stable GARCH(1,1) (generalizable to a GARCH(p, q) by including additional lags). Unfortunately, the heavier-than-normal tails that both GARCH and stable models capture forced us to constraint some parameters in order to avoid the explosion of the series simulated under the true model. In particular, one needs to constraint the α-stable shocks to be symmetric, since the asymmetry parameter of α-stable distributions affects considerably the tail-thickness. These findings also indicate that the conditions needed for stationarity of stable GARCH models are stricter than the Gaussian GARCH ones[14].

However, the Monte Carlo studies employed suggest that the models we have

[14]Nelson (1990)[32] shows that the conditions for stationarity in a GARCH model are stricter when the error term follows a Cauchy distribution ($\alpha = 1$) than when the error follows a normal distribution.

entertained have good performances both with artificial and real data.
The simulation studies have been carried out using the econometric programming language GAUSS. This sofware provided us with some advantages: for instance, the *skew* library permits to readily compute the pseudo-likelihood function of the auxiliary model. Furthermore, efficient optimization routines are already written and allow one to incorporate the inequality restrictions on the parameters. Unfortunately, there is a trade-off between these conveniences and the computational time. For this reason, further researches will consist in more extensive simulation experiments, to see wheter - using a more basic programming language such as Fortran or C/C++ - the promising results will be confirmed.

Appendix A

Derivation of the asymptotic results

The goal of this appendix is to hint why indirect inference estimators are consistent, to get the form of their asymptotic variance-covariance matrices and their asymptotic expansion, in order to study their asymptotic equivalence. For more precise proofs, see Gouriéroux et al. (1993)[21].

Identification

(i) In the just identified case, $\hat{\theta}(\Omega)$ is the solution of the system

$$\hat{\beta} = \hat{\beta}(\theta)$$

since, for such a choice, the criterion function is equal to 0. It follows that $\hat{\theta}(\Omega)$ does not depend of Ω.

(ii) Similarly, $\check{\theta}(\Sigma)$ is the solution of the system

$$\sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta); \hat{\beta}] = 0$$

and it is independent of Σ.

(iii) If

$$\sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\check{\theta}); \beta] = 0$$

has a unique solution, this solution has to be $\hat{\beta}(\check{\theta})$; from (ii) one can deduce that this solution is equal to $\hat{\beta}$, from (i) we know that $\hat{\beta} = \hat{\beta}(\hat{\theta})$, therefore $\check{\theta} = \hat{\theta}$.

□

Consistency

Let us first consider the two intermediate estimators $\hat{\beta}$ and $\hat{\beta}(\theta)$. We have:

$$\hat{\beta} = \arg\max_{\beta} \psi_T(y_t; \beta) \rightarrow \arg\max_{\beta} \psi_\infty(\theta_0; \beta) = b(\theta_0)$$

$$\hat{\beta}(\theta) = \arg\max_{\beta} \sum_{s=1}^{S} \psi_T(y_t^s(\theta); \beta) \rightarrow \arg\max_{\beta} S\psi_\infty(\theta; \beta) = b(\theta)$$

Therefore $\hat{\beta}$ converges to $b(\theta_0)$ and $\hat{\beta}(\cdot)$ converges to the binding function $b(\cdot)$. Then:

$$\begin{aligned}
\hat{\theta}(\Omega) &= \arg\min_{\theta} [\hat{\beta} - \hat{\beta}(\theta)]' \, \Omega \, [\hat{\beta} - \hat{\beta}(\theta)] \\
&\rightarrow \arg\min_{\theta} [b(\theta_0) - b(\theta)]' \, \Omega \, [b(\theta_0) - b(\theta)] \\
&= \{\theta : b(\theta) = b(\theta_0)\} \quad \text{(being } \Omega \text{ positive definite)} \\
&= \theta_0 \quad \text{(from (A4))}.
\end{aligned}$$

□

Asymptotic normality of $\hat{\theta}(\Omega)$

Asymptotic expansion of $\hat{\beta}$ and $\hat{\beta}(\theta)$

From the first order conditions for $\hat{\beta}(\theta)$ we have:

$$\sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta} [y_t^s(\theta); \hat{\beta}(\theta)] = 0$$

By means of Taylor's formula we get, around $\theta = \theta_0$:

$$\sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta} [y_t^s(\theta_0); b(\theta_0)] + \sum_{s=1}^{S} \frac{\partial^2 \psi_T}{\partial \beta \partial \beta'} [y_t^s(\theta_0); b(\theta_0)] \sqrt{T} \left[\hat{\beta}(\theta_0) - b(\theta_0)\right] \approx 0$$

$$\sqrt{T}[\hat{\beta}(\theta_0) - b(\theta_0)] \approx \Big\{ - \sum_{s=1}^{S} \frac{\partial^2 \psi_T}{\partial \beta \partial \beta'}[y_t^s(\theta_0); b(\theta_0)] \Big\}^{-1} \times \sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0); b(\theta_0)]$$

$$\approx \frac{1}{S}\Big\{ - \frac{\partial^2 \psi_\infty}{\partial \beta \partial \beta'}[\theta_0; b(\theta_0)] \Big\}^{-1} \sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0); b(\theta_0)],$$

$$\sqrt{T}\left[\hat{\beta}(\theta_0) - b(\theta_0)\right] \approx \frac{J_0^{-1}}{S} \sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0); b(\theta_0)] \tag{A.1}$$

Analogously, we get

$$\sqrt{T}\left[\hat{\beta} - b(\theta_0)\right] \approx J_0^{-1} \sqrt{T} \sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t; b(\theta_0)] \tag{A.2}$$

Asymptotic expansion of $\hat{\theta}(\Omega)$

The first order conditions for $\hat{\theta}(\Omega)$ is $\frac{\partial}{\partial \theta}[\hat{\beta} - \hat{\beta}(\hat{\theta})]\,\Omega\,[\hat{\beta} - \hat{\beta}(\hat{\theta})] = 0$, which can be equivalently written as

$$\frac{\partial \hat{\beta}}{\partial \theta}(\hat{\theta})\,\Omega\left[\hat{\beta} - \hat{\beta}(\hat{\theta})\right] = 0$$

An expansion around the limit value θ_0 gives:

$$\frac{\partial \hat{\beta}}{\partial \theta}(\theta_0)\,\Omega\,\sqrt{T}[\hat{\beta} - \hat{\beta}(\theta_0)] - \frac{\partial \hat{\beta}'}{\partial \theta}(\theta_0)\,\Omega\,\frac{\partial \hat{\beta}}{\partial \theta'}(\theta_0)\sqrt{T}\left[\hat{\theta}(\Omega) - \theta_0\right] \approx 0$$

$$\sqrt{T}\left[\hat{\theta}(\Omega) - \theta_0\right] \approx \left[\frac{\partial b'}{\partial \theta}(\theta_0)\,\Omega\,\frac{\partial b}{\partial \theta'}(\theta_0)\right]^{-1} \frac{\partial b'}{\partial \theta}(\theta_0)\,\Omega\,\sqrt{T}\left[\hat{\beta} - \hat{\beta}(\theta_0)\right] \tag{A.3}$$

Asymptotic distribution of $\sqrt{T}(\hat{\beta} - \hat{\beta}(\theta))$

Using (A.1) and (A.2), one can obtain:

$$\sqrt{T}\left[\hat{\beta} - \hat{\beta}(\theta_0)\right] = J_0^{-1}\sqrt{T}\Big\{\frac{\partial \psi_T}{\partial \beta}[y_t; b(\theta_0)] - \frac{1}{S}\sum_{s=1}^{S} \frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0); b(\theta_0)]\Big\}$$

Therefore under assumption (A7), (A8), this difference has an asymptotic zero mean normal distribution, with variance-covariance matrix given by:

$$
\begin{aligned}
&\mathrm{V}_{as}\big\{\sqrt{T}[\hat{\beta}-\hat{\beta}(\theta_0)]\big\} \\
&= J_0^{-1}\mathrm{V}_{as}\Big\{\sqrt{T}\big\{\frac{\partial \psi_T}{\partial \beta}[y_t;b(\theta_0)] - \frac{1}{S}\sum_{s=1}^{S}\frac{\partial \psi_T}{\partial \beta}[y_t^s(\theta_0);b(\theta_0)]\big\}\Big\}J_0^{-1}
\end{aligned}
$$

and, after some algebra, we get

$$
\mathrm{V}_{as}\big\{\sqrt{T}[\hat{\beta}-\hat{\beta}(\theta_0)]\big\} = \Big(1+\frac{1}{S}\Big)J_0^{-1}(I_0-K_0)J_0^{-1}
$$

Finally, one can use (A.3) to write

$$
\sqrt{T}\left[\hat{\theta}(\Omega)-\theta_0\right] \xrightarrow[T\to\infty]{d} \mathrm{N}\big(0, W(S,\Omega)\big)
$$

where $W(S,\Omega)$, given by Proposition 3.3, is easily obtained applying the "δ-method". $\square$

The optimality of the matrix $\Omega = \Omega^*$ (Proposition 3.4), is a consequence of Gauss-Markov theorem.

Asymptotic equivalence of the two estimators

Asymptotic expansion of $\check{\theta}(\Sigma)$

The optimization problem defining $\check{\theta}(\Sigma)$ implies that

$$
\begin{aligned}
&\frac{\partial}{\partial \theta}\Big\{\sum_{s=1}^{S}\frac{\partial \psi_T}{\partial \beta'}[(y_t^s(\check{\theta});\hat{\beta}]\Big\}\,\Sigma\,\Big\{\sum_{s=1}^{S}\frac{\partial \psi_T}{\partial \beta}[(y_t^s(\check{\theta});\hat{\beta}]\Big\} \\
&= \Big\{\sum_{s=1}^{S}\frac{\partial^2 \psi_T}{\partial \theta \partial \beta'}[(y_t^s(\check{\theta});\hat{\beta}]\Big\}\,\Sigma\,\Big\{\sum_{s=1}^{S}\frac{\partial \psi_T}{\partial \beta}[(y_t^s(\check{\theta});\hat{\beta}]\Big\} = 0
\end{aligned}
$$

An expansion around the values $\big(\theta_0, b(\theta_0)\big)$ of the parameters provides:

$$
\begin{aligned}
&\Big\{\sum_{s=1}^{S}\frac{\partial^2 \psi_T}{\partial \theta \partial \beta'}[(y_t^s(\theta_0);b(\theta_0)]\Big\}\,\Sigma\,\Big\{\sqrt{T}\sum_{s=1}^{S}\frac{\partial \psi_T}{\partial \beta}[(y_t^s(\theta_0);b(\theta_0)] \\
&+ S\frac{\partial^2 \psi_\infty}{\partial \beta \partial \beta'}[\theta_0, b(\theta_0)]\sqrt{T}\big[\hat{\beta}-b(\theta_0)\big] + S\frac{\partial^2 \psi_\infty}{\partial \beta \partial \theta'}[\theta_0, b(\theta_0)]\sqrt{T}\big[\check{\theta}(\Sigma)-\theta_0\big]\Big\} \approx 0
\end{aligned}
$$

Now, using (A.1) we get asymptotically

$$S\frac{\partial^2\psi_\infty}{\partial\theta\partial\beta'}[\theta_0, b(\theta_0)]\,\Sigma\left\{SJ_0\sqrt{T}\big[\hat{\beta}(\theta_0) - b(\theta_0)\big]\right.$$
$$\left. + S\frac{\partial^2\psi_\infty}{\partial\beta\partial\beta'}[\theta_0, b(\theta_0)]\sqrt{T}\big[\hat{\beta} - b(\theta_0)\big] + S\frac{\partial^2\psi_\infty}{\partial\beta\partial\theta'}[\theta_0, b(\theta_0)]\sqrt{T}\big[\check{\theta}(\Sigma) - \theta_0\big]\right\} \approx 0$$

and, remembering that $\frac{\partial^2\psi_\infty}{\partial\beta\partial\beta'}[\theta_0; b(\theta_0)] = -J_0$,

$$\frac{\partial^2\psi_\infty}{\partial\theta\partial\beta'}[\theta_0, b(\theta_0)]\,\Sigma\left\{J_0\sqrt{T}\big[\hat{\beta}(\theta_0) - \hat{\beta}\big] + \frac{\partial^2\psi_\infty}{\partial\beta\partial\theta'}[\theta_0, b(\theta_0)]\sqrt{T}\big[\check{\theta}(\Sigma) - \theta_0\big]\right\} \approx 0$$

Finally, we obtain:

$$\sqrt{T}\big[\check{\theta}(\Sigma) - \theta_0\big] \approx \left\{\frac{\partial^2\psi_\infty}{\partial\theta\partial\beta'}[\theta_0, b(\theta_0)]\,\Sigma\,\frac{\partial^2\psi_\infty}{\partial\beta\partial\theta'}[\theta_0, b(\theta_0)]\right\}^{-1}$$
$$\times\,\frac{\partial^2\psi_\infty}{\partial\theta\partial\beta'}[\theta_0, b(\theta_0)]\,\Sigma\,J_0\sqrt{T}\,[\hat{\beta} - \hat{\beta}(\theta_0)]\Big\} \tag{A.4}$$

From (3.12) we know that:

$$\frac{\partial b}{\partial\theta'}(\theta_0) = J_0^{-1}\frac{\partial^2\psi_\infty}{\partial\beta\partial\theta'}[\theta_0, b(\theta_0)]$$

Therefore the asymptotic expansion of $\check{\theta}(\Sigma)$ given in (A.4) may also be written as

$$\sqrt{T}\big[\check{\theta}(\Sigma) - \theta_0\big]$$
$$\approx \Big[\frac{\partial b'}{\partial\theta}(\theta_0)J_0\,\Sigma\,J_0\frac{\partial b}{\partial\theta'}(\theta_0)\Big]^{-1}\frac{\partial b'}{\partial\theta}(\theta_0)J_0\,\Sigma\,J_0\sqrt{T}\big[\hat{\beta} - \hat{\beta}(\theta_0)\big]$$

A comparison with expansion (A.3) directly gives Proposition 3.5:

$$\sqrt{T}\big[\check{\theta}(\Sigma) - \hat{\theta}(J_0\,\Sigma\,J_0)\big] \approx 0.$$

□

Appendix B

GAUSS code for the simulation of a Stable GARCH(1,1)

```
/* Simulation of a GARCH(1,1) with standard stable shocks */

proc sim_StableGARCH(_theta,_T,_S);
 local _w,_alpha1,_beta1,_a,_b,_z,_sigma2,_eps,_y;
 _w = _theta[1];
 _alpha1 = _theta[2];
 _beta1 = _theta[3];
 _a = _theta[4];
 _b = _theta[5];
 _sigma2 = ones(_T+1,_S)*_w/(1-_alpha1-_beta1);
 _z = zeros(_T+1,_S);
 _eps = zeros(_T+1,_S);
 for j (1,_S,1);
    _z[2:(_T+1),j] = rstab(_a|_b, _T);
 endfor;
 for j (1,_S,1);
    for i (2,_T+1,1);
       _sigma2[i,j] = _w +_alpha1*_eps[i-1,j]^2 +
                      _beta1*_sigma2[i-1,j];
       _eps[i,j] = _z[i,j]*sqrt(_sigma2[i,j]);
    endfor;
 endfor;
 _y = _z[2:(_T+1),.].*sqrt(_sigma2[2:(_T+1),.]);
```

```
  retp(_y);
  endp;

/* Pseudo-random generation from a S(_a,_b,1,0) */

 proc rstab(_thetaStable,_n);
  local _a,_b,_z,_csi,_U,_W;
  _a = _thetaStable[1];
  _b = _thetaStable[2];
  _z = zeros(_n,1);
  _csi = atan(_b*tan(pi*_a/2))/_a;
  for i (1,_n,1);
     _U = pi*(rndu(1,1)-0.5);
     _W = -ln(rndu(1,1));
     if (_b == 0);
        if (_a == 1);
           _z[i] = tan(_U);
        else;
           _z[i] = (sin(_a*_U)/((cos(_U))^(1/_a)))*
                   ((cos((_a-1)*_U)/_W)^((1-_a)/_a));
        endif;
     else;
        if(_a == 1);
           _z[i] = (2/pi)*((pi/2+_b*_U)*tan(_U)-_b*
                   ln((2/pi)*_W*cos(_U)/(pi/2+_b*_U)));
        else;
           _z[i] = ((sin(_a)*(_csi+_U))/(cos(_a*_csi)*cos(_U))^(1/_a))*
                   ((cos(_a*_csi+_a*_U-_U)/_W)^((1-_a)/_a));
        endif;
     endif;
  endfor;
  retp(_z);
 endp;
```

Bibliography

[1] Azzalini, A. (1985), "A Class of Distributions which Includes the Normal One", *Scandinavian Journal of Statistics* **12**, 171–178.

[2] Azzalini, A. & Capitanio, A. (2003), "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution", *Journal of the Royal Statistical Society B* **65**, 367–389.

[3] Belsley, D. A. & Kontoghiorghes, E. J. (2009), *Handbook of Computational Econometrics*, John Wiley & Sons, Chichester, England.

[4] Beltratti, A. & Liera, M. (2009), *Capire la Borsa*, Il Sole 24 ore.

[5] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* **31**, 307–327.

[6] Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *The Review of Economics and Statistics*, **69(3)**, 542-547.

[7] Calzolari, G., Di Iorio, F. & Fiorentini, G. (1999), "Indirect Estimation of Just-Identified Models with Control Variates", *Quaderni del Dipartimento di Statistica "G.Parenti"* 46.

[8] Cameron, C. A. & Trivedi, P. K. (2005), *Microeconometrics – Methods and Applications*, Cambridge University Press, New York.

[9] Chambers, J., Mallows, C. & Stuck, B. (1976), "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association* **71**, 340–344.

[10] de Vries, C. G. (1991), "On the Relation between GARCH and Stable Processes", *Journal of Econometrics* **48(3)**, 313-324.

[11] Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica* **50**, 987–1008.

[12] Engle, R. (2001), "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics", *Journal of Economic Perspectives* **15**, 157–168.

[13] Fiorentini, G., Sentana, E. & Calzolari, G. (2003), "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student-t Innovations", *Journal of Business and Economic Statistics* **21(4)**, 532-546.

[14] Feller, W. (1966), *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, New York.

[15] Francq, C. & Zakoian, J. M. (2010), *GARCH Models - Structure, Statistical Inference and Financial Applications*, Wiley & Sons, Chichester, England.

[16] Gallant, A. & Tauchen, G. (1996), "Which Moments to match?", *Econometric Theory* **12**, 657-681.

[17] Gallo, G. & Pacini, B. (2002), *Metodi Quantitativi per i Mercati Finanziari*, Carocci Università.

[18] Garcia, R., Renault, E. & Veredas, D. (2011), "Estimation of Stable Distributions by Indirect Inference", *Journal of Econometrics* **161(2)**, 325-337.

[19] Ghose, D. & Kroner, K. F. (1995), "The Relationship between GARCH and Symmetric Stable Processes: Finding the Source of Fat Tails in Financial Data", *Journal of Empirical Finance* **2**, 225-251.

[20] Gouriéroux, C. & Monfort, A. (1996), *Simulation Based Econometric Methods*, Oxford University Press, New York.

[21] Gouriéroux, C., Monfort, A. & Renault, E. (1993), "Indirect Inference", *Journal of Applied Econometrics* **8**, S85-S118.

[22] Halbleib (Chiriac), R. & Voev, V. (2011) "Modelling and forecasting multivariate realized volatility", *Journal of Applied Econometrics*, **26(6)**, 922-947.

[23] Halbleib (Chiriac), R., Nolte, I., Pohlmeier, W. & Voev, V. (2010), *Financial Econometrics*, Department of Economics, University of Konstanz, Konstanz, Germany.

[24] Lambert, P. & Laurent, S., (2002), "Modeling Skewness Dynamics in Series of Financial Data Using Skewed Location-scale Distributions", Working Paper, Université Catholique de Louvain and Universite de Liege, Belgium.

[25] Liu, S. M. & Brorsen, B. W. (1995), "Maximum Likelihood Estimation of a GARCH-Stable Model", *Journal of Applied Econometrics* **10(3)**, 273-285.

[26] Lombardi, M. J. & Calzolari, G. (2008), "Indirect Estimation of α-stable Distributions and Processes", *Econometrics Journal* **11(1)**, 193-208.

[27] Lombardi, M. J. & Calzolari, G. (2009), "Indirect Estimation of α-stable Stochastic Volatility Models", *Computational Statistics and Data Analysis* **53(6)**, 2298-2308.

[28] Maddala, G. S. (2001), *Introduction to Econometrics*, Third Edition, Wiley, Chichester, England.

[29] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business* **36**, 394-419.

[30] McCulloch, J. H. (1985), "Interest-risk Sensitive Deposit Insurance Premia: Stable ACH Estimates", *Journal of Banking and Finance* **9**, 137–156.

[31] McCullough, B. D. & Renfro C. G. (1998), "Benchmarks and Software Standards: A Case Study of GARCH Procedures", *Journal of Economic and Social Measurement* **25**, 59–71.

[32] Nelson, D. B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model", *Econometric Theory* **6**, 318-334.

[33] Nolan, J. P. (2003), "Modeling Financial Distributions with Stable Distributions", Chapter 3 of *Handbook of Heavy Tailed Distributions in Finance*, Elsevier Science, Amsterdam.

[34] Nolan, J. P. (2009), "*Stable Distributions - Models for Heavy Tailed Data*", Unfinished manuscript, Chapter 1 online at academic2.american.edu/ jpnolan/stable/stable.html.

[35] Panorska, A. K., Mittnik, S., & Rachev, S. T. (1995), "Stable GARCH Models for Financial Time Series", *Applied Mathematics Letters*, **8(5)**, 33–37.

[36] Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, S. M. & Jašić, T. (2007), *Financial Econometrics - From Basics to Advanced Modeling Techniques*, John Wiley & Sons, Hoboken, New Jersey.

[37] Roncalli, T. & Lagache, T. (2004), "A Gauss Implementation of Skew Normal/Student Distributions - The Skew Library", online at www.thierry-roncalli.com.

[38] Samorodnitsky, G. & Taqqu, M. (1994), *Stable Non-Gaussian Random Processes*, Chapman & Hall, Boca Raton.

[39] Taleb, N. (2007), "Black Swans and the Domains of Statistics", *The American Statistician*, **61(3)**, 198-200.

[40] Smith, T. (2008), "Indirect Inference", in *The New Palgrave Dictionary of Economics*, 2nd edition.

[41] Tsay, R. S. (2010), *Analysis of Financial Time Series - Third Edition*, Joun Wiley & Sons, New Jersey.

[42] Weron, R. (1996), "On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables", *Statistics and Probability Letters* **28(2)**, 165-171.

[43] Weron, R. (1996), "Correction to: "On the Chambers–Mallows–Stuck Method for Simulating Skewed Stable Random Variables", *MPRA Paper* 20761, University Library of Munich, Germany.

[44] Yang, S. R. & Brorsen, B. W. (1993), "Nonlinear Dynamics of Daily Futures Prices: Conditional Heteroskedasticity or Chaos?", *Journal of Futures Markets* **13(2)**, 175-191.

[45] Zolotarev, V. (1986), *One-dimensional Stable Distributions*, American Mathematical Society, Providence.