Ali Mahmoud Mohamed Eltaybe Abdelrahman

Asymptotic Theory for the Sample Autocorrelation Function and the Extremes of Stochastic Volatility Models

Ph.D. Thesis

2008

Thesis Advisor: Professor Thomas Mikosch
University of Copenhagen

Thesis Committee: Professor Michael Sørensen
University of Copenhagen
Professor Søren Asmussen
University of Aarhus
Professor Gennady Samorodnitsky
Cornell University

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

Asymptotic theory for the sample autocorrelation and the extremes of stochastic volatility models

Ali

June 14, 2008

Preface

This thesis is a partial fulfillment of the requirements for achieving the Ph.D. degree in Mathematics at the Department of Mathematical Sciences under the Faculty of Science at the University of Copenhagen.

First I wold like to thank the Mission Department of the Egyptian Ministry of High Education for its financial support throughout my studies in Denmark. I also thank the Department of Mathematical Sciences at Copenhagen University for accepting me.

On a personal level, I am very grateful to Professor Thomas Mikosch for accepting to supervise my Ph.D. studies and for his help and support. Staff and Faculty of the Department of Mathematical Sciences at Copenhagen University helped me a lot since I arrived in Denmark, and they were very friendly and supportive with me. The Egyptian Culture and Educational Bureau in Berlin gave me a special Social and Fellow Support for me and my family. A special thank goes to Jesper Lund Pedersen for helping me in translating the summary to Danish. Finally, I will forever be grateful to the lecturers at the Department whose courses I enjoyed. A special thank goes to Professor Martin Jacobsen for his advice and help.

In particular, I would like to thank my wife for her strong personal support. She chose to be away from Egypt and her parents in the last 4 years in order to be close to me.

Summary

This thesis is concerned with one of the nonlinear financial time series models, the stochastic volatility model. For financial time series, nonlinear time series models are better suited to describe their behavior. In contrast to linear time series models such as ARMA, the investigation of nonlinear models is still work in progress.

We start by collecting some of the standard properties of a stochastic volatility process. Under mild assumptions, the stochastic volatility sequence is a strictly stationary ergodic martingale difference sequence. Using this fact, we derive the central limit theorem for the sample mean of this sequence. Another property of the stochastic volatility model is strong mixing. This fact is helpful to establish a central limit theorem for the sample variance of this sequence. In both cases, under the assumption of finite variance innovations, the limiting distribution is a Gaussian distribution. In contrast to this case, under the assumption of infinite variance stable innovations, these estimators have a limiting distribution which is not Gaussian and rather unfamiliar. The autocorrelation function is an important tool in time series analysis. We study its estimator, the sample autocorrelation function and its limit Gaussian distribution via a multivariate central limit theorem.

Another important characterization of a time series is provided by its spectral density. We estimate the spectral density of a stochastic volatility process in some heavy– and light–tailed cases by using the raw periodogram. We derive the pointwise limit distribution of the periodogram. The limit of the periodogram in the case of iid Gaussian noise is exponential as in the case of an iid sequence. On the other hand, the limit in the case of $\alpha-$stable infinite variance noise is rather unfamiliar and, in particular, depends on the fact whether the frequency is a rational or irrational multiple of π.

In the second part of this thesis, we study the extremes in the stochastic volatility model and compare the results with the GARCH model. Under the assumption of regular variation of the noise of the stochastic volatility model, we show that the distributional limits of the maxima and order statistics are the same as in the iid case with the same marginal distributions as in the stochastic volatility model.

In addition, the stochastic volatility model itself inherits regular variation from the iid regular variation noise. The stochastic volatility model has extremal index one and upper tail dependence coefficient zero. These results show that the extremal behavior of a stochastic volatility process and of an iid sequence with regularly varying marginals is very much the same.

Resumé

I denne afhandling studerer vi en ikke-lineær model for finansielle tidsrækker, den stokastiske volatilitetsmodel. Ikke-lineære modeller giver et godt fit til finansielle tidsrækker, og forklarer tidsrækkenernes egenskaber bedre end lineære tidsrækkemodeller. Analysen af ikke-lineære tidsrækkemodeller er ikke afsluttet, og der eksisterer stadigvæk mange interessante problemer.

Først opsummerer vi basale egenskaber for stokastiske volatilitetsmodeller. Vi viser under milde betingelser, at en stokastisk volatilitetsmodel er en stærkt stationær og ergodisk martingaldifferens. Vi bruger dette til at bevise den centrale grænseværdisætning for middelværdien af stikprøven. Den stokastiske volatilitetsmodel er også strong mixing. Med hjælp af denne egenskab beviser vi en central grænseværdisætning for stikprøvevariansen. Hvis støjen i modellen har endelig varians, så er grænsefordelingerne gaussiske. Hvis vi antager, at den multiplikative støj i modellen har uendelig varians og stabil fordeling, beviser vi, at grænsefordelingerne er ukendte ikke-gaussiske. Autokorrelationsfunktionen er en af de vigtigste objekter i tidsrækkeanalyse. Vi undersøger stikprøveautokorellationsfunktionen, og beviser en multivariat central grænseværdisætning for en vektor af stikprøveautokorellationerne.

En stationær tidsrække er også karakteriseret af sin spektraltæthed. Vi estimerer spektraltætheden af den stokastiske volatilitetsmodel i tilfælde med både tunge of lette haler med hjælp af periodogrammet. Vi bestemmer grænsefordelingen af periodogrammet for en fast frekvens. Denne fordeling er exponentiel, når tidsrækken er iid gaussisk. Hvis støjen er stabil med uendelig varians, så er grænsefordelingen en ukendt fordeling, som afhænger af formen af frekvensværdien.

I den anden del af afhandlingen undersøger vi ekstremværdierne i en stokastisk volatilitetsmodel, og vi sammenligner også resultaterne med GARCH modellen. Hvis vi antager, at støjen er regulært varierende, viser vi at grænsefordelingerne af maksima og order statistics er de samme, som i tilfældet hvor tidsrækken er iid og har samme marginalfordeling som den stokastiske volaltilitetsmodel. Blandt andet er den stokastiske volatilitetsmodel også regulært varierende. Den stokastiske volatilitetsmodel har ekstremalindeks 1 og den øvre haleafhængighedskoefficient er 0. Alle resultater viser, at de ekstremale

egenskaber af den stokastiske volatilitetsmodel er de samme som de ekstremale egenskaber af en iid følge med den samme marginalfordeling.

Contents

Chapter 1

Introduction

1.1 Objectives of time series analysis

In practice, observations of various phenomena are often recorded sequentially over time. Values in the future depend, usually in a stochastic manner, on the observations available at the present. Such dependence makes it worthwhile to predict the future from its past. Time series analysis deals with such records that are collected over time. The time order of data is important. We write $(X_t)_{t \in \mathbb{Z}}$ for any time series. The unit of the time scale is usually implicit in this notation.

The following example of a real-life time series is often used in the financial time series literature.

Example 1.1.1. (The Standard and Poor's 500 Index)
The Standard and Poor's 500 index (S&P 500) is a value-weighted index based on the prices of the 500 stocks that account for approximately 70% of the total U.S. equity market capitalization. The selected companies tend to be the leading companies in leading industries within the U.S. economy. The index is a market capitalization-weighted index (shares outstanding multiplied by stock price)-the weighted average of the stock price of the 500 companies. In 1980, the S&P 500 became a component of the U.S. Department of Commerce's Index of Leading Economic Indicators, which are used to gauge the health of the U.S. economy. It serves as a benchmark of stock market performance against which the performance of many mutual funds is compared. It is also a useful financial instrument for hedging the risks of market portfolios. The S&P index began in 1923 when the Standard and Poor's Company introduced a series of indices, which included 323 companies and covered 26 industries. The current S&P 500 Index was introduced in 1957.

This example is one of a multitude of time series data existing in astronomy, biology, economics, finance, environmental studies, engineering, and other areas.

Depending on the background of applications, the objectives of time series analysis are diverse. Statisticians usually view a time series as a realization from a stochastic process. A fundamental task is to unveil the probability law that governs the observed time series. With such a probability law, we can understand the underlying dynamics, forecast future events, and control future events via intervention. These are the three main objectives of time series analysis.

Time series analysis rests on proper statistical modeling. In selecting a model, interpretability, simplicity, and feasibility play important roles. A selected model should reasonably reflect the physical law that governs the data. Everything else being equal, a simple model is usually preferable.

1.2 Linear and nonlinear time series

A very popular class of time series models consists of the autoregressive moving average (ARMA) models. ARMA models are frequently used to describe linear dynamic structures, to depict linear relationships among lagged variables (see Example 2.1.6). It is one of the most frequently used families of parametric models in time series analysis. This is due to their flexibility in approximating many stationary processes. From the pioneering work of Yule [41] on AR modeling of the sunspot numbers to the work of Box and Jenkins [6] that marked the maturity of ARMA modeling in terms of theory and methodology, linear Gaussian time series models flourished and dominated both theoretical explorations and practical applications.

However, there is no universal key that can open every door. Moran, in his classical paper in 1953 [31], on modeling the Canadian lynx data, hinted at a limitation of linear models. He drew attention to the "curious feature" that the residuals for the sample points greater than the mean were significantly smaller than those for the sample points smaller than the mean. This can be well-explained in terms of the so-called "regime effect" at different stages of population fluctuation.

Another application which challenges the linear time series model are financial time series (see Section 2.2 for stylized facts about financial time series). Among the stylized facts of financial time series (log −returns), we have:

- Zero sample autocorrelations $\widehat{\rho}_X(h)$ for (X_t) at almost all lags $h > 0$, with a possible exception at the first lag although the estimated autocorrelation $\widehat{\rho}_X(1)$ is usually rather small (often about 0.1).

- Very slowly decaying sample autocorrelations of $(|X_t|)$ and (X_t^2). In this context, one often refers to long memory in the volatility.

- Occurrence of extremely large and small X_t's clustered at certain instants of time, caused by turbulences in the market due to financial crashes, political decisions, war, etc.

If we wanted to explain the dependence structure of such a model by an ARMA model with iid noise (Z_t), we would have to restrict our attention to models of the form $X_t = Z_t$ or moving average models of very low order. Indeed, for an MA(q) (moving average process of order q) model, X_t and X_{t+q+1} are independent, hence $|X_t|^r$ and $|X^r_{t+q+1}|$ are independent for any $r > 0$ and therefore $\rho_{|X|^r}(h) = 0$ for $|h| > q$. This means that the effect of non-vanishing autocorrelations of the $(|X_t|^r)$ processes for $r = 1, 2$ cannot be explained by an MA(q) model with iid noise (Z_t).

Beyond the linear domain, there are infinitely many non–linear forms to be explored. In the econometrics literature, the ARCH processes (autoregressive processes with conditional heteroscedasticity) and their numerous modifications have attracted significant attention. One of the 2003 Bank of Sweden Prizes for Economics, better known under the name of Nobel Prize for Economics, was awarded to Robert Engle who introduced the ARCH model in the celebrated 1982 paper [12]. We also refer to the collection of papers on the theme "ARCH" edited by Engle [13]. A generalization of ARCH is given by the popular GARCH models (generalized ARCH). For example, GARCH(1,1) is given by

$$\begin{aligned} X_t &= \sigma_t Z_t, \qquad t \in \mathbb{Z}, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

where (Z_t) is iid noise and $\alpha_0, \alpha_1, \beta_1$ are positive parameters.
There are some reasons for the wide use of GARCH. Mikosch [27] summarized these reasons which can be given by:

- Its relation to ARMA processes suggests that the theory behind it might be closely related to ARMA process theory which is well studied, widely known and seemingly "easy".

  This opinion is, however, wishful thinking. The difference to standard ARMA processes is due to the fact that the noise sequence in the ARMA representation of (X_t^2) depends on the X_t's themselves, so a complicated non–linear relationship of the X_t's builds up.

- A second argument for the use of GARCH models is that, even for a GARCH(1,1) model with three parameters one often gets a reasonable fit to real-life financial data, provided that the sample has not been chosen from a too long period making the stationarity assumption questionable. Tests for the residuals of GARCH(1,1) models with estimated parameters $\alpha_0, \alpha_1, \beta_1$ give the impression that the residuals very much behave like an iid sequence. Some evidence on this issue can be found in the paper of Mikosch and Stărică [29].

- The most powerful argument in favor of GARCH models, from an applied point of view, is the fact that the statistical estimation of the parameters of a GARCH process is rather uncomplicated. This attractive property has led S+ to provide us with a module for the statistical inference and simulation of GARCH models, called S+FinMetrics.

Another non–linear model for financial time series is the stochastic volatility model, given by

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z},$$

where the noise process (Z_t) and the volatility process (σ_t) are assumed to be independent. Although this model is much easier than the GARCH in understanding its probabilistic properties, the lack of estimation procedures for the stochastic volatility model made it less attractive than the GARCH family. Recently, the attitude towards the stochastic volatility model has changed and a variety of estimation techniques (such as GMM and quasi-MLE) have been developed as well. We study some properties of the stochastic volatility model in Section 2.4, including stationarity, mixing properties, moments.

1.3 Objectives and methodology of the thesis

As mentioned before, financial time series have some characteristic features which should be taken into consideration when analyzing them. By now, there is no "perfect" time series model for returns. Some of these models score high in some cases and not so well in other cases. This leads to many different models for financial time series. Some of these models are useful as regards certain aspects of time series analysis, such as estimation of the parameters, capturing a large variety of real–life financial time series: they give a good fit in many examples, their probabilistic properties are easy to study as regards some other aspects. On the other hand, the available models usually suffer from some shortage as regards some other aspects. Therefore, in general we use different models according to the slogan "This model is correct, having some shortage and the other model is wrong, having some advantages". This is the problem of statistics in general.

We already mentioned some time series models, ARMA, GARCH and stochastic volatility models. The (linear) ARMA process is not really suited for modeling financial time series. The GARCH model is very attractive because of its simplicity in estimation problems and its capability to fit real–life return data. There exists an enormous body of literature on the properties of the GARCH model. This work aims at studying strict and weak stationarity, power law tails, extremes, mixing and the statistical properties of parameter estimators (such as consistency and asymptotic normality under general assumptions on the distribution of Z_t). Despite the many papers and books which have been written on the "ARCH" theme, the GARCH model is still not completely understood as regards its probabilistic properties. The stochastic volatility model is still in the frame of the picture when we speak about financial time series analysis. The problem of estimation of the stochastic volatility model parameters was a serious threat for this model. New methods for estimation made the stochastic volatility model more interesting for financial time series analysis.

1.3.1 Asymptotic theory for the sample autocorrelation for the stochastic volatility model

In this thesis our interest focuses on the stochastic volatility model as a tool for financial time series analysis. Our first aim will be to derive an asymptotic theory for the stochastic volatility model. If one estimates the parameters of the model from the sample data, the estimators often include the sample mean $\overline{X}_n$, the sample variance and the sample autocorrelations for (X_t) and $(|X_t|^p)$ for some $p > 0$, e.g. for $p = 2$. We start our study, as usual in time series analysis, by exploring the probabilistic properties of the stochastic volatility process.

In the beginning, an introduction to the probabilistic tools which are used throughout the thesis are given. The ergodic theorem and the central limit theorem play an important role for the asymptotic theory of the estimators. Martingale properties are crucial for the central limit theorem of (X_t). For the $(|X_t|^p)$, $p > 0$, sequence, mixing conditions are needed for the application of the central limit theorem for dependent variables. We introduce different types of mixing conditions, see Section 2.1.4. In the case of the stochastic volatility model, strong mixing is convenient. Using the Cramér -Wold device, we get a multivariate version of the central limit theorem, see Section 2.1.5. The multivariate central limit theorem is useful, e.g. when studying the joint convergence of the sample mean and the sample variance, but also when we consider the asymptotic behavior for a lagged vector of sample autocorrelations or sample autocorrelations. This is the topic of Section 2.7.

Stationarity is an important property in time series analysis. For linear time series analysis often weak stationarity suffices, whereas for non–linear time series analysis strict stationarity is important. In practice, financial time series in general are non–stationary. If there is a doubt about stationarity we can apply some transformations to make the data stationary. The stochastic volatility sequence is a strictly stationary sequence under some mild conditions, see Section 2.4.2.

One of the advantages of the stochastic volatility model is that we can easily study its probabilistic properties. Its moments are used to compute variances, autocovariances, and autocorrelation functions. In Section 2.4.4, we study the moments of the stochastic volatility sequence (X_t) and its powers $(|X_t|^p)$. We introduce the results in the case $\log \sigma_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}$ where (η_t) is an iid sequence and (ψ_t) are suitable constants and in the special case that η has a Gaussian distribution.

We need a central limit theorem to study the asymptotic behavior of the estimators of moments and covariances. The central limit theorem for iid sequences is not applicable for financial time series as the data are dependent through time. For a stochastic volatility model, the sequence (X_t) is a strictly stationary ergodic martingale difference sequence under mild assumptions, see Section 2.4.4. Hence the central limit theorem for strictly stationary ergodic martingale difference sequences can be applied to the sample mean of a stochastic volatility model. Another important sequence in financial time series analysis consists of the series of the powers $(|X_t|^p)$. In practice, it is common to study the sequences $(|X_t|)$ and (X_t^2) in order to detect non–linearities in the sequence of the returns (X_t). If the sequence (X_t) is strongly mixing the sequence $(|X_t|^p)$, $p > 0$, is also strongly mixing sequence. Thus the central limit theorem for strictly stationary strongly mixing sequences can be applied to get the asymptotic behavior of the sample mean of $(|X_t|^p)$. Using these central limit theorems and the Cramér–Wold device, we get multivariate versions of the central limit theorem. Section 2.5 studies the joint central limit theorem for the sample mean and the estimated variance of the stochastic volatility sequence (X_t) and the power sequence $(|X_t|^p)$.

First, we are interested in the case of finite variance innovations Z_t, where we can use the standard central limit theorems from Section 2.1.5. We also consider the non–standard case, when Z_t has infinite variance. As a special case we study iid $\alpha-$stable innovations (Z_t). We study the joint asymptotic behavior of $(\overline{X}_n, \widehat{s}_n^2)$, where $\overline{X}_n$ is the sample mean and $\widehat{s}_n^2$ is the sample variance, under the assumption that (Z_t) is iid symmetric stable $(s\alpha s)$ for some $\alpha \in (0, 2)$. The limiting joint mixed characteristic function–Laplace–Stieltjes transformation of $(\overline{X}_n, \widehat{s}_n^2)$ is derived. We also study the limiting distribution of the standardized sample mean $\sqrt{n}\overline{X}_n/\widehat{s}_n$. It is symmetric and has unit variance but it is not standard normal. This result is in contrast to the case when $var(X)$ is finite, when the limit is normal.

The other side of the coin in time series analysis is spectral analysis. In this thesis we study the asymptotic theory for the periodogram of a stochastic volatility sequence. The spectral representation of a stationary process (X_t) essentially decomposes (X_t) into a sum of sinusoidal components with uncorrelated random coefficients. In conjunction with this decomposition there is a corresponding decomposition into sinusoids of the spectral density of (X_t). The spectral decomposition is thus an analogue for stationary stochastic processes of the more familiar Fourier representation of deterministic functions. Herglotz's theorem and the spectral distribution function of a stationary process are important building blocks in this area. The standard methods for estimating the spectral density are based on the periodogram. We again study the asymptotic behavior of the periodogram for the stochastic volatility model when either (Z_t) is iid normal or (Z_t) is iid infinite variance $s\alpha s$, for some $\alpha < 2$, see Section 2.6. We also study the self-normalized periodogram.

An important fact about financial time series is related to the autocovariance/autocorrelation functions, which are usually estimated by their sample autocovariance/autocorrelation functions. Of course, as for any statistical estimator, we need to study its asymptotic behavior. The stochastic volatility sequence (X_t) is a strictly stationary ergodic martingale difference sequence, so the central limit theorem for strictly stationary ergodic martingale difference sequences is applicable in this case. We study the joint limit distribution for the sample autocovariance and the sample autocorrelation functions in the case of Gaussian and non-Gaussian $\log \sigma_t$. For the sequence $(|X_t|^p)$, $p > 0$, its autocorrelation function is relevant in financial time series analysis. For example, for real–life data (X_t), the autocorrelation function of the time series $(|X_t|)$ does not decay very fast for large lags h. The sequence $(|X_t|^p)$ is a strictly stationary strongly mixing sequence, so the central limit theorem for this kind of sequence and its multivariate versions can be applied to the sample autocovariances and autocorrelations. As the autocovariance for lag $h > 0$ does not equal zero, the variance covariance matrix for the limiting distribution is more complicated than in the case when $p = 1$. These are the topics of Section 2.7.

1.3.2 Extremes in the stochastic volatility model

When reading a newspaper and studying the financial index on the first page, one expects to find big losses or big profits in the financial market. This is a simple example when extremes matter. In general, extreme value theory takes a special interest in very different areas, and financial time series analysis is just one of them. Some time series analysts consider extremes as outliers. This means, that they do not consider them as belonging to the probability distribution underlying the data. Dealing with these extremes as outliers, leaves one no choice for a quantitative analysis of these values. Another point of view, especially for long records of phenomena, is to model the extremes. Extreme value theory deals with this topic. In Section 3.1.1, we give a short review of classical extreme value theory for iid random variables.

Real–life financial returns are heavy–tailed in the sense that their distributional tails exhibit power law behavior. This means in particular that sufficiently high moments of returns might be infinite. For modeling heavy tails of financial time series, the class of regularly varying distributions is particularly attractive. A random variable X is said to be regularly varying with index $\alpha \geq 0$ if there exist constants $p, q \geq 0,\ q = 1 - p$ satisfies

$$P(X > x) = (p + o(1))x^{-\alpha}L(x) \quad \text{and} \quad P(X \leq -x) = (q + o(1))x^{-\alpha}L(x),$$

for every $x > 0$, $L(\cdot)$ is a slowly varying function, i.e. $L(cx)/L(x) \to 1$, as $x \to \infty$, for every $c > 0$. In many real–life data, there is evidence in favour of regularly varying distributions, see [11]. In addition, mathematical reasons for using regularly varying distributions for modeling extremal events come from their relation with extreme value theory and the fluctuation theory of modern models. A probability distribution is regularly varying with index $\alpha \in (0, 2)$ if and only if it belongs to the domain of attraction of an infinite variance stable distribution. Moreover, a regularly varying distribution with index $\alpha > 0$ belongs to the maximum domain of attraction of one particular extreme value distribution, the Fréchet distribution

$$\Phi_\alpha(x) = e^{-x^{-\alpha}}, \qquad x > 0.$$

In Sections 3.1.3 and 3.1.4, we introduce univariate/multivariate regular variation, major properties of regularly varying variables and vectors, and give examples. Two functions of regularly varying vectors, product and sum, are particularly important for our purposes. We study them in these sections.

An important application of regular variation is modeling extremal events in financial time series. In Section 3.2.1 we study the regular variation of the GARCH model. In particular, we study the so–called spectral distribution of the GARCH$(1, 1)$ case under mild conditions. This distribution is a quantitative measure of the likelihood that multivariate extremes occur in a certain direction. The knowledge of the spectral distribution enables one to conclude various limiting conditional probabilities, among which the tail dependence coefficient has gained some popularity in quantitative risk management, see e.g. McNeil et al [25]. This leads us to study some consequences for the extremes of the GARCH$(1, 1)$. We also study regular variation in the stochastic volatility model in Section 3.2.2, which is a less demanding task.

In general, sequences of iid random variables are not very often met in time series practice. An important fact about financial time series is that their extremes usually occur in clusters. Classical extreme value theory does not apply to the stochastic volatility model; we need some extra conditions on the dependence structure. There are different models for dependence. In the case of heavy–tailed distributions, the autocorrelation function is not a perfect measure of dependence. To measure the dependence in the tails, the extremal index of a stationary sequence is a useful tool. Smaller values of the extremal index indicate stronger clustering of large fluctuations, i.e. more dependence in the tails.

Perhaps surprisingly (and in contrast to the GARCH model), the extremal index of a stochastic volatility model is one, i.e. it has the same value as for an iid sequence. In Section 3.3 we study the asymptotic behavior of the extremes in a stochastic volatility process. Under the assumption of regular variation on the noise (Z_t) we show that the limits of the maxima and ordered statistics are the same as in the iid case with the same marginal distributions as in the stochastic volatility model, and we also show that the weak limit of the point processes of exceedances is a homogeneous Poisson process.

1.4 Summary

The thesis contributes results in the following areas:

1) We give an overview of the asymptotic behavior of the sample mean and sample variance for a stochastic volatility model under strong mixing of the volatility sequence. Whereas the results in the finite variance case follow standard patterns, the results, where the noise is iid symmetric $\alpha-$stable, are new.

2) We study the asymptotic behavior of the periodogram of the stochastic volatility model in the cases when the noise is either iid Gaussian or infinite variance $\alpha-$stable. We also consider the self–normalized (or standardized) periodogram at a fixed frequency. Whereas the limit of the periodogram in the case of iid Gaussian noise is exponential as in the case of an iid sequence, the limits in the case of $\alpha-$stable infinite variance noise are rather unusual and, in particular, depend on whether the frequency is a rational or irrational multiple of 2π. Whereas this phenomena has been observed in Klüppelberg and Mikosch [23] for an iid sequence, the results are new for the stochastic volatility model. Since the self–normalized periodogram has a limit distribution in both cases, the finite and the infinite variance ones, the limits of the smoothed self–normalized periodogram can be interpreted as a spectral density irrespective of whether the spectral density of the model is well defined or not.

3) We provide a limit theory for the sample autocovariance and autocorrelation functions of the stochastic volatility model and its absolute values and any positive power. Since we exclusively made use of strong mixing, the results follow by standard arguments from the theory of strong mixing.

4) We study the extremal behavior of a stochastic volatility model under the assumption that the iid noise be regularly varying. We show that the stochastic volatility model itself inherits regular variation in the sense that its finite–dimensional distributions are regularly varying with a spectral distribution concentrated at the coordinate axes. This is again analogous to the case of an iid sequence. In particular,

the stochastic volatility sequence has extremal index one and has upper tail zero. These are two other properties it shares with an iid regularly varying sequence.

In since, the results of this thesis show that the asymptotic results for sum–like functionals and extremes of a strongly mixing stochastic volatility sequence very much parallel the theory for an iid sequence.

Chapter 2

Asymptotic theory for stochastic volatility processes

This chapter deals with the stochastic volatility model. We study the properties of this model including its moments. The main aim is to study the asymptotic behavior of its estimated autocorrelations.

2.1 Preminaries on time series

In this section the emphasis is on some tools which play an important role in the study of the stochastic volatility model.

2.1.1 Weak and strict stationarity

When looking at a time series we hope to see some sort of "regularity". In particular, when looking at different segments of the series we might expect to discover similar patterns or similar behavior. This can be made precise by introducing the notion of "stationarity". Before we can do that we need another fundamental quantity:

Definition 2.1.1. (The autocovariance function (ACVF))
Let $(X_t)_{t\in\mathbb{Z}}$ be a process such that $var(X_t) < \infty$ for all $t \in \mathbb{Z}$. The function

$$\gamma_X(s,t) = cov(X_s, X_t) = E[(X_s - EX_s)(X_t - EX_t)]; \quad s,t \in \mathbb{Z}, \tag{2.1.1}$$

is called the autocovariance function of the process (X_t). We write ACVF for short.

Definition 2.1.2. ((Weak) stationarity)
The time series $(X_t)_{t\in\mathbb{Z}}$ is said to be stationary if the following relations hold:

- $E|X_t|^2 < \infty, t \in \mathbb{Z}$.

- $EX_t = m, t \in \mathbb{Z}$, *for a constant* m.
- $\gamma_X(s,t) = \gamma_X(s+h, t+h)$ *for all* $s, t, h \in \mathbb{Z}$.

Definition 2.1.3. (Strict stationarity)
The time series $(X_t)_{t\in\mathbb{Z}}$ *is said to be strictly stationary if for any* $h \in \mathbb{Z}$ *and* $t \geq 0$, *the random vectors* $(X_h, \ldots, X_{t+h})$ *and* $(X_0, \ldots, X_t)$ *have the same distribution.*

Any strictly stationary sequence will always be indexed by the integers $\mathbb{Z}$. We will usually drop the index set. For any strictly stationary sequence (X_t) we write X for a generic element of the sequence.

- If $(X_t)_{t\in\mathbb{Z}}$ is stationary then the autocovariance function of (X_t) can be written as

$$\gamma_X(h) \equiv cov(X_t, X_{t+h}) = \gamma_X(0,h)$$

- A strictly stationary process with finite second moments is (weakly) stationary. The converse of the previous statement is not true.
- There is an important case in which (weak) stationarity implies strict stationarity. It is a Gaussian time series (if the finite dimensional distributions of (X_t) are all multivariate normal).

The weak stationarity assumes that only the first two moments of time series are time invariant provided that the process has finite second moments. Weak stationarity is primely used for linear time series, such as ARMA processes, where we are mainly concerned with the linear relationships among variables at different times. In fact, the assumption of weak stationarity suffices for most linear time series analysis, such as in spectral analysis. In contrast, we have to look beyond the first two moments if our focus is on nonlinear relationships. This explains why strict stationarity is often required in the context of nonlinear time series analysis.

2.1.2 Autocovariance, autocorrelation functions and their sample analogs of a stationary process

Definition 2.1.4. (Autocorrelation function ACF)
Let $(X_t)_{t\in T}$ *be a stationary process such that* $var(X_t) < \infty$ *for all* $t \in T$. *The function*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \tag{2.1.2}$$

is called the autocorrelation function at lag $h \in \mathbb{Z}$.

Since the ACVF and ACF are unknown for a real-life data set, they have to be estimated by statistical means. Standard estimators are given by the sample autocovariances $\hat{\gamma}_X(h)$ and the sample autocorrelations $\hat{\rho}_X(h)$ at lag $h \in \mathbb{Z}$.

Definition 2.1.5. (The sample autocovariance function and the sample autocorrelation function)
The sample ACVF and sample ACF of a stationary process (X_t) are given by

$$\hat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), \quad \hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}, \quad 0 \le h < n, \tag{2.1.3}$$

respectively, where $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$ is the sample mean.

Autocovariances, autocorrelations and their sample versions are relevant for the study of the dependence structure and for building theoretical time series models. Whenever we work with these quantities we are in the time domain of time series analysis. Another way of looking at time series is the frequency domain where one studies the spectral properties of such series.

Example 2.1.6. (Linear Process, ARMA)
The time series (X_t) is said to be an ARMA(p, q) process if it is stationary and satisfies the ARMA difference equations

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad t \in \mathbb{Z}, \tag{2.1.4}$$

for given real numbers $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and a white noise sequence (Z_t) with $var(Z) > 0$. The process (X_t) is linear if it has representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}.$$

An ARMA(p, q) process is said to be causal if it has representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}, \tag{2.1.5}$$

for constants ψ_j satisfying

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \tag{2.1.6}$$

This means all causal ARMA(p, q) processes have a linear series representation. The following formula can be obtained for any linear process, hence in particular for ARMA(p,q) process

$$var(X) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2, \tag{2.1.7}$$

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}, \quad h \in \mathbb{Z}, \tag{2.1.8}$$

$$\rho_X(h) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}}{\sum_{j=0}^{\infty} \psi_j^2}, \quad h \in \mathbb{Z}. \tag{2.1.9}$$

For causal ARMA processes, there exist $k > 0, a < 1$, such that $|\psi_j| \leq ka^j$ for all $j \geq 0$. Hence for all $h \geq 0$

$$\begin{aligned} |\gamma_X(h)| \leq \sigma^2 \sum_{j=0}^{\infty} |\psi_j||\psi_{j+h}| &\leq \sigma^2 k^2 \sum_{j=0}^{\infty} a^j a^{j+h} \\ &= \sigma^2 k^2 a^h \frac{1}{1-a^2}. \end{aligned}$$

Therefore both $\gamma_X(h)$ and $\rho_X(h)$ converge to zero exponentially fast as $h \to \infty$.

2.1.3 Ergodicity

If $(X_t, t \geq 0)$ are independent identically distributed, then their sample mean converges to EX if $E|X| < \infty$. When (X_t) is merely strictly stationary, convergence to a nonconstant limit is possible. In order to be able to study the limiting random variable in the stationary case ergodicity and invariance play an important role (see for example [8], Chapter 6 and [38]).

Definition 2.1.7. (Measure-preserving function)
Let $(\Omega, \mathcal{F}, P)$ be a probability space. A transformation T from Ω to Ω is measure-preserving (alternatively "T preserves P") if it is measurable and if $P[T^{-1}A] = P(A)$ for all $\mathcal{A} \in \mathcal{F}$.

Every measure-preserving transformation generates a strictly stationary sequence and any strictly stationary sequence can be represented by means of a measure-preserving transformation [38, p. 168].

Definition 2.1.8. (Invariant event)
Given a measure-preserving transformation T, a measurable event A is said to be invariant if $T^{-1}A = A$. If $P(T^{-1}A \triangle A) = 0$ then A is said to be almost invariant.

The collection of almost invariant events $\overline{\mathcal{I}}$ forms a σ-field which is the completion of $\mathcal{I}$ with respect to $\mathcal{F}$ and P (that is, every almost invariant event differs from an invariant event by a measurable event of probability 0.)

Definition 2.1.9. (Ergodic transformation) *[38, p. 172]*

A measure-preserving transformation is ergodic if for all $A \in \mathcal{T}$ either $P(A) = 0$ or $P(A) = 1$.

Lemma 2.1.10. (The mean ergodic theorem) *[38, p. 178]*

Let T be a measure- preserving transformation. Then $E|X| < \infty$ implies

$$E\left|\sum_{k=0}^{n-1} X(T^k)/n - E[X|\mathcal{T}]\right| \to 0$$

as $n \to \infty$ (that is, convergence in $\mathfrak{L}_1$).

Theorem 2.1.11. (The pointwise ergodic theorem for a strictly stationary sequence) *[38, p. 181]*
Let $(X_t, t \geq 1)$ be a strictly stationary sequence with $E|X| < \infty$. Then $\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X|\mathcal{T}]$ a.s. If in addition $(X_t, t \geq 1)$ is ergodic, $\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X]$ a.s.

In particular, so-called Bernoulli shifts $Y_t = f((X_{t+h})_{h\in\mathbb{Z}})$, $t \in \mathbb{Z}$, of a strictly stationary ergodic sequence (X_t) are again strictly stationary ergodic sequences. Here f is any measurable real-valued function f, and (X_t) can be an ergodic process with values in some abstract space, see for example [24]. In particular, (Y_t) is ergodic if (X_t) is an iid sequence.

2.1.4 Mixing

The classical asymptotic theory in statistics is built on the central limit theorem and the law of large numbers for sequences of independent random variables. For time series, there are rather complicated dependence structures. Certain asymptotic independence conditions are needed in order to derive large sample properties for time series inferences. A mixing time series can be viewed as a sequence of random variables for which the past and distant future are asymptotically independent. For mixing sequences, both the law of large numbers (i.e., ergodic theorem) and central limit theorem can be established. Mixing conditions for strictly stationary processes can be given by defining mixing coefficients. These coefficients measure the strength of dependence for the two segments of a time series that are apart from each other in time.

Definition 2.1.12. (Mixing conditions) *[15, p. 68-69]*
Let $\mathfrak{F}_a^b$ be the $\sigma-$field generated by $X_t, a \leq t \leq b$, and $\mathfrak{L}^2(\mathfrak{F}_a^b)$ consists of the $\mathfrak{F}_a^b$-measurable random variables with finite second moment. A strictly stationary sequence $(X_t)_{t\in Z}$ is said to be

α-mixing, also called strongly mixing *if*

$$\sup_{B\in\mathfrak{F}_{-\infty}^0, C\in\mathfrak{F}_t^\infty} |P(B \cap C) - P(B)P(C)| =: \alpha_t \longrightarrow 0, \quad t \longrightarrow \infty, \tag{2.1.10}$$

β-mixing, also called absolute regular *if*

$$E\left(\sup_{B\in\mathfrak{F}_t^\infty}|P(B)-P(B|X_0,X_{-1},X_{-2},\cdots)|\right)=:\beta_t\longrightarrow 0,\quad t\longrightarrow\infty, \tag{2.1.11}$$

ρ-mixing *if*

$$\sup_{X\in\mathfrak{L}^2(\mathfrak{F}_{-\infty}^0),Y\in\mathfrak{L}^2(\mathfrak{F}_t^\infty)}|Corr(X,Y)|=:\rho_t\longrightarrow 0,\quad t\longrightarrow\infty, \tag{2.1.12}$$

φ-mixing *if*

$$\sup_{B\in\mathfrak{F}_{-\infty}^0,C\in\mathfrak{F}_t^\infty,P(B)>0}|P(C)-P(C|B)|=:\varphi_t\longrightarrow 0,\quad t\longrightarrow\infty, \tag{2.1.13}$$

and

ψ-mixing *if*

$$\sup_{B\in\mathfrak{F}_{-\infty}^0,C\in\mathfrak{F}_t^\infty,P(B)P(C)>0}\left|1-\frac{P(C|B)}{P(C)}\right|=:\psi_t\longrightarrow 0,\quad t\longrightarrow\infty. \tag{2.1.14}$$

The following diagram illustrates the relationships between the five mixing conditions:

$$\psi-mixing\quad\rightarrow\quad\varphi-mixing\quad\begin{matrix}\nearrow & \beta-mixing & \searrow\\ & & \\ \searrow & \rho-mixing & \nearrow\end{matrix}\quad\alpha-mixing.$$

In general the α-mixing condition, also called strong mixing, is the weakest among the five, which is implied by any one of the four other mixing conditions. On the other hand, ψ-mixing is the strongest condition. However, e.g. for Gaussian processes, ρ-mixing is equivalent to α-mixing and therefore is weaker than the β-mixing condition [15, p. 69]. Usually strong mixing plays a fundamental role in time series analysis.

The decay rate of the mixing coefficients to zero as $t\to\infty$ is a measure of the range of dependence or of the memory in the sequence (X_t). If it decays to zero at an exponential rate, then (X_t) is said to be mixing with geometric rate.

The rate function (α_t) is closely related to the rate of decay of the ACF ρ_X of the stationary process (X_t) [17, p. 309]. For example, the following classical result holds.

Lemma 2.1.13. (ACF upper limit)
Assume $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary and strongly mixing with rate function (α_h) and $E(|X|^{2+\delta})<\infty$ for some $\delta>0$. Then the relation

$$|\rho_X(h)|\le c\alpha_h^{\frac{\delta}{2+\delta}},\quad h\ge 0 \tag{2.1.15}$$

holds for some constant $c>0$.

In particular, if (α_h) decays to zero exponentially fast as $h \to \infty$ so does $(\rho_X(h))$.

Example 2.1.14. (The rates of decays for ARMA)
Assume that (X_t) is a causal Gaussian ARMA process. Then it follows from [33] that (α_h) decays exponentially fast. Of course, in this case it is well known (also for non-Gaussian causal ARMA processes) that $\rho_X(h)$ decays to zero exponentially fast, see Example 2.1.6.

Since strong mixing is defined via the $\sigma-$fields generated by the random variables X_t, it remains valid for any sub-σ fields. In particular, if one considers the time series

$$Y_t = g(X_{t-k}, \ldots, X_{t+k}), \quad t \in \mathbb{Z},$$

for any $k \geq 0$ and a measurable function g assuming values in d-dimensional Euclidean space, then (Y_t) is again strictly stationary and strongly mixing with a rate function $\alpha_h(g) \leq \alpha_h$. This follows from the fact that the Borel σ-fields generated from the Y_t's are sub-σ-fields of those generated from $(X_{t-k}, \ldots, X_{t+k})$.

We will typically be interested in measurable transformations of the form

$$Y_t = |X_t|^p, \quad t \in \mathbb{Z},$$

for some $p > 0$, or

$$Y_t = (X_t^2, X_t X_{t+1}, \ldots, X_t X_{t+k}),$$

for some $k \geq 0$ and the corresponding analogs for $|X_t|^p$, some $p > 0$. These processes have essentially the same strong mixing properties as the original (X_t) sequence.

2.1.5 Central limit theorems

Here the focus will be on the central limit theorem for dependent variables.

Theorem 2.1.15. (Central limit theorem for strictly stationary ergodic martingale difference sequences [2, p. 206])
Let (X_t) be a strictly stationary, ergodic sequence for which

$$E(X_n | X_1, \ldots X_{n-1}) = 0, \quad a.s. \tag{2.1.16}$$

and for which $EX^2 = \sigma^2 \in (0, \infty)$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n} X_i \xrightarrow{d} N(0, \sigma^2). \tag{2.1.17}$$

This result is remarkable insofar that it holds under a second moment condition as in the iid case. For the iid case $\sigma^2 < \infty$ is necessary and sufficient for relation (2.1.17).

Theorem 2.1.16. (Central limit theorem for strictly stationary strongly mixing sequences [17, p. 346–347])

Let the strictly stationary sequence (X_t) *satisfy the strong mixing condition with mixing coefficients* α_t, *and let* $EX = 0$ *and* $E|X_t|^{2+\delta} < \infty$ *for some* $\delta > 0$. *If* $\sum_{i=1}^{\infty} \alpha_i^{\frac{\delta}{2+\delta}} < \infty$, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n} X_i \xrightarrow{d} N(0, \sigma^2 + 2 \sum_{j=1}^{\infty} \gamma_X(j)). \tag{2.1.18}$$

The asymptotic variance in (2.1.18) is finite. This is an immediate consequence of (2.1.15) and

$$\sum_{i=1}^{\infty} \alpha_i^{\delta/(2+\delta)} < \infty.$$

Theorem 2.1.17. (Cramér-Wold device [21, p. 150])

Suppose that $X, X_1, X_2, \ldots$ *are k-dimensional random vectors. Then,* $X_n \xrightarrow{d} X$ *if and only if for all choices of* $a = (a(1), \ldots, a(k)) \in \mathbb{R}^k$,

$$\sum_{i=1}^{k} a(i) X_n(i) \xrightarrow{d} \sum_{i=1}^{k} a(i) X(i). \tag{2.1.19}$$

Lemma 2.1.18. (The multivariate central limit theorem for strictly stationary ergodic martingale difference sequences)

Let $X, X_1, X_2, \ldots$ *be a strictly stationary* $k-$*dimensional ergodic martingale difference sequence satisfying* (2.1.16) *and* $E|X|^2 < \infty$. *Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} N(\underline{0}, \Sigma), \tag{2.1.20}$$

where Σ *is the variance-covariance matrix of* X.

Proof. Since $E(X_n | X_1, X_2, \ldots, X_{n-1}) = 0$,

$$E(a' X_n | X_1, X_2, \ldots, X_{n-1}) = a' E(X_n | X_1, X_2, \ldots, X_{n-1}) = 0,$$

for all $a \in \mathbb{R}^k$. An application of Theorem 2.1.15 yields that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a' X_i \xrightarrow{d} N(0, var(a'X)).$$

Notice that

$$var(a'X) = a' \Sigma a.$$

By the Cramér-Wold device (Theorem 2.1.17), Relation (2.1.20) holds. □

Lemma 2.1.19. (The multivariate central limit theorem for strictly stationary strongly mixing sequences)

Let $X, X_1, X_2, \ldots$ be a strictly stationary $k-$dimensional strongly mixing sequence satisfying $E|X|^{2+\delta} < \infty$ for some $\delta > 0$ and $\sum_{i=1}^{\infty} \alpha_i^{\frac{\delta}{2+\delta}} < \infty$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} N(\underline{0}, \Sigma + 2 \sum_{h=1}^{\infty} \Sigma_{0h}), \tag{2.1.21}$$

where Σ is the variance-covariance matrix of X and

$$\Sigma_{0h} = (cov(X_0(i), X_h(j)))_{i,j=1,\ldots,k}.$$

Proof. Since (X_i) is strongly mixing with rate function α_h, the measurable function $f(X_i) = a'X_i$ is strongly mixing with the same rate function. An application of Theorem 2.1.16 yields that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a'X_i \xrightarrow{d} N(0, var(a'X) + 2 \sum_{h=1}^{\infty} cov(a'X_0, a'X_h)),$$

and the asymptotic variance is finite. Notice that

$$var(a'X) = a'\Sigma a',$$

and

$$cov(a'X_0, a'X_h) = a'\Sigma_{0h} a.$$

By the Cramér-Wold device (Theorem 2.1.17), this is equivalent to relation (2.1.21). □

2.2 Financial time series and their stylized facts

One branch of financial mathematics is pricing theory. It is the part of financial mathematics which is related to the pricing of derivatives such as options. The Black-Scholes option pricing formula stands as a synonym for this theory. It has been developed since 1973 when two fundamental papers of Black, Scholes [3] and Merton [26] appeared. The basic version of the much quoted Black-Scholes model for log prices, $p_t = \log(P_t)$, is given by the simple stochastic differential equation,

$$dp_t = \mu dt + \sigma dB_t,$$

with t denoting time and B a Brownian motion. Furthermore, μ is the drift coefficient and σ is commonly referred to as the volatility which is essential in option pricing for example. This model has two basic assumptions (which are often known to fail):

- The volatility σ is constant over time.

- p_t is normally distributed.

Brownian motion and Itô stochastic integral, Girsanov transformation and change of measure are basic notions in this framework. A deep knowledge of the theory of stochastic processes is the basis for anybody who wants to conduct serious research in this area.

Another approach to modeling financial phenomena is via time series. Time series are discrete-time processes, and the link to continuous-time processes is not obvious and in general difficult to establish, i.e., the embedding of a discrete-time process in a continuous-time one is by no means an easy matter. The aims of time series analysis are different from those of pricing. Whereas the latter requires a continuous-time framework in order to make martingale and stochastic integration techniques applicable, time series analysis has always been directed towards the understanding of the mechanism that drives a given series of data with the aim of possibly predicting future values in the series.

This does not mean that stochastic differential equations, which are commonly used to model price movements for derivative pricing, do not describe a certain physical mechanism of the price evolution. This approach, however, does not primarily aim at the most realistic model for prices. Its basic goal is to get a reasonable model that is mathematically tractable and can be understood or interpreted by financial practitioners. Thus, financial time series analysis focuses on the "truth" behind the data meaning that one is interested in finding physical models that explain, at least to some extent, the empirically observed features of real-life data.

Because there are many financial data, it might be difficult to say anything about some common properties. Surprisingly, a large variety of financial data which we denote consistently by $P_t, t = 0, 1, 2, \ldots,$ (t can be minutes, hours, days, etc.) exhibits similar properties after the transformation

$$X_t = \log P_t - \log P_{t-1} = \log(1 + \frac{P_t - P_{t-1}}{P_{t-1}}),$$

at least if one focuses on share prices (of Microsoft, say), stock indices (DAX, Nikkei, Dow Jones, etc.) or foreign exchange rates (such as USD/JPY, USD/DEM, or JPY/DEM). In fact, these similar properties depend on the time scale chosen. Depending on whether the time unit is a second, half an hour, or a day, a month, or a year, qualitative differences in the time series can be expected and different models need to be introduced. E.g., if the time scale is too small P_t lives on a grid and varies little; one often observes that P_t does not change over longer (relative to the time unit) periods. This would imply that the distribution of X_t has an atom at zero. This would be an unacceptable assumption if X_t was calculated on a daily basis, in which case one often requires that X_t has a density. Since we do not want to get into too large or small time scales; in what follows we think of t in units of hours, days, or weeks.

The resulting (X_t) is the time series of log-returns which, by a Taylor series argument, is close to the relative returns series $\frac{P_t - P_{t-1}}{P_t}$. The relative returns give some more intuition on the log-returns (the relative returns are very small and therefore almost indistinguishable from the log-returns), i.e., they describe the relative change over time of the price process. In what follows, we often refer, for short, to returns instead of log-returns.

The log-differences X_t have the advantage that they are free of any unit, therefore, comparable among each other. Moreover, there is an important mathematical issue as well: it is believed that the time series (X_t) can be modeled by a stationary (in the strict or wide senses) stochastic process, i.e., this transformation yields one realization of a stationary process. In turn, stationarity is a basic assumption for any kind of time series analysis. There are several references which discuss the stylized facts of financial time series, for example Mikosch [27].

2.2.1 Distribution and tails

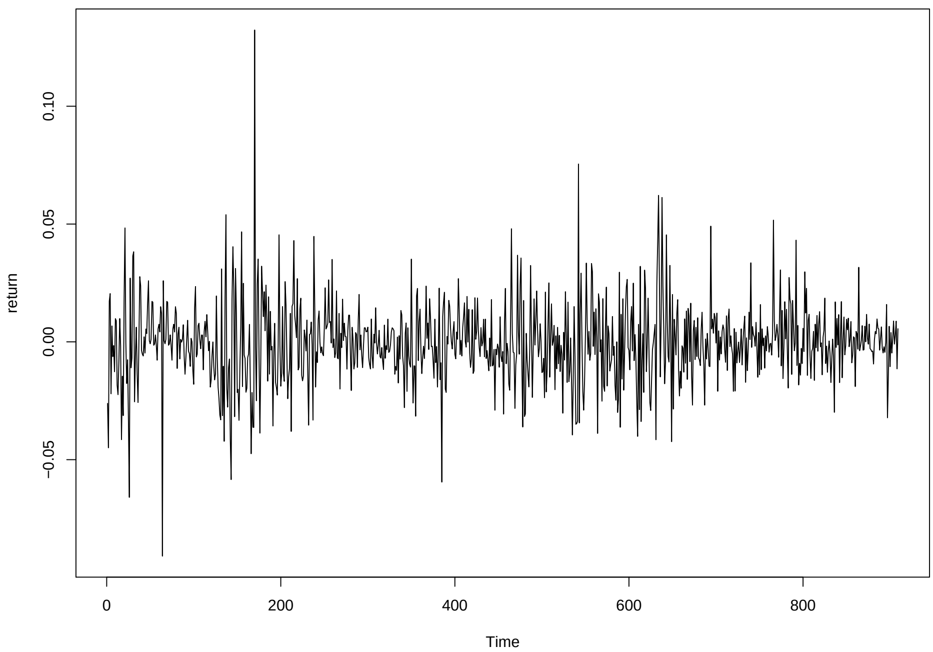

Figure 2.1: The Nikkei daily closing log-returns over a period of 4 years

Samples of returns, $X_1, \ldots, X_n$ have the following stylized facts in common:

- The sample mean of the data is close to zero; the sample variance is of the order 10^{-4} or smaller. This is due to the fact that price changes are in general very small; a daily change of 1%, 2% or more is very unusual.

- A density plot of the data, see Figure 2.2, shows that the distribution of the data is roughly symmetric in its center, sharply peaked around zero with heavy tails on both sides. The shape of

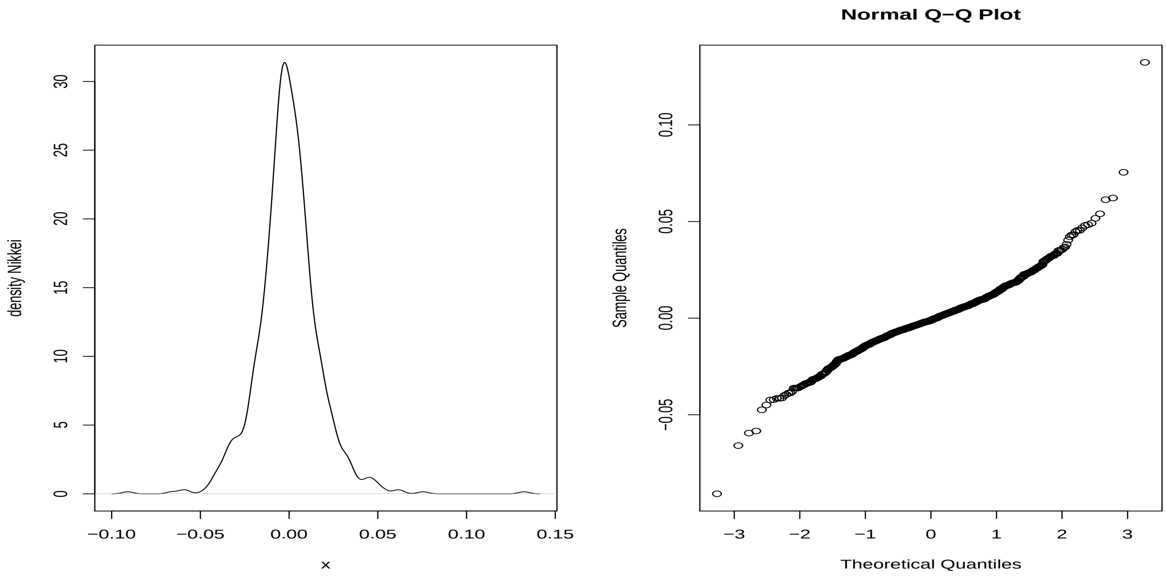

Figure 2.2: Left: Density plot of the Nikkei data. Right: QQ-plot of the Nikkei data against the normal distribution with mean and variance estimated from the Nikkei data.

the density and the QQ-plot of the data against the normal distribution in Figure 2.2 indicate that the normal distribution is not a perfect distribution to fit the returns. This is in contrast to the assumptions in the Black-Scholes model, which is most widely used for modeling stock prices.

- The log-returns have heavy-tailed marginal distribution. This is clear from the very large and very small values in the density plot and the shape of the QQ-plot (which curves up at the right and curves down at the left). Examples of distributions more suitable for fitting return data are e.g. the Pareto and the t-distribution. Both distributions are regularly varying in the sense defined in Definition 3.1.4. This means that their tails are power-like, hence very heavy-tailed.

2.2.2 Dependence and autocorrelations

Using the ACF and its estimator, the sample ACF, the following dependence properties of log-return series (X_t) are commonly observed, see Figure 2.3 for an illustration. They are observed for many financial

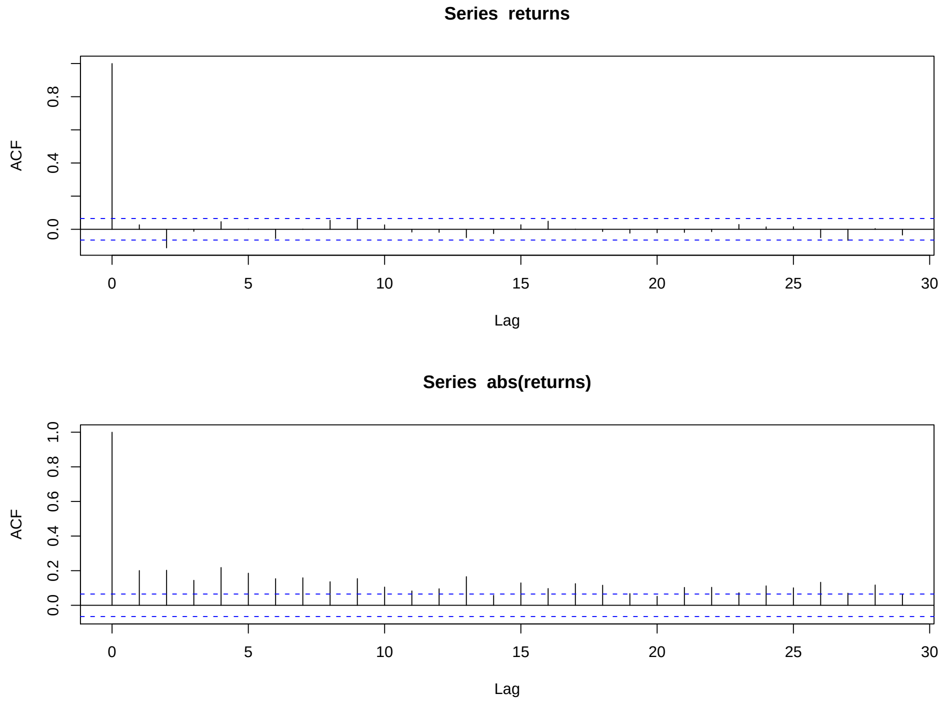

Figure 2.3: Sample ACFs for the log-returns (top) and absolute log-returns (bottom) of the Nikkei composite stock index. The confidence band is set as the 95 % asymptotic confidence interval corresponding to the sample ACF of iid Gaussian noise.

time series.

- The sample ACF is negligible at all lags. An exception can be the first lag.
- The sample ACF for the absolute values $|X_t|$ and also for the squares X_t^2 are different from zero for a large number of lags and stay almost constant and positive for large lags.
- Occurrence of extremely large and small X_t's clustered at certain instants of time, caused by turbulences in the market due to financial crashes, political decisions, war, etc.

The slow decay of the sample ACF for the absolute log-returns is typical for longer time series. The sample ACF is not negligible even for large lags, this is often interpreted as long memory of the absolute returns.

Autocorrelations are not good tools for explaining large and small values in a time series. Indeed, covariance and correlations are moments, hence they are integrated characteristics of the distribution of the underlying time series; the contribution of the probabilities in the tails of the distributions is averaged out and disappears.

2.2.3 Dependence and extremes

The dependence of extremal return values is obvious if one looks, for example, at pairs $|X_t|, |X_{t+1}|$ exceeding a high threshold, see Figures 2.4 and 2.6. For comparison, we include the graph of the joint exceedances of the same threshold for the successive values $|X_t|, |X_{t+1}|$ in an iid sequences (X_t) with a student $t-$distribution with four degrees of freedom which often fits returns nicely. It is obvious that joint pairwise exceedances of a high threshold occur in clumps for log-return data due the dependence of the extreme values. We refer to dependence in the tails. For an iid sequence, exceedances of high positive or low negative thresholds (relatively to the size of the data) occur separated over time, roughly according to a homogeneous Poisson process. For returns, this is not true. This leads to the next stylized fact:

- The large and small values in the log-return sample occur in clusters. There is dependence in the tails.

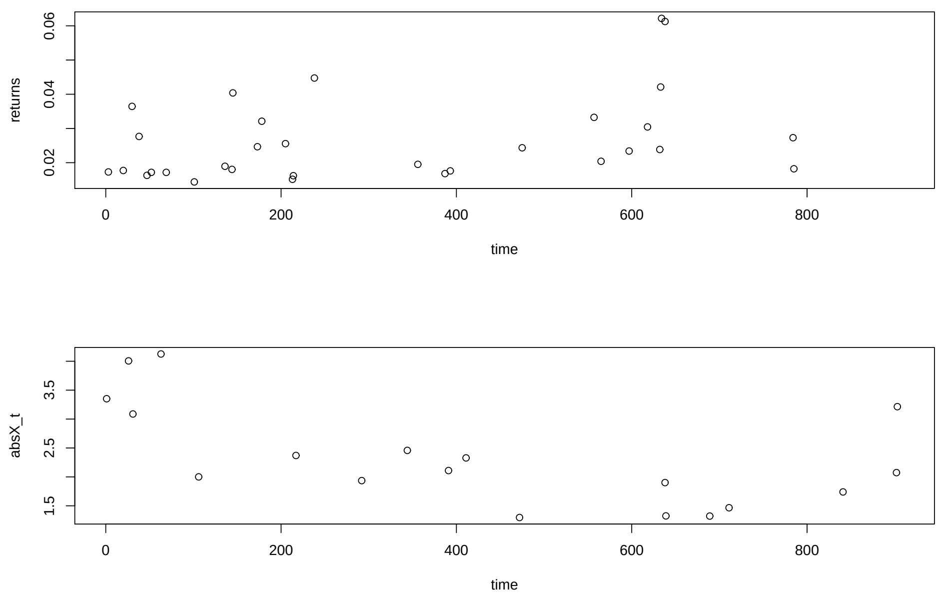

Figure 2.4: Top: Absolute returns $|X_t|$ of the Nikkei composite stock index series for which both $|X_t|$ and $|X_{t+1}|$ exceed the 85% quantile of the data. Bottom: The same kind of plot for an iid sequence from a student distribution with 4 degrees of freedom. In the former case pairwise exceedances occur in clusters, in the latter case exceedances appear uniformly scattered over time.

Since autocorrelations are not appropriate for describing the dependence of large and small X_t-values, further tools have been considered. One of them we want to explain now. For an iid sequences (X_t) we

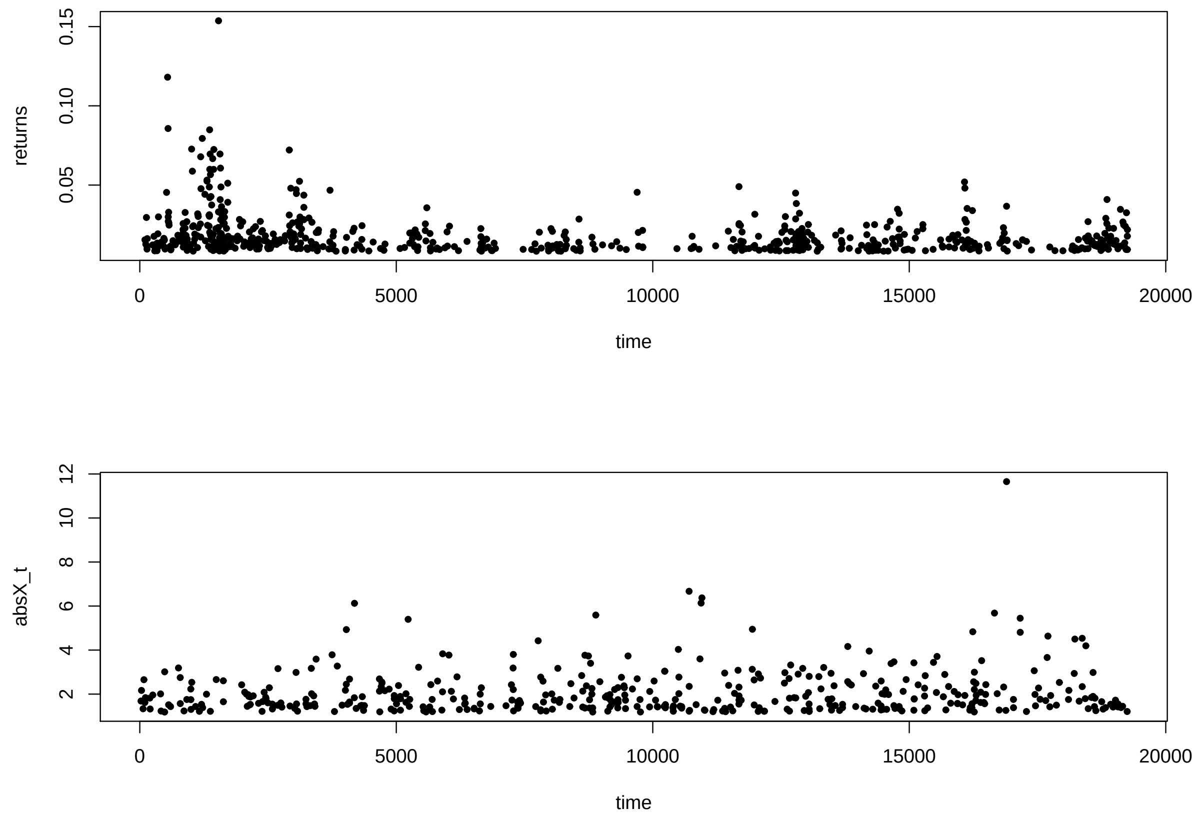

Figure 2.5: Top: Absolute returns $|X_t|$ of the S&P500 composite stock index series for which both $|X_t|$ and $|X_{t+1}|$ exceed the 85% quantile of the data. Bottom: The same kind of plot for an iid sequence from a student distribution with 4 degrees of freedom. In the former case pairwise exceedances occur in clusters, in the latter case exceedances appear uniformly scattered over time.

know that

$$P(M_n \le x) = [P(X \le x)]^n, \quad n = 1, 2, \ldots,$$

where

$$M_n = \max(X_1, \ldots, X_n).$$

For large classes of strictly stationary sequences (X_t) one can show the existence of a number $\theta \in [0, 1]$ such that

$$P(M_n \le x_n) = [P(X \le x_n)]^\theta + o(1),$$

where (x_n) is a suitable sequence converging to the right endpoint of the distribution of X. This number θ is the extremal index of (X_t). It can be estimated from the data by statistical methods, see [11], Section

8.1. By its definition, θ describes the reciprocal of the expected cluster size in a sample $X_1, \ldots, X_n$ above high thresholds.

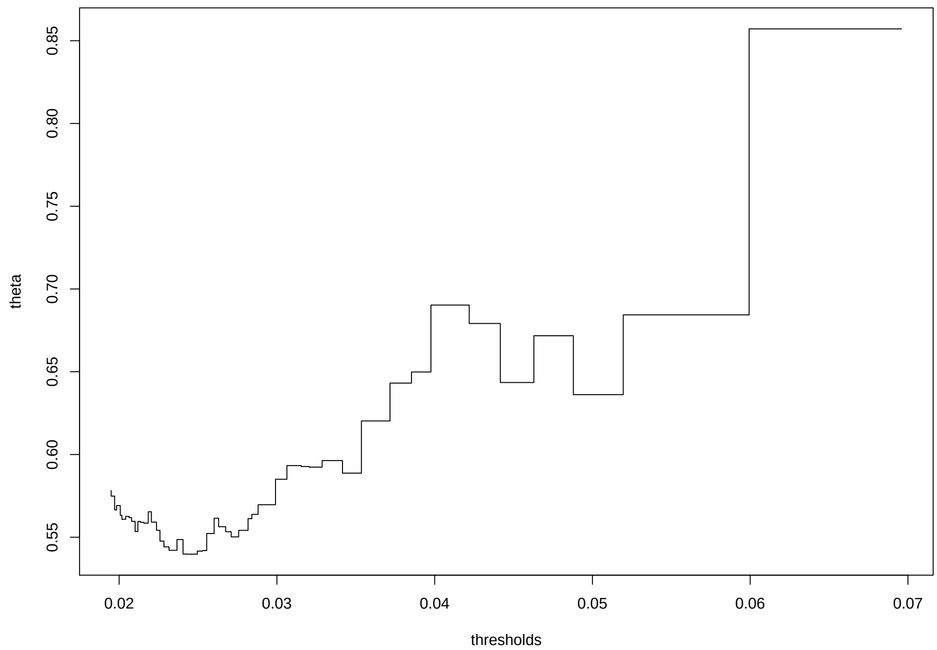

Figure 2.6: Point estimation of the extremal index for the Nikkei data. The estimators are based on the upper order statistics exceeding the threshold u. The smallest u is the 97% quantile of the data. The estimator estimates θ in a u-region, where the plot does not change much (between 0.02 and 0.03) resulting in an extremal index of about 0.55. If u is too high (above 0.03 say), the estimator is based on too few order statistics and not reliable.

2.3 Some standard financial time series models

Time series can be roughly divided into two groups, linear and nonlinear time series. Examples of linear time series models are autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) processes. ARMA models and their variations play an active role in analyzing time series data due to their simplicity, feasibility, and flexibility [15, p. 14]. On the other hand, as regards the properties of financial time series, the ARMA family is not suitable for fitting the data. Beyond the linear domain, there are infinitely many nonlinear forms to be explored.

Most models for return data used in practice are of a multiplicative form

$$X_t = \mu + \sigma_t Z_t, \quad t \in \mathbb{Z}, \tag{2.3.1}$$

where (Z_t) is an iid noise or innovation sequence, (σ_t) is a stochastic process with nonnegative values such that σ_t and Z_t are independent for fixed t. The volatility process (σ_t) and the return process (X_t) are usually assumed to be strictly stationary. Sometimes Z_t is assumed to be symmetric. We will often assume that $EZ = 0$ and $var(Z) = 1$. We will also assume that μ can be estimated from the data and therefore it will be convenient to assume $\mu = 0$ as well. If Z_t is iid symmetric, the direction of price changes is modeled by the sign of Z_t, independent of the order of magnitude of this change, which is directed by the volatility σ_t. This is in agreement with the empirical observation that it is difficult to predict the sign of price changes. Since σ_t and Z_t are independent, σ_t^2 is then the conditional variance of X_t given σ_t. Most models assume that σ_t is a function of $X_{t-1}, X_{t-2}, \ldots$ and $\sigma_{t-1}, \sigma_{t-2}, \ldots$. We consider two of the most popular ones.

2.3.1 The ARCH family

One of the successful examples of nonlinear time series models which has a multiplicative form is the ARCH process (autoregressive process with conditional heteroscedasticity) and its numerous modifications.

Definition 2.3.1. (ARCH(p) process)
An ARCH process of order p is defined as

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = a_0 + b_1 X_{t-1}^2 + \cdots + b_p X_{t-p}^2, \quad t \in \mathbb{Z} \tag{2.3.2}$$

where $a_0 > 0, b_j \geq 0, b_p > 0$ *and* Z_t *are iid with* $EZ = 0$ *and* $var(Z) = 1$*.*

The ARCH model was introduced by Engle [12] in 1982 to model the varying conditional variance or volatility of a return time series. It is one of the properties of financial time series that the larger values of the past lead to instability (i.e., larger variances at the present), which is termed (conditional) heteroscedasticity.

Bollerslev [4] in 1986 introduced the generalized autoregressive conditionally heteroscedastic (GARCH) process of order (p, q) by replacing the second equation in (2.3.2) with

$$\sigma_t^2 = a_0 + a_1\sigma_{t-1}^2 + \cdots + a_q\sigma_{t-q}^2 + b_1 X_{t-1}^2 + \cdots + b_p X_{t-p}^2, \quad t \in \mathbb{Z}, \tag{2.3.3}$$

where $a_j \geq 0$, $b_j \geq 0$, $a_0 > 0, a_q > 0$ and $b_p > 0$.

There exists a constantly increasing number of references to ARCH-GARCH and related process. Relevant references on different issues are the following ones:

- **On strict stationarity** Bougerol and Picard [5], Nelson [32] give conditions for strict stationarity. Those are in general not easily established. They depend on the coefficients a_i, b_j and on the distribution of Z.

- **On the tails** GARCH models, under general conditions, have power law tails. This is explored in [11], Section 8.4 for ARCH(1), in [29] for GARCH(1,1) and for general GARCH(p, q) in [1].

- **On extremes** Power law tails of an iid sequence (X_t) imply that the distribution of X is in the domain of attraction of the Fréchet distribution, see [11], Section 3. This remains true for GARCH process. See [29] and [1]. In contrast to iid sequences, GARCH extremes occur in clusters. See also Section 3.2.1 below.

- **On mixing** The GARCH process has nice mixing properties. Given that Z_t has a positive density in some neighborhood of the origin (such as the normal or t-densities), (X_t) is β-mixing with geometric rate. This was proved by Mokkadem [30], see also [10, p. 108].

- **On estimation** An advanced estimation theory for GARCH and related processes can be found in [39]. Estimation of the GARCH parameters is typically based on the Gaussian quasi-maximum likelihood procedure. This means that one maximizes the likelihood function of the sample $(X_1, \ldots, X_n)$ under the assumption that the Z_t's are iid $N(0, 1)$. The estimators are a.s. consistent and asymptotically normal under very general assumptions on the distribution of Z_t.

2.3.2 The stochastic volatility model

The stochastic volatility model is an alternative model to the celebrated GARCH model. The GARCH probabilistic properties (existence of stationary solution, dependence structure, tails, etc.) are by no means easy to derive and not in all cases well understood. We will consider another multiplicative model, the stochastic volatility process.

Definition 2.3.2. (Stochastic volatility model)
The time series $(X_t)_{t\in\mathbb{Z}}$ is said to be a stochastic volatility process if it satisfies the equations

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z},$$

where the volatility sequence $(\sigma_t)_{t\in\mathbb{Z}}$ is a strictly stationary sequence of positive random variables and $(Z_t)_{t\in\mathbb{Z}}$ is an iid noise sequence. Moreover, $(\sigma_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$ are mutually independent.

In the stochastic volatility model, the volatility sequence (σ_t) and the noise sequence (Z_t) are independent. This is in contrast to the GARCH model where σ_t is a measurable function of the past values of the multiplicative noise $(Z_s)_{s<t}$ and, for fixed t, σ_t and $(Z_s)_{s\geq t}$ are independent, i.e., in contrast to GARCH processes, there is no feedback between the noise (Z_t) and the volatility process (σ_t),i.e., there are two independent sources of randomness. For this reason, the stochastic volatility model is sometimes considered as unnatural. However, the probabilistic properties of a stochastic volatility process are much better understood than the properties of GARCH processes.

Usually (σ_t) is given by a parametric model such as a Gaussian ARMA process for $(\log\sigma_t)$. By first squaring X_t and then taking logarithms, one can see that the ARMA process $2\log\sigma_t$ gets perturbed by the extra noise $2\log|Z_t|$ which makes estimation more complicated because it is impossible to give an explicit expression for the likelihood function. We still can use some estimation methods like GMM and quasi-MLE. Often one needs to resort to simulation based methods to calculate efficient estimates.

2.4 Properties of the stochastic volatility model

2.4.1 Some elementary properties

In what follows, we will assume that the following conditions are satisfied. It is common use to assume that $EZ = 0$ and $EZ^2 = 1$. Then, if $E\sigma < \infty$, EX equals zero. This is in agreement with return data of price series such as foreign exchange rates, share prices and stock indices, see page 20.

Moreover, if $\mathrm{var}(Z)$ and $\mathrm{var}(\sigma)$ are finite, $\mathrm{var}(X) < \infty$ and, by the Cauchy-Schwarz inequality,

$$|\mathrm{cov}(X_t, X_{t+h})| \leq \sqrt{\mathrm{var}(X_t)\mathrm{var}(X_{t+h})}\,.$$

In particular, we have

$$\mathrm{cov}(X_t, X_{t+h}) = E(X_t X_{t+h}) = E(Z_t)E(Z_{t+h})E(\sigma_t\sigma_{t+h}) = 0$$

for all $t, h \in \mathbb{Z}$. This shows that (X_t) is a white noise sequence. This is in agreement with the stylized facts in Section 2.2.2.

2.4.2 Strict stationarity

The assumption of stationarity of the time series (X_t) is basic for the statistical analysis of the data.

Lemma 2.4.1. (The stochastic volatility sequence is a strictly stationary ergodic sequence)
Under the conditions of Definition 2.3.2, (X_t) is strictly stationary. Moreover, it is ergodic if (σ_t) is ergodic.

Proof. Condition on (Z_n). By strict stationarity of (σ_n), the sequences (σ_n) and (σ_{n+h}) have the same distribution for any $h \in \mathbb{Z}$. Hence $(Z_n\sigma_{n+h}) \stackrel{d}{=} (Z_n\sigma_n)$. Now condition $(Z_n\sigma_{n+h})$ on (σ_{n+h}). The iid property of (Z_n) yields that $(Z_n) \stackrel{d}{=} (Z_{n+h})$. Hence $(Z_n\sigma_{n+h})$ has the same distribution as $(Z_{n+h}\sigma_{n+h})$. Finally, $(X_n) \stackrel{d}{=} (X_{n+h})$. The ergodicity of (X_n) follows from the fact that both (σ_t) and (Z_t) are ergodic and mutually independent. □

2.4.3 Strong mixing

Lemma 2.4.2. (The stochastic volatility sequence is a strongly mixing sequence)
Consider the stochastic volatility model with a strongly mixing sequence (σ_n) with rate function $(\alpha_h(\sigma))$. Then the sequence (X_t) is strongly mixing with rate function $\alpha_h \leq 4\alpha_h(\sigma)$.

Proof. Let $\mathfrak{F}_a^b = \sigma(X_t, a \leq t \leq b)$ and choose $A \in \mathfrak{F}_{-\infty}^0, B \in \mathfrak{F}_t^\infty$. Then

$$\begin{aligned} |P(A \cap B) - P(A)P(B)| &= P((X_s)_{s\leq 0} \in C, (X_s)_{s\geq t} \in D) \\ &\quad - P((X_s)_{s\leq 0} \in C)P((X_s)_{s\geq t} \in D), \end{aligned}$$

where C, D are suitable Borel sets in $\mathbb{R}^\infty$. Conditioning on (σ_s) yields

$$\begin{aligned} P(A \cap B) &= E[P((X_s)_{s\leq 0} \in C, (X_s)_{s\geq t} \in D)|(\sigma_s)] \\ &= E[P(((X_s)_{s\leq 0} \in C)|\sigma_s)P(((X_s)_{s\geq t} \in D)|(\sigma_s))] \end{aligned}$$

Since (σ_t) and (Z_t) are independent we can write

$$f(\ldots, \sigma_{-1}, \sigma_0) = P((\ldots, X_{-1}, X_0) \in C|\sigma_s, s \leq 0),$$

$$g(\sigma_t, \sigma_{t+1}, \ldots) = P((X_t, X_{t+1}, \ldots) \in D|\sigma_s, s \geq t),$$

and then

$$\begin{aligned} |P(A \cap B) - P(A)P(B)| &= |E(f(\ldots, \sigma_{-1}, \sigma_0)g(\sigma_t, \sigma_{t+1}, \ldots)) \\ &\quad - E(f(\ldots, \sigma_{-1}, \sigma_0))E(g(\sigma_t, \sigma_{t+1}, \ldots))|. \end{aligned}$$

Notice that $f(\ldots, \sigma_{-1}, \sigma_0)$ and $g(\sigma_t, \sigma_{t+1}, \ldots)$ are less than 1. Standard results about strong mixing show that the right-hand side in the previous equation is bounded by $4\alpha_h(\sigma)$ [15, Proposition 2.5, p. 71–72]. □

A common way of constructing a positive strictly stationary volatility sequence (σ_t) is to assume a particular form of the strictly stationary log-volatility sequence $Y_t = \log \sigma_t, t \in \mathbb{Z}$. In the literature, it is often assumed that (Y_t) is a linear process

$$Y_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}, \quad t \in \mathbb{Z}, \tag{2.4.1}$$

where (ψ_t) is a sequence of deterministic coefficients with $\psi_0 = 1$ and $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ and (η_t) is an iid mean zero and finite variance sequence of random variables. By Kolmogorov's three series theorem [21, p. 185] the latter conditions ensure that the infinite series (2.4.1) converges a.s. As a Bernoulli shift of the iid sequence (η_t), the process (σ_t) is strictly stationary ergodic (see Section 2.1.3), and (X_t) inherits ergodicity from ergodicity of (σ_t), the iid property of (Z_t) and the independence of (σ_t) and (Z_t). Moreover, if (Y_t) is strongly mixing with rate function (α_h) then (σ_t) has the same rate function and by Lemma 2.4.2, (X_t) has rate function $\alpha_h \le 4\alpha_h(\sigma)$.

2.4.4 Martingale properties

We consider the filtration $\mathbb{G}_t = \sigma(Z_s, \eta_s, s \le t)$ in the model given by Equation (2.4.1). Then (X_t) is adapted to $(\mathbb{G}_t)$ and

$$E(X_t|\mathbb{G}_{t-1}) = e^{\sum_{i=1}^{\infty} \psi_i \eta_{t-i}} E(Z_t e^{\eta_t}) = e^{\sum_{i=1}^{\infty} \psi_i \eta_{t-i}} EZEe^{\eta_t} = 0 \quad a.s., \tag{2.4.2}$$

provided that $E\sigma < \infty$. Then (X_t) constitutes a centered finite variance strictly stationary ergodic martingale difference sequence.

Lemma 2.4.3. *The sequence $(X_t X_{t+h})$ is a mean zero finite variance strictly stationary ergodic martingale difference sequence with respect to $(\mathbb{G}_{t+h})$. If $var(\sigma^2) < \infty$ it is also stationary.*

Proof. For $h > 0$, (X_t, X_{t+h}) is adapted to the filtration $(\mathbb{G}_{t+h})$ and by (2.4.2)

$$E(X_t X_{t+h}|\mathbb{G}_{t+h-1}) = X_t E(X_{t+h}|\mathbb{G}_{t+h-1}) = 0 \quad a.s.$$

Hence $(X_t X_{t+h})$ is a strictly stationary ergodic mean zero martingale difference sequence. In addition, since $EZ^2 = 1$ and if $var(\sigma^2) < \infty$, then

$$var(X_t X_{t+h}) = E({\sigma_0}^2 {\sigma_h}^2) < \infty.$$

Therefore $(X_t X_{t+h})$ is a mean zero finite variance strictly stationary ergodic martingale difference sequence. □

Lemma 2.4.4. *Assume $p \ge 1$. If $E\sigma^p < \infty$ and $E|Z|^p < \infty$, the sequence $(\sigma_t^p(|Z_t|^p - E|Z_t|^p))$ is a strictly stationary ergodic mean zero martingale difference sequence with respect to $(\mathbb{G}_t)$. If $E\sigma^{2p} < \infty$ and $E|Z|^{2p} < \infty$ it is also stationary.*

Proof. The sequence $(\sigma_t^p(|Z_t|^p - E|Z_t|^p))$ is adapted to the filtration $(\mathbb{G}_t)$. Moreover,

$$\begin{aligned} E(\sigma_t^p(|Z_t|^p - E|Z_t|^p)|\mathbb{G}_{t-1}) &= e^{p\sum_{i=1}^\infty \psi_i\eta_{t-i}} E((e^{p\eta_t}(|Z_t|^p - E|Z|^p))|\mathbb{G}_{t-1}) \\ &= e^{p\sum_{i=1}^\infty \psi_i\eta_{t-i}} Ee^{p\eta_t} E(|Z|^p - E|Z|^p) = 0. \end{aligned}$$

Thus $(\sigma_t^p(|Z_t|^p - E|Z_t|^p))$ is a mean zero strictly stationary ergodic martingale difference sequence with respect to $(\mathbb{G}_t)$. If $E\sigma^{2p} < \infty$ and $E|Z|^{2p} < \infty$,

$$var(\sigma^p(|Z_t|^p - E|Z_t|^p)) = E\sigma^{2p}(E|Z_t|^{2p} - (E|Z|^p)^2) = E\sigma^{2p} var(|Z_t|^p) < \infty.$$

□

Another sequence of interest in the stochastic volatility model is

$$(\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p))), \quad h > 0.$$

The last sequence is adapted to the filtration $(\mathbb{G}_{t+h})$.

Lemma 2.4.5. *Assume $p \geq 1$. For every $h > 0$ the sequence $(\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p)))$ is a strictly stationary ergodic martingale difference with respect to $(\mathbb{G}_{t+h})$ if $E\sigma^{2p} < \infty$ and $E|Z|^p < \infty$. It is stationary if $E|Z|^{2p} < \infty$ and $E\sigma^{4p} < \infty$.*

Proof. For $h > 0$, the sequence $(\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p)))$ is adapted to the filtration $(\mathbb{G}_{t+h})$. Therefore

$$\begin{aligned} E((\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p))|\mathbb{G}_{t+h-1})) &= \\ (|Z_t|^p - E|Z|^p)\sigma_t^p e^{p\sum_{i=1}^\infty \psi_i\eta_{t+h-i}} E(e^{p\eta_{t+h}}(|Z_{t+h}|^p - E|Z_t|^p)) &= \\ (|Z_t|^p - E|Z|^p)\sigma_t^p e^{p\sum_{i=1}^\infty \psi_i\eta_{t+h-i}} Ee^{p\eta_{t+h}}(E|Z_{t+h}|^p - E|Z_t|^p) &= 0. \end{aligned}$$

Since $E\sigma^{2p} < \infty$ and $E|Z|^p < \infty$, the sequence $(\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p)))$ constitutes a strictly stationary ergodic martingale difference sequence with respect to $(\mathbb{G}_{t+h})$. Moreover, if $E\sigma^{4p} < \infty$ and $E|Z|^{2p} < \infty$,

$$var(\sigma_t^p\sigma_{t+h}^p((|Z_t|^p - E|Z|^p)(|Z_{t+h}|^p - E|Z|^p))) = E(\sigma_t^{2p}\sigma_{t+h}^{2p})(var(|Z|^p))^2 < \infty.$$

□

2.4.5 Moments of the stochastic volatility model

Throughout this section, we assume (X_t) is a stochastic volatility process with specification given in Section 2.4.3, in particular $EZ = 0$ and $EZ^2 = 1$, and (σ_t) is defined through equation (2.4.1) where (η_t) is an iid sequence with $E(\eta) = 0, var(\eta) = E(\eta_t^2) = \tau^2 < \infty$ and the moment generating function of η, $m_\eta(s) = Ee^{s\eta}$, is finite for all $s \in \mathbb{R}$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ and $\psi_0 = 1$.

Remark 2.4.6. The fact that (X_t) is a white noise sequence agrees with real-life return data. However, this observation is not very informative. Therefore it has become common in financial time series analysis to study the ACVF and ACF of the absolute values, squares and other powers of absolute return data as well. In contrast to the GARCH process, in the stochastic volatility model one can exploit the independence between (σ_t) and (Z_t) in order to get explicit formulas for $\gamma_{|X|^p}$.

Lemma 2.4.7. *Assume $p > 0$. Then the following relations hold.*

$$E(\sigma^p) = \prod_{i=0}^{\infty} m_\eta(p\psi_i), \tag{2.4.3}$$

$$var(\sigma^p) = \prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2. \tag{2.4.4}$$

The ACVF of (σ_t^p) at lag $h > 0$ is given by

$$\gamma_{\sigma^p}(h) = \prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2 \tag{2.4.5}$$

and the ACF of (σ_t^p) at lag $h > 0$ is given by

$$\rho_{\sigma^p}(h) = \frac{\prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2}{\prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2}. \tag{2.4.6}$$

Moreover if η has a Gaussian distribution with mean zero and variance $\tau^2 > 0$ then

$$E(\sigma^p) = e^{\frac{p^2\tau^2}{2}\sum_{i=0}^{\infty}\psi_i^2}, \tag{2.4.7}$$

$$var(\sigma^p) = e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}(e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2} - 1), \tag{2.4.8}$$

$$\gamma_{\sigma^p}(h) = e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}\left(e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}} - 1\right), \tag{2.4.9}$$

$$\rho_{\sigma^p}(h) = \frac{e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}} - 1}{e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2} - 1} = \frac{e^{p^2\gamma_Y(h)} - 1}{e^{p^2\gamma_Y(0)} - 1}. \tag{2.4.10}$$

Proof. The volatility sequence (σ_t) in the stochastic volatility model is given by $\sigma_t = e^{\sum_{i=0}^{\infty}\psi_i\eta_{t-i}}$. Hence by independence of the η_i's and since $m_\eta(s)$ exists for all $s \in \mathbb{R}$,

$$\begin{aligned}
E(\sigma^p) &= Ee^{p\sum_{i=0}^{\infty}\psi_i\eta_{-i}} = \prod_{i=0}^{\infty} m_\eta(p\psi_i),\\
var(\sigma^p) &= E(\sigma^{2p} - (E\sigma^p)^2) = \prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2,\\
E(\sigma_t^p\sigma_{t+h}^p) &= E(\sigma_0^p\sigma_h^p) = Ee^{p(\sum_{i=0}^{\infty}\psi_i\eta_{-i}+\sum_{i=0}^{\infty}\psi_i\eta_{h-i})}\\
&= Ee^{p(\sum_{i=0}^{\infty}(\psi_i+\psi_{i+h})\eta_{-i}+\sum_{j=-h}^{-1}\psi_{j+h}\eta_{-j})}\\
&= \prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(p\psi_i), \quad h>0.
\end{aligned}$$

From the above results, formulas (2.4.5) and (2.4.6) are straightforward. For the special case when η has a Gaussian distribution $N(0,\tau^2), \tau^2>0$, the moment generating function is given by $m_\eta(s) = e^{\frac{1}{2}\tau^2 s^2}$. In this case (2.4.7) and (2.4.8) are straightforward from (2.4.3) and (2.4.4). Direct calculation with (2.4.5) yields

$$\begin{aligned}
E(\sigma_0^p\sigma_h^p) &= e^{\frac{p^2\tau^2}{2}\left(\sum_{i=0}^{\infty}(\psi_i+\psi_{i+h})^2+\sum_{i=0}^{h-1}\psi_i^2\right)}\\
&= e^{\frac{p^2\tau^2}{2}\left(\sum_{i=0}^{\infty}\psi_i^2+\sum_{i=0}^{\infty}\psi_{i+h}^2+2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}+\sum_{i=0}^{h-1}\psi_i^2\right)}\\
&= e^{p^2\tau^2\sum_{i=0}^{\infty}\left(\psi_i^2+\psi_i\psi_{i+h}\right)},\\
\gamma_{\sigma^p}(h) &= cov(\sigma_t^p\sigma_{t+h}^p) = e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}\left(e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}}-1\right),\\
\rho_{\sigma^p}(h) &= \frac{e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}}-1}{e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}-1}.
\end{aligned}$$

□

Remark 2.4.8. Notice that when η comes from a Gaussian distribution $N(0,\tau^2)$, (Y_t) presents a linear process as introduced in Example 2.1.6. Thus (Y_t) satisfies

$$\begin{aligned}
\gamma_Y(0) &= \tau^2\sum_{j=0}^{\infty}\psi_j^2,\\
\gamma_Y(h) &= \tau^2\sum_{j=0}^{\infty}\psi_j\psi_{j+h}, \quad h>0,\\
\rho_Y(h) &= \frac{\sum_{j=0}^{\infty}\psi_j\psi_{j+h}}{\sum_{j=0}^{\infty}\psi_j^2}, \quad h>0.
\end{aligned}$$

For the ACF of (σ_t^p) we then have

$$\gamma_{\sigma^p}(h) = e^{p^2\gamma_Y(0)}(e^{p^2\gamma_Y(h)}-1), \quad h>0.$$

Since $\gamma_Y(h) \to 0$ as $h \to \infty$ a Taylor series expansion yields

$$\gamma_{\sigma^p}(h) \sim e^{p^2\gamma_Y(0)} p^2 \gamma_Y(h), \quad h \to \infty. \tag{2.4.11}$$

In addition,

$$\rho_{\sigma^p}(h) = \frac{e^{p^2\gamma_Y(h)} - 1}{e^{p^2\gamma_Y(0)} - 1} \sim \frac{p^2}{e^{p^2\gamma_Y(0)} - 1}\gamma_Y(h). \tag{2.4.12}$$

In particular, if (Y_t) is strongly mixing with geometric rate, $\gamma_Y(h) \to 0$ as $h \to \infty$ exponentially fast, hence $\rho_{\sigma^p}(h) \to 0$ exponentially fast too. This is in agreement with the theory of strong mixing, see Lemma 2.1.13.

Lemma 2.4.9. *Assume $p > 0$. The ACVF of $(\sigma_s^p \sigma_{s+h}^p)$ at lag $t > 0$ is given by*

- *if $h \le t$*

$$\begin{aligned}\gamma_{\sigma_0^p\sigma_h^p}(t) &= \prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{h+i} + \psi_{t+i} + \psi_{t+h+i})) \\ &\cdot \prod_{i=0}^{h-1} m_\eta(p(\psi_i + \psi_{i+t-h} + \psi_{i+t}) \prod_{i=0}^{t-h-1} m_\eta(p(\psi_i + \psi_{h+i})) \\ &\cdot \prod_{i=0}^{h-1} m_\eta(p\psi_i) - \left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) . \prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2,\end{aligned}$$

- *if $h > t$*

$$\begin{aligned}\gamma_{\sigma_0^p\sigma_h^p}(t) &= \prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{h+i} + \psi_{t+i} + \psi_{t+h+i})) \\ &\cdot \prod_{i=0}^{t-1} m_\eta(p(\psi_i + \psi_{i+h-t} + \psi_{i+h}) \prod_{i=0}^{h-t-1} m_\eta(p(\psi_i + \psi_{t+i})) \\ &\cdot \prod_{i=0}^{t-1} m_\eta(p\psi_i) - \left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) . \prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2.\end{aligned}$$

Moreover, if η has a Gaussian distribution $N(0, \tau^2)$ the above ACVF is

$$\gamma_{\sigma_0^p\sigma_h^p}(t) = e^{2p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2 + \psi_i\psi_{i+h})}[e^{\sum_{i=0}^{\infty}(\psi_i\psi_{t+i} + \psi_i\psi_{t+h+i})} - 1]. \tag{2.4.13}$$

Proof. If $h \leq t$ we have

$$
\begin{aligned}
E(\sigma_0^p\sigma_h^p\sigma_t^p\sigma_{t+h}^p) &= Ee^{p(\sum_{i=0}^{\infty}\psi_i\eta_{-i}+\sum_{i=0}^{\infty}\psi_i\eta_{h-i}+\sum_{i=0}^{\infty}\psi_i\eta_{t-i}+\sum_{i=0}^{\infty}\psi_i\eta_{t+h-i})} \\
&= E\Big[e^{p(\sum_{i=o}^{\infty}(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t-h-i})\eta_{-i})} \\
&\quad e^{p(\sum_{j=-h}^{-1}(\psi_{j+h}+\psi_{j+t}+\psi_{t+h+j})\eta_{-j}+\sum_{l=-t}^{-h-1}(\psi_{l+t}+\psi_{t+h+l})\eta_{-l}+\sum_{m=-t-h}^{-t-1}\psi_{m+t+h}\eta_{-m})}\Big] \\
&= \prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t-h-i})).\prod_{i=-h}^{-1} m_\eta(p(\psi_{j+h}+\psi_{j+t}+\psi_{t+h+j})) \\
&\quad .\prod_{i=-t}^{-h-1} m_\eta(p(\psi_{l+t}+\psi_{t+h+l})).\prod_{i=-t-h}^{-t-1} m_\eta(p\psi_{t+h+i}).
\end{aligned}
$$

From the above result and Lemma 2.4.7 we get

$$
\begin{aligned}
\gamma_{\sigma_0^p\sigma_h^p}(t) &= E(\sigma_0^p\sigma_h^p\sigma_t^p\sigma_{t+h}^p) - (E(\sigma_0^p\sigma_h^p))^2 \\
&= \prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t+h+i})) \\
&\quad .\prod_{i=-h}^{-1} m_\eta(p(\psi_{i+h}+\psi_{i+t}+\psi_{i+t+h})) \prod_{i=-t}^{-h-1} m_\eta(p(\psi_{i+t}+\psi_{t+h+i})) \\
&\quad .\prod_{i=-t-h}^{-t-1} m_\eta(p\psi_{t+h+i}) - \left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h})).\prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2.
\end{aligned}
$$

For the special case we have

$$
\begin{aligned}
\gamma_{\sigma_0^p\sigma_h^p}(t) &= e^{\frac{p^2\tau^2}{2}[\sum_{i=0}^{\infty}(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t+h+i})^2+\sum_{i=-h}^{-1}(\psi_{i+h}+\psi_{i+t}+\psi_{i+t+h})^2+\sum_{i=-t}^{-h-1}(\psi_{i+t}+\psi_{t+h+i})^2]} \\
&\quad .e^{\frac{p^2\tau^2}{2}\sum_{i=-t-h}^{-t-1}\psi_{t+h+i}^2} - [e^{p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h})}]^2 \\
&= e^{2p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h}+\psi_i\psi_{i+t}+\psi_i\psi_{i+t+h})} - e^{2p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h})}.
\end{aligned}
$$

In the same way we can get the results for $h > t$. □

Remark 2.4.10. When η comes from a Gaussian distribution $N(0, \tau^2)$, the following result holds

$$\gamma_{\sigma_0^p\sigma_h^p}(t) = e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}[e^{\gamma_Y(t)+\gamma_Y(t+h)} - 1]. \tag{2.4.14}$$

A Taylor expansion yields

$$\gamma_{\sigma_0^p\sigma_h^p} \sim e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}(\gamma_Y(t)+\gamma_Y(t+h)), \quad t \to \infty.$$

The variance of $(\sigma_0^p\sigma_h^p)$ is given by

$$
\begin{aligned}
var(\sigma_0^p\sigma_h^p) &= E(\sigma_0^{2p}\sigma_h^{2p}) - (E(\sigma_0^p\sigma_h^p))^2 \\
&= e^{4p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h})} - (e^{p^2\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h})})^2 \\
&= e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}(e^{2p^2(\gamma_Y(0)+\gamma_Y(h))} - 1).
\end{aligned}
$$

In addition, the ACF has the form

$$\rho_{\sigma_0^p\sigma_h^p}(t) = \frac{e^{\gamma_Y(t)+\gamma_Y(t+h)}-1}{e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}-1} \sim \frac{e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}}{e^{2p^2(\gamma_Y(0)+\gamma_Y(h))}-1}(\gamma_Y(t)+\gamma_Y(t+h)) \tag{2.4.15}$$

In particular, if (Y_t) is strongly mixing with geometric rate, $\gamma_Y(t) \to 0$ as $t \to \infty$ exponentially fast, hence $\gamma_{\sigma_0^p\sigma_h^p}(t) \to 0$ exponentially fast too.

Lemma 2.4.11. *Assume $p > 0$. If $E|Z|^p < \infty$, we have*

$$E|X_t|^p = E(|Z|^p)\prod_{i=0}^{\infty} m_\eta(p\psi_i). \tag{2.4.16}$$

If $E|Z|^p < \infty$ the ACVF for $(|X|^p)$ at lag $h > 0$ is given by

$$\gamma_{|X|^p}(h) = (E(|Z|^p))^2\left(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2\right). \tag{2.4.17}$$

If $E|Z|^{2p} < \infty$ the ACF of $(|X|^p)$ at lag $h > 0$ is given by

$$\rho_{|X|^p}(h) = \frac{(E(|Z|^p))^2\left(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h}))\prod_{i=-h}^{-1} m_\eta(p\psi_{i+h}) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2\right)}{E(|Z|^{2p})\prod_{i=0}^{\infty} m_\eta(2p\psi_i) - \left[E(|Z|^p)\prod_{i=0}^{\infty} m_\eta(p\psi_i)\right]^2}. \tag{2.4.18}$$

When η has a Gaussian distribution $N(0,\tau^2)$ we have

$$E(|X|^p) = E(|Z|^p)e^{\frac{p^2\tau^2}{2}\sum_{i=0}^{\infty}\psi_i^2} = E|Z|^p e^{\frac{p^2}{2}\gamma_Y(0)}, \tag{2.4.19}$$

$$\begin{aligned} var(|X|^p) &= e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}\left(E|Z|^{2p}e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2} - (E|Z|^p)^2\right) \\ &= e^{p^2\gamma_Y(0)}\left(E|Z|^{2p}e^{p^2\gamma_Y(0)} - (E|Z|^p)^2\right), \end{aligned} \tag{2.4.20}$$

$$E(|X_t|^p|X_{t+h}|^p) = e^{p^2\gamma_Y(0)(1+\rho_Y(h))}(E|Z|^p)^2, \tag{2.4.21}$$

$$\gamma_{|X|^p}(h) = (E|Z|^p)^2 e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2}\left(e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}} - 1\right), \tag{2.4.22}$$

$$\rho_{|X|^p}(h) = \frac{(E|Z|^p)^2\left(e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}} - 1\right)}{E|Z|^{2p}e^{p^2\tau^2\sum_{i=0}^{\infty}\psi_i^2} - (E|Z|^p)^2}. \tag{2.4.23}$$

Proof. For the sequence (X_t) we have the following straightforward calculations.

$$\begin{aligned} E|X|^p &= E(\sigma^p)E(|Z|^p) = E(|Z|^p)\prod_{i=0}^{\infty} m_\eta(p\psi_i), \\ var(|X|^p) &= E(|Z|^{2p})\prod_{i=0}^{\infty} m_\eta(2p\psi_i) - \left[E(|Z|^p)\prod_{i=0}^{\infty} m_\eta(p\psi_i)\right]^2 \\ \gamma_{|X|^p}(h) &= E(\sigma_0^p|Z_0|^p\sigma_h^p|Z_h|^p) - (E|Z|^p)^2E(\sigma_0^p\sigma_h^p) = (E(|Z|^p))^2\gamma_{\sigma^p}(h). \end{aligned}$$

Now we use (2.4.5) to get the formulas for $\gamma_{|X|^p}(h)$ and $\rho_{|X|^p}(h)$.

For the special Gaussian case, the moment generating function is given by $m_\eta(t) = e^{t^2\tau^2/2}$. We have from the previous calculations

$$
\begin{aligned}
E(|X|^p) &= E(|Z|^p) e^{\frac{p^2\tau^2}{2}\sum_{i=0}^\infty \psi_i^2}, \\
var(|X|^p) &= E|Z|^{2p} e^{2p^2\tau^2\sum_{i=0}^\infty \psi_i^2} - (E|Z|^p)^2 e^{p^2\tau^2\sum_{i=0}^\infty \psi_i^2} \\
&= e^{p^2\tau^2\sum_{i=0}^\infty \psi_i^2}\left(E|Z|^{2p} e^{p^2\tau^2\sum_{i=0}^\infty \psi_i^2} - (E|Z|^p)^2\right), \\
\gamma_{|X|^p}(h) &= (E|Z|^p)^2 e^{p^2\tau^2\sum_{i=0}^\infty \psi_i^2}\left(e^{p^2\tau^2\sum_{i=0}^\infty \psi_i\psi_{i+h}} - 1\right), \\
\rho_{|X|^p}(h) &= \frac{(E|Z|^p)^2\left(e^{p^2\tau^2\sum_{i=0}^\infty \psi_i\psi_{i+h}} - 1\right)}{E|Z|^{2p} e^{p^2\tau^2\sum_{i=0}^\infty \psi_i^2} - (E|Z|^p)^2}.
\end{aligned}
$$

□

Remark 2.4.12. Note that in the above lemma $var(|X|^p)$ cannot be derived by formally setting $h = 0$ in the ACVF $\gamma_{|X|^p}(h)$ in Equation (2.4.17).

Remark 2.4.13. The ACVF $\gamma_{|X|^p}$ in the case of a Gaussian distribution $N(0, \tau^2)$ for η can be approximated by using a Taylor expansion as follows:

$$
\begin{aligned}
\gamma_{|X|^p}(h) &= (E|Z|^p)^2 e^{p^2\gamma_Y(0)}(e^{p^2\gamma_Y(h)} - 1) \\
&\sim (E|Z|^p)^2 p^2 e^{p^2\gamma_Y(0)}\gamma_Y(h), \quad h \to \infty. \qquad (2.4.24)
\end{aligned}
$$

This follows from the fact that $(\gamma_Y(h))$ decays to zero at $h \to \infty$. The ACF is approximated by

$$\rho_{|X|^p}(h) \sim \frac{(E|Z|^p)^2 p^2}{E|Z|^{2p} e^{p^2\gamma_Y(0)} - (E|Z|^p)^2}\gamma_Y(h), \quad h \to \infty. \qquad (2.4.25)$$

Remark 2.4.14. We conclude from Remarks 2.4.8 and 2.4.13 that γ_{σ^p} inherits the asymptotic behavior of the ACVF γ_Y and, in turn, $\gamma_{|X|^p}$ inherits the asymptotic behavior of γ_{σ^p} and γ_Y. In particular, if (Y_t) is strongly mixing with a geometric rate function (α_t) then we know from page 17 that $(|X_t|^p)$ and (σ_t^p) inherit strong mixing with the same rate function. We conclude from (2.4.24) and (2.4.25) that $(\gamma_{|X|^p}(h))$ and $(\gamma_{\sigma^p}(h))$ decay exponentially fast as $h \to \infty$. This is in agreement with Remark 2.4.13.

Lemma 2.4.15. *Assume $p > 0$ and $E|Z|^p < \infty$. The ACVF of the sequence $(|X_0 X_h|^p)$ for $h > 0$ at lag $t > 0$ is given by*

- *if* $h < t$

$$\begin{aligned}\gamma_{|X_0X_h|^p}(t) &= (E|Z|^p)^4\Bigg(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t+h+i}))\\ &\quad .\prod_{i=0}^{h-1} m_\eta(p(\psi_i+\psi_{i+t-h}+\psi_{i+t})\prod_{i=0}^{t-h-1} m_\eta(p(\psi_i+\psi_{h+i}))\\ &\quad .\prod_{i=0}^{h-1} m_\eta(p\psi_i) - \left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h})).\prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2\Bigg),\end{aligned} \tag{2.4.26}$$

- *if* $h > t$

$$\begin{aligned}\gamma_{|X_0X_h|^p}(t) &= (E|Z|^p)^4\Bigg(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{h+i}+\psi_{t+i}+\psi_{t+h+i}))\\ &\quad .\prod_{i=0}^{t-1} m_\eta(p(\psi_i+\psi_{i+h-t}+\psi_{i+h})\prod_{i=0}^{h-t-1} m_\eta(p(\psi_i+\psi_{t+i}))\\ &\quad .\prod_{i=0}^{t-1} m_\eta(p\psi_i) - \left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h})).\prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2\Bigg),\end{aligned} \tag{2.4.27}$$

- *if* $h = t$ *and* $E|Z|^{2p} < \infty$

$$\begin{aligned}\gamma_{|X_0X_h|^p}(h) &= (E|Z|^p)^2 E|Z|^{2p}\Bigg(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+2h}+2\psi_{h+i}))\\ &\quad .\prod_{i=0}^{h-1} m_\eta(p(2\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(p\psi_i)\Bigg)\\ &\quad -(E|Z|^p)^4\left[\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h})).\prod_{i=0}^{h-1} m_\eta(p\psi_i)\right]^2.\end{aligned} \tag{2.4.28}$$

Proof. In the stochastic volatility model we have for $t \neq h$

$$\begin{aligned}\gamma_{|X_0X_h|^p}(t) &= E(|Z_0|^p\sigma_0^p|Z_h|^p\sigma_h^p|Z_t|^p\sigma_t^p|Z_{t+h}|^p\sigma_{t+h}^p) - (E(|Z_0|^p\sigma_0^p|Z_h|^p\sigma_h^p))^2\\ &= (E|Z|^p)^4\gamma_{\sigma_0^p\sigma_h^p}(t), \quad t \neq h.\end{aligned}$$

The ACVF for $(\sigma_0^p\sigma_h^p)$ was given in Lemma 2.4.9. This yields formulas (2.4.26) and (2.4.27). If $t = h$ the ACVF is given by

$$\begin{aligned}\gamma_{|X_0X_h|^p}(h) &= (E|Z|^p)^2E|Z|^{2p}E(\sigma_0^p\sigma_h^{2p}\sigma_{2h}^p) - (E|Z|^p)^4(E(\sigma_0^p\sigma_h^p))^2,\\ E(\sigma_0^p\sigma_h^{2p}\sigma_{2h}^p) &= \prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+2h}+2\psi_{h+i}))\prod_{i=0}^{h-1} m_\eta(p(2\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(p\psi_i).\end{aligned}$$

Direct calculation leads to formula (2.4.28). □

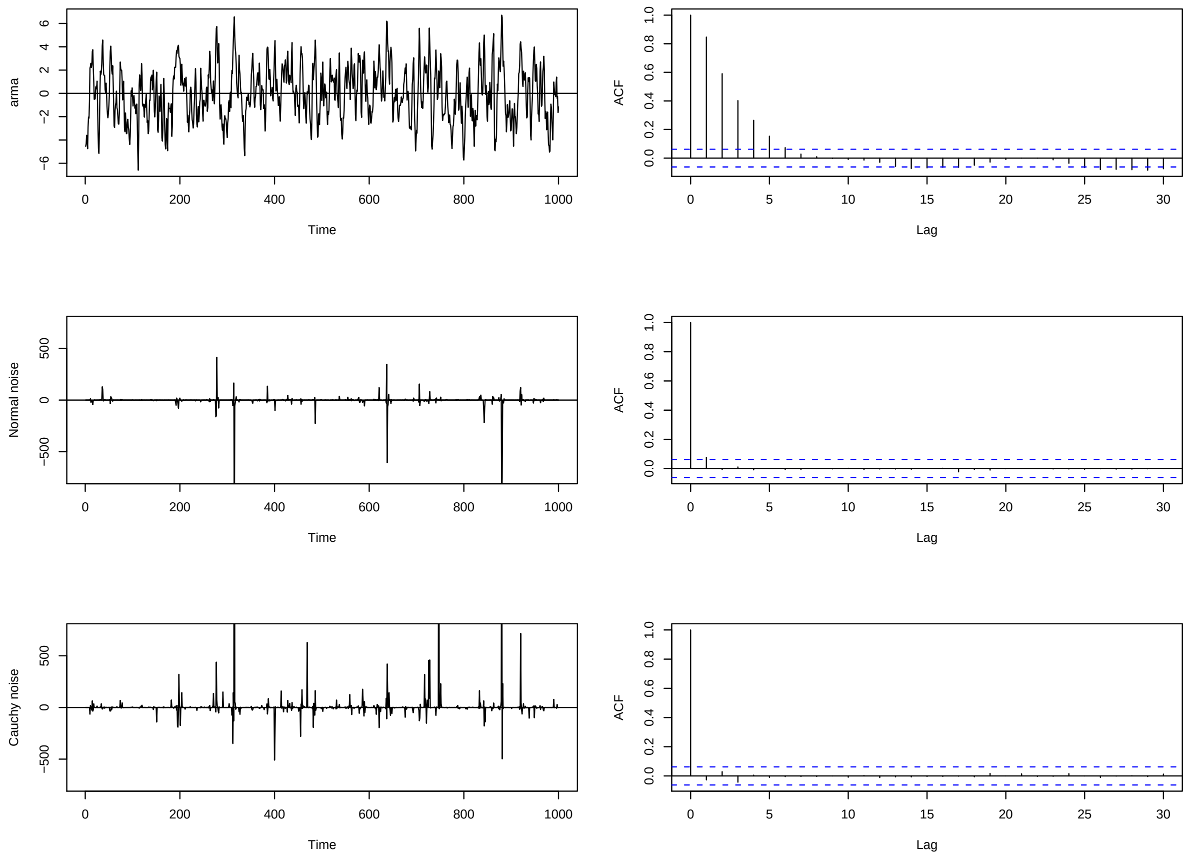

Figure 2.7: Top: A simulated time series and the corresponding sample ACF of 1000 observations from the ARMA(1,1) process given by $V_t = 0.7V_{t-1} + 0.7\delta_{t-1} + \delta_t$, for iid (δ_t) with common distribution $N(0,1)$. Middle: A simulated time series and the corresponding sample ACF of 1000 observations from the stochastic volatility model given by $Q_t = Z_t\sigma_t$ where $\log(\sigma_t) = V_t$ in the top case and Z comes from the Gaussian distribution $N(0,1)$. Bottom: A simulated time series and the corresponding sample ACF of 1000 observations from the stochastic volatility model when Z comes from a Cauchy distribution.

2.5 Asymptotic theory for the sample mean and sample variance in the stochastic volatility model

2.5.1 Asymptotic theory for the sample mean and sample variance with finite variance innovations Z_t

In what follows, we assume (X_t) is a stochastic volatility process with specification given in Section 2.4.3, in particular $EZ = 0$ and $EZ^2 = 1$, and (σ_t) is defined through equation (2.4.1) where (η_t) is an iid sequence with $E(\eta) = 0, var(\eta) = E(\eta_t^2) = \tau^2 < \infty$ and the moment generating function of η, $m_\eta(s) = Ee^{s\eta}$, is finite for all $s \in \mathbb{R}$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ and $\psi_0 = 1$.

Proposition 2.5.1. *Assume (X_t) is a stochastic volatility process. The sample mean $\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$*

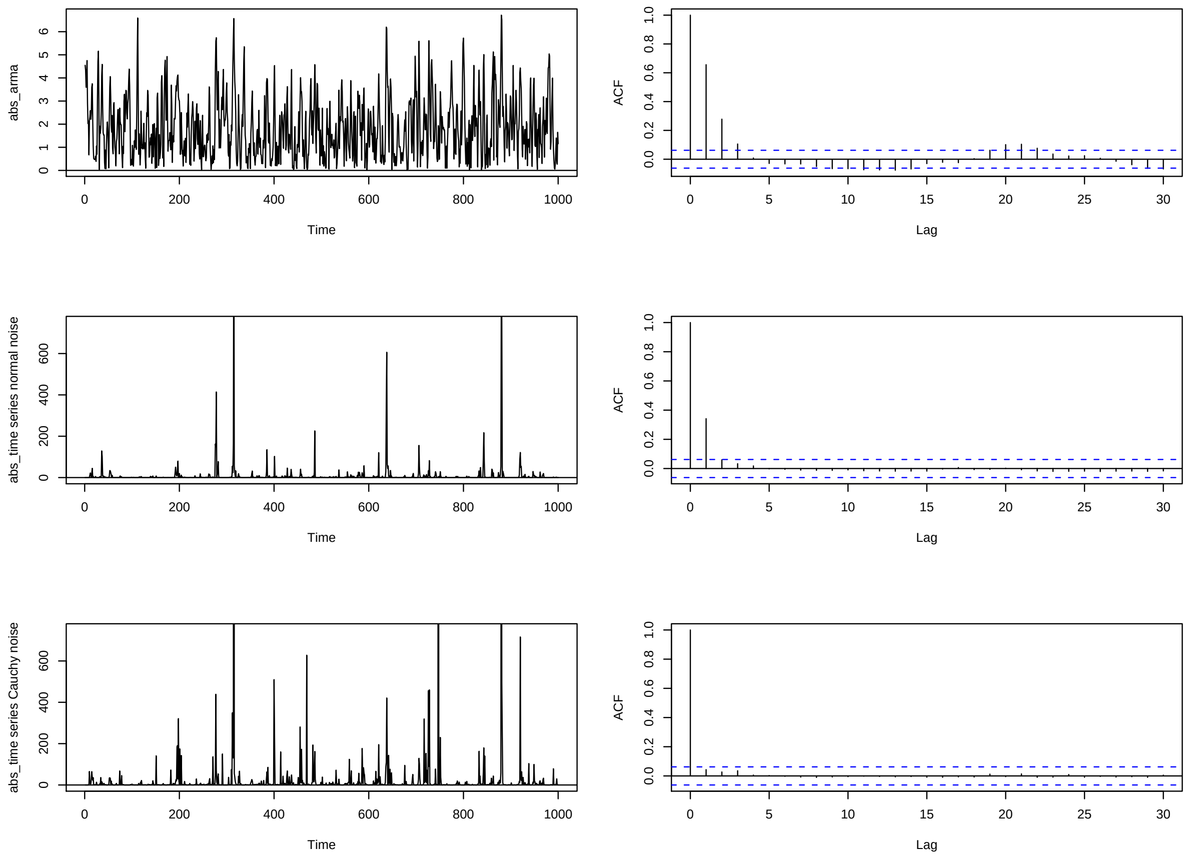

Figure 2.8: Top: The time series of the absolute values and the corresponding sample ACF of 1000 observations from the ARMA(1,1) process given in Figure 2.7. Middle: The time series of absolute values and the corresponding sample ACF of 1000 observations from the stochastic volatility model given in Figure 2.7 when Z comes from the Gaussian distribution $N(0,1)$. Bottom: The time series of absolute values and the corresponding sample ACF of 1000 observations from the stochastic volatility model given in Figure 2.7 when Z comes from a Cauchy distribution.

satisfies the following property

$$\sqrt{n}\ \overline{X}_n \xrightarrow{d} N(0, E(\sigma^2)). \tag{2.5.1}$$

In particular, if η *has Gaussian* $N(0,\tau^2)$ *distribution then*

$$\sqrt{n}\ \overline{X}_n \xrightarrow{d} N(0, e^{2\tau^2 \Sigma_{i=0}^{\infty}\psi_i^2}).$$

Proof. The sequence (X_t) constitutes a centered finite variance strictly stationary ergodic martingale difference sequence (see relation (2.4.2)). Therefore the central limit theorem for strictly stationary ergodic martingale difference sequences can be applied (see Theorem 2.1.15)

$$\sqrt{n}\ \overline{X}_n \xrightarrow{d} N(0, var(X)),$$

where $var(X) = E(\sigma^2) = \prod_{i=0}^{\infty} m_\eta(2\psi_i)$. If η has a Gaussian distribution with mean 0 and variance $\tau^2 > 0$ then $E\sigma^2 = e^{2\tau^2 \sum_{i=0}^{\infty} \psi_i^2}$, see Lemma 2.4.7. □

Proposition 2.5.2. *Assume that* $(Y_t) = (\log \sigma_t)$ *is a strictly stationary strongly mixing sequence with rate function* (α_t) *satisfying the condition* $\sum_{t=1}^{\infty} \alpha_t^{\delta/(2+\delta)} < \infty$*. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n} (\sigma_t^p - E\sigma^p) \xrightarrow{d} N\left(0, var(\sigma^p) + 2\sum_{i=1}^{\infty} \gamma_{\sigma^p}(i)\right), \tag{2.5.2}$$

where

$$var(\sigma^p) = \prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2$$

and

$$\gamma_{\sigma^p}(h) = \prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2.$$

Moreover, the infinite series $\sum_{i=1}^{\infty} \gamma_{\sigma^p}(i)$ *is finite. If* η *is Gaussian* $N(0, \tau^2)$ *distributed the above results simplify to*

$$\begin{aligned} var(\sigma^p) &= e^{p^2\gamma_Y(0)}(e^{p^2\gamma_Y(0)} - 1), \\ \gamma_{\sigma^p}(i) &= e^{p^2\gamma_Y(0)}(e^{p^2\gamma_Y(i)} - 1). \end{aligned}$$

Proof. The strictly stationary ergodic sequence (σ_t^p) inherits strong mixing from the sequence $(\log \sigma_t)$ with the same rate function. A direct application of the central limit theorem in Theorem 2.1.16 gives the above result. The variance and the ACVF of (σ_t^p) were calculated in Lemma 2.4.7 and Remark 2.4.8. □

Proposition 2.5.3. *Under the same assumptions for* $(\log \sigma_t)$ *given in Proposition 2.5.2 and if* $E|Z|^{2p+\varepsilon} < \infty$ *for some* $\varepsilon > 0$*, we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (|X_t|^p - E|X|^p) \xrightarrow{d} N(0, \nu^2), \tag{2.5.3}$$

where

$$\begin{aligned} \nu^2 = \quad & 2(E(|Z|^p))^2 \sum_{h=1}^{\infty} \left(\prod_{i=0}^{\infty} m_\eta(p(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2 \right) \\ & + E|Z|^{2p} \prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (E|Z|^p \prod_{i=0}^{\infty} m_\eta(p\psi_i))^2. \end{aligned} \tag{2.5.4}$$

If η *is Gaussian* $N(0, \tau^2)$ *then*

$$\nu^2 = e^{p^2\gamma_Y(0)}(E|Z|^{2p} e^{p^2\gamma_Y(0)} - (E|Z|^p)^2) + 2(E|Z|^p)^2 e^{p^2\gamma_Y(0)} \sum_{h=1}^{\infty} (e^{p^2\gamma_Y(h)} - 1). \tag{2.5.5}$$

Proof. The strictly stationary ergodic sequence (X_t) inherits the strong mixing properties from the sequence (Y_t) with the same rate function (α_t). Moreover, $E|X|^{2p+\varepsilon} = E\sigma^{2p+\varepsilon}E|Z|^{2p+\varepsilon} < \infty$ and $\sum_{i=1}^{\infty} \alpha_i^{\delta/(2+\delta)} < \infty$, by assumption. Hence, we can use Theorem 2.1.16 to get the central limit theorem.

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^p - E|X|^p) \xrightarrow{d} N(0,\nu^2),$$

with variance

$$\nu^2 = var(|X|^p) + 2\sum_{i=1}^{\infty}\gamma_{|X|^p}(i) < \infty.$$

From Lemma 2.4.11 we have

$$\begin{aligned} var(|X|^p) &= E|Z|^{2p}\prod_{i=0}^{\infty} m_\eta(2p\psi_i) - (E|Z|^p\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2 \\ \gamma_{|X|^p}(h) &= (E(|Z|^p))^2\left(\prod_{i=0}^{\infty} m_\eta(p(\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(p\psi_i) - (\prod_{i=0}^{\infty} m_\eta(p\psi_i))^2\right), \end{aligned}$$

This yields ν^2 in (2.5.4). If η is Gaussian $N(0,\tau^2)$ then from Lemma 2.4.11 we also have

$$\begin{aligned} var(|X_t|^p) &= e^{p^2\gamma_Y(0)}(E|Z|^{2p}e^{p^2\gamma_Y(0)} - (E|Z|^p)^2), \\ \gamma_{|X|^p}(h) &= (E|Z|^p)^2 e^{p^2\gamma_Y(0)}(e^{p^2\gamma_Y(h)} - 1). \end{aligned}$$

□

Remark 2.5.4. Another proof for Proposition 2.5.3 using the multivariate CLT for strongly mixing sequences can be derived as follows. We can write $\frac{1}{\sqrt{n}}\sum_{t=1}^{\infty}(|X_t|^p - E|X|^p)$ as follows:

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^p - E|X|^p) &= \frac{1}{\sqrt{n}}(\sum_{t=1}^{n}(\sigma_t^p|Z_t|^p - E(\sigma^p Z^p)) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_t^p(|Z_t|^p - E|Z|^p) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^p - E\sigma^p)E|Z|^p \\ &= \begin{pmatrix}1 & , & E|Z|^p\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_t^p(|Z_t|^p - E|Z|^p) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^p - E\sigma^p)\end{pmatrix}. \end{aligned} \tag{2.5.6}$$

The sequence $(\sigma_t^p - E\sigma^p)$ is a strongly mixing strictly stationary ergodic sequence. The CLT for this sequence was given in Proposition 2.5.2. The sequence $(\sigma_t^p(|Z_t|^p - E|Z|^p))$ is a mean zero finite variance strictly stationary ergodic martingale sequence with respect to the filtration $(\mathbb{G}_t)$ (see Lemma 2.4.4). Hence this sequence satisfies the CLT in Theorem 2.1.15. Moreover, $(\sigma_t^p(|Z_t|^p - E|Z|^p))$ and σ_t^p are uncorrelated and $(\sigma_t^p(|Z_t|^p - E|Z|^p))$ is strongly mixing with the same rate as (σ_t^p). Then the multivariate central limit theorem in Theorem 2.1.18 yields that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\sigma_t^p(|Z_t|^p - E|Z|^p) \\ \sigma_t^p - E\sigma^p\end{pmatrix} \xrightarrow{d} N(0,\Sigma), \tag{2.5.7}$$

where

$$\Sigma = \begin{pmatrix} E\sigma^{2p} var(|Z|^p) & 0 \\ 0 & var(\sigma^p) + 2\sum_{h=1}^{\infty} \gamma_{\sigma^p}(h) \end{pmatrix}.$$

Finally, (2.5.6) and (2.5.7) and the continuous mapping theorem imply that:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\infty} (|X_t|^p - E|X|^p) \xrightarrow{d} N(0, \nu^2),$$

where

$$\begin{aligned} \nu^2 &= \begin{pmatrix} 1 & , & E|Z|^p \end{pmatrix} \Sigma \begin{pmatrix} 1 \\ E|Z|^p \end{pmatrix} \\ &= E\sigma^{2p} var(|Z|^p) + (E|Z|^p)^2 (var(\sigma^p) + 2\sum_{h=1}^{\infty} \gamma_{\sigma^p}(h)). \end{aligned}$$

□

This approach shows nicely that the CLT for the sample mean of $(|X_t|^p)$ is essentially determined by the CLT for the sample mean of (σ_t^p) and the CLT for the asymptotically independent sample mean of $(\sigma_t^p(|Z_t|^p - E|Z|^p))$.

Lemma 2.5.5. *If (X_t) is a stochastic volatility sequence, $(\log \sigma_t)$ satisfies the conditions in Proposition 2.5.2 and $E|Z|^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$, then*

$$\sqrt{n}(\frac{1}{n} \sum_{t=1}^{n} (X_t - \overline{X}_n)^2 - var(X)) \xrightarrow{d} N(0, \nu^2), \tag{2.5.8}$$

where

$$\begin{aligned} \nu^2 &= EZ^4 \prod_{i=0}^{\infty} m_\eta(4\psi_i) - [EZ^2 \prod_{i=0}^{\infty} m_\eta(2\psi_i)]^2 \\ &\quad + 2(EZ^2)^2 \sum_{h=1}^{\infty} \left(\prod_{i=0}^{\infty} m_\eta(2(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(2\psi_i) - (\prod_{i=0}^{\infty} m_\eta(2\psi_i))^2 \right). \end{aligned}$$

If η is a Gaussian distribution $N(0, \tau^2)$ then

$$\nu^2 = e^{4\gamma_Y(0)}(EZ^4 e^{4\gamma_Y(0)} - (EZ^2)^2) + 2(EZ^2)^2 e^{4\gamma_Y(0)} \sum_{h=1}^{\infty} (e^{4\gamma_Y(h)} - 1).$$

Proof. We observe that

$$\begin{aligned}
&\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(X_t-\overline{X}_n)^2-var(X)) \\
&= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}X_t^2-(\overline{X}_n)^2-var(X)) \\
&= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(X_t^2-EX^2)-\overline{X}_n^2+(EX)^2) \\
&= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(X_t^2-EX^2)-(\overline{X}_n-EX)(\overline{X}_n+EX)) \\
&= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(X_t^2-EX^2))-(\overline{X}_n+EX)\sqrt{n}(\overline{X}_n-EX) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(X_t^2-EX^2)+o_p(1).
\end{aligned}$$

In the last step we used the SLLN

$$\overline{X}_n \to EX = 0 \quad a.s.,$$

and the CLT

$$\sqrt{n}(\overline{X}_n-EX) \xrightarrow{d} N(0, var(X)),$$

from Proposition 2.5.1. Hence $\sqrt{n}((\overline{X}_n)^2-(EX)^2) \xrightarrow{P} 0$ and therefore it suffices to prove a CLT for (X_t^2). We may apply the CLT in Theorem 2.1.16 to obtain

$$\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(X_t^2-EX^2)) \quad \xrightarrow{d} \quad N(0, var(X^2)+2\sum_{h=1}^{\infty}\gamma_{X^2}(h)),$$

since (X_t^2) is a strongly mixing ergodic sequence with rate function (α_t) satisfying $\sum_{t=1}^{\infty}\alpha_t^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ and since $E|X|^{4+\varepsilon} = E|Z|^{4+\varepsilon}E\sigma^{4+\varepsilon} < \infty$. Using Lemma 2.4.11, we get

$$\begin{aligned}
var(X^2) &= EZ^4\prod_{i=0}^{\infty}m_\eta(4\psi_i)-[EZ^2\prod_{i=0}^{\infty}m_\eta(2\psi_i)]^2, \\
\gamma_{X^2}(h) &= (EZ^2)^2\left(\prod_{i=0}^{\infty}m_\eta(2(\psi_i+\psi_{i+h}))\prod_{i=-h}^{-1}m_\eta(2\psi_{i+h})-(\prod_{i=0}^{\infty}m_\eta(2\psi_i))^2\right).
\end{aligned}$$

If η has a Gaussian distribution $N(0,\tau^2)$ then by Lemma 2.4.11

$$\begin{aligned}
var(X^2) &= e^{4\gamma_Y(0)}(EZ^4e^{4\gamma_Y(0)}-(EZ^2)^2), \\
\gamma_{X^2}(h) &= (EZ^2)^2e^{4\gamma_Y(0)}(e^{4\gamma_Y(h)}-1).
\end{aligned}$$

□

In the same way, it can be shown for the strictly stationary strongly mixing ergodic sequence (σ_t^p) that the following result holds.

Lemma 2.5.6. *If the volatility sequence (σ_t) satisfies the assumptions in Proposition 2.5.2, the following CLT holds*

$$\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^p - \frac{1}{n}\sum_{j=1}^{n}\sigma_j^p)^2 - var(\sigma^p)) \xrightarrow{d} N(0,\nu^2), \qquad (2.5.9)$$

where

$$\begin{aligned}\nu^2 &= 4(E\sigma^p)^2 var(\sigma^p) + var(\sigma^{2p}) - 4E\sigma^p cov(\sigma_t^{2p},\sigma_t^p) + 2\sum_{h=1}^{\infty}\gamma_{\sigma^{2p}}(h)\\ &\quad +8(E\sigma^p)^2\sum_{h=1}^{\infty}\gamma_{\sigma^p}(h) - 8E\sigma^p\sum_{h=1}^{\infty}cov(\sigma_0^{2p},\sigma_h^p).\end{aligned}$$

Proof. In the same way as in the proof of Lemma 2.5.5 we start with the following decomposition

$$\begin{aligned}&\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^p - \frac{1}{n}\sum_{j=1}^{n}\sigma_j^p)^2 - var(\sigma^p))\\ &= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^{2p} - E\sigma^{2p})) - (\frac{1}{n}\sum_{t=1}^{n}\sigma_t^p + E\sigma^p)\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}\sigma_t^p - E\sigma^p)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^{2p} - E\sigma^{2p}) - 2E\sigma^p\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}\sigma_t^p - E\sigma^p) + o_p(1) \qquad (2.5.10)\\ &= \begin{pmatrix}1 & , & -2E\sigma^p\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^{2p} - E\sigma^{2p})\\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^p - E\sigma^p)\end{pmatrix},\end{aligned}$$

Here we used the SLLN

$$\frac{1}{n}\sum_{t=1}^{n}\sigma_t^p \rightarrow E\sigma^p \ \ a.s.,$$

to get that $\frac{1}{n}\sum_{t=1}^{n}\sigma_t^p + E\sigma^p = 2E\sigma^p + o_p(1)$. Note that the sequences (σ_t^{2p}) and (σ_t^p) are strongly mixing stationary ergodic sequences. The CLT in Lemma 2.1.19 can be applied.

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}(\sigma_t^{2p} - E\sigma^{2p})\\ (\sigma_t^p - E\sigma^p)\end{pmatrix} \xrightarrow{d} N(\underline{0},\Sigma),$$

where

$$\Sigma = \begin{pmatrix} var(\sigma^{2p}) + 2\sum_{h=1}^{\infty}\gamma_{\sigma^{2p}}(h) & cov(\sigma_0^{2p},\sigma_0^p) + 2\sum_{t=1}^{\infty}cov(\sigma_0^{2p},\sigma_t^p)\\ cov(\sigma_0^{2p},\sigma_0^p) + 2\sum_{t=1}^{\infty}cov(\sigma_0^{2p},\sigma_t^p) & var(\sigma^p) + 2\sum_{h=1}^{\infty}\gamma_{\sigma^p}(h)\end{pmatrix}.$$

Therefore

$$\begin{pmatrix}1 & , & -2E\sigma^p\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^{2p} - E\sigma^{2p})\\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^p - E\sigma^p)\end{pmatrix} \xrightarrow{d} N(\underline{0},a'\Sigma a)$$

where

$$
\begin{aligned}
a' &= \begin{pmatrix}1 & , & -2E\sigma^p\end{pmatrix}, \\
a'\Sigma a &= \begin{pmatrix}1 & , & -2E\sigma^p\end{pmatrix}\Sigma\begin{pmatrix}1\\ -2E\sigma^p\end{pmatrix} \\
&= 4(E\sigma^p)^2 var(\sigma^p) + var(\sigma^{2p}) - 4E\sigma^p cov(\sigma_t^{2p}, \sigma_t^p) + 2\sum_{h=1}^{\infty}\gamma_{\sigma^{2p}}(h) \\
&\quad +8(E\sigma^p)^2\sum_{h=1}^{\infty}\gamma_{\sigma^p}(h) - 8E\sigma^p\sum_{h=1}^{\infty}cov(\sigma_0^{2p}, \sigma_h^p).
\end{aligned}
$$

□

Remark 2.5.7. Alternatively the proof of Lemma 2.5.6 can be given as follows. Starting from (2.5.10), we have

$$
\begin{aligned}
&\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^p - \frac{1}{n}\sum_{j=1}^{n}\sigma_j^p)^2 - var(\sigma^p)) \\
&= \sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}[\sigma_t^{2p} - E\sigma^{2p} - 2E\sigma^p(\sigma_t^p - E\sigma^p)]) + o_p(1).
\end{aligned}
$$

The sequence $(\sigma_t^{2p} - 2\sigma_t^p E\sigma^p)$ is a strongly mixing sequence with rate function (α_t) such that $\sum_{t=1}^{\infty}\alpha_t^{\delta/(2+\delta)} < \infty$ and σ_t is supposed to have all moments finite. Theorem 2.1.16 can be applied:

$$
\sqrt{n}(\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^p - \frac{1}{n}\sum_{j=1}^{n}\sigma_j^p)^2 - var(\sigma^p)) \xrightarrow{d} N(0, \nu^2),
$$

where

$$
\begin{aligned}
\nu^2 &= var(\sigma_t^{2p} - E\sigma^{2p} - 2E\sigma^p(\sigma_t^p - E\sigma^p)) \\
&\quad +2\sum_{h=1}^{\infty}cov(\sigma_0^{2p} - E\sigma^{2p} - 2E\sigma^p(\sigma_0^p - E\sigma^p), (\sigma_h^{2p} - E\sigma^{2p} - 2E\sigma^p(\sigma_h^p - E\sigma^p))) \\
&= var(\sigma^{2p}) + 4(E\sigma^p)^2 var(\sigma^p) - 4E\sigma^p cov(\sigma^{2p}, \sigma^p) + 2\sum_{h=1}^{\infty}\gamma_{\sigma^{2p}}(h) \\
&\quad +8(E\sigma^p)^2\sum_{h=1}^{\infty}\gamma_{\sigma^p}(h) - 8E\sigma^p\sum_{h=1}^{\infty}cov(\sigma_0^{2p}, \sigma_h^p).
\end{aligned}
$$

□

Lemma 2.5.8. *Assume (X_t) is a stochastic volatility sequence satisfying the conditions in Proposition 2.5.2 and $E|Z|^{4p+\varepsilon} < \infty$ for some $\varepsilon > 0$. Let $\overline{|X|^p}_n = \frac{1}{n}\sum_{t=1}^{n}|X_t|^p$. Then*

$$
\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n}(|X_t|^p - \overline{|X|^p}_n)^2 - var(|X|^p)\right) \xrightarrow{d} N(0, \nu^2), \tag{2.5.11}
$$

where

$$\begin{aligned}\nu^2 &= var(|X|^{2p}) + 2\sum_{h=1}^{\infty}\gamma_{|X|^{2p}}(h) - 4E|X|^p cov(|X|^{2p},|X|^p)\\ &\quad -8E|X|^p\sum_{r=1}^{\infty}cov(|X_0|^{2p},|X_r|^p) + 4(E|X|^p)^2 var(|X|^p) + 8(E|X|^p)^2\sum_{h=1}^{\infty}\gamma_{|X|^p}(h).\end{aligned}$$

Proof. We observe that

$$\begin{aligned}&\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n}(|X_t|^p - \overline{|X|^p})^2 - var(|X|^p)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ &\quad -\sqrt{n}(\overline{|X|^p}_n - E|X|^p)(\overline{|X|^p} + E|X|^p)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ &\quad -2E|X|^p\sqrt{n}(\overline{|X|^p}_n - E|X|^p) + o_p(1)\\ &= \begin{pmatrix}1 & , & -2E|X|^p\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ \sqrt{n}(\overline{|X|^p}_n - E|X|^p)\end{pmatrix}\\ &= a'\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ \sqrt{n}(\overline{|X|^p}_n - E|X|^p)\end{pmatrix}.\end{aligned}$$

Notice that we used the SLLN $\overline{|X|^p}_n \to E|X|^p$ a.s. Hence

$$(\overline{|X|^p} + E|X|^p) = 2E|X|^p + o_p(1).$$

The sequences $(|X_t|^{2p})$ and $(|X_t|^p)$ are strongly mixing stationary ergodic sequences with rate functions (α_t) such that $\sum_{t=1}^{\infty}\alpha_t^{\delta/(2+\delta)} < \infty$. Moreover $E|X|^{4p+\varepsilon} = E\sigma^{4p+\varepsilon}E|Z|^{4p+\varepsilon} < \infty$. Lemma 2.1.19 gives that

$$\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ \sqrt{n}(\overline{|X|^p}_n - E|X|^p)\end{pmatrix} \xrightarrow{d} N(\underline{0},\Sigma),$$

where

$$\Sigma = \begin{pmatrix} var(|X|^{2p}) + 2\sum_{h=1}^{\infty}\gamma_{|X|^{2p}}(h) & cov(|X|^{2p},|X|^p) + 2\sum_{r=1}^{\infty}cov(|X_0|^{2p},|X_r|^p)\\ cov(|X|^{2p},|X|^p) + 2\sum_{r=1}^{\infty}cov(|X_0|^{2p},|X_r|^p) & var(|X|^p) + 2\sum_{h=1}^{\infty}\gamma_{|X|^p}(h)\end{pmatrix}.$$

Then

$$a'\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_t|^{2p} - E|X|^{2p})\\ \sqrt{n}(\overline{|X|^p}_n - E|X|^p)\end{pmatrix} \xrightarrow{d} N(0, a'\Sigma a),$$

where

$$\begin{aligned}a'\Sigma a &= var(|X|^{2p}) - 4E|X|^p cov(|X|^{2p},|X|^p) - 8E|X|^p\sum_{r=1}^{\infty}cov(|X_0|^{2p},|X_r|^p)\\ &\quad +4(E|X|^p)^2 var(|X|^p) + 8(E|X|^p)^2\sum_{h=1}^{\infty}\gamma_{|X|^p}(h) + 2\sum_{h=1}^{\infty}\gamma_{|X|^{2p}}(h).\end{aligned}$$

□

Remark 2.5.9. The same idea of the proof in Remark 2.5.7 can be repeated for Lemma 2.5.8 to get the above results.

2.5.2 Asymptotic theory for the sample mean and sample variance with infinite variance α-stable innovations (Z_t)

Any time series which exhibits sharp spikes or occasional bursts of outlying observations suggests the possible use of an infinite variance model (see [14]). In this section we study the asymptotic behavior of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{t=1}^n X_t$ and of the sample variance $\widehat{s}_n^2 = \frac{1}{n}\sum_{t=1}^n (X_t - \overline{X})^2$ in the case when (σ_t) is a strictly stationary ergodic sequence independent of the iid noise (Z_t), but we give up the assumption of finite variance, $var(Z) < \infty$.

Definition 2.5.10. (Stable distribution) [37, p. 5]
A random variable Z is said to be stable, or to have a stable distribution, if for every positive integer n there exist constants, $a_n > 0$ and b_n, such that the sum $Z_1 + \cdots + Z_n$ has the same distribution as $a_n Z + b_n$ for iid random variables $Z_1, \ldots, Z_n$, with the same distribution as Z.

Properties of a stable random variable, Z

- The characteristic function, $\phi(u) = Ee^{iuZ}$, is given by

$$\phi(u) = \begin{cases} e^{iu\beta - c|u|^{\alpha}(1 - i\theta sgn(u)\tan(\pi\alpha/2))}, & \text{if } \alpha \neq 1; \\ e^{iu\beta - c|u|(1 + i\theta(2\pi)sgn(u)\log|u|)}, & \text{if } \alpha = 1; \end{cases} \tag{2.5.12}$$

  where $sgn(u)$ is $u/|u|$ if $u \neq 0$, and zero otherwise. The parameters $\alpha \in (0, 2]$, $\beta \in \mathbb{R}$, $c \in [0, \infty)$ and $\theta \in [-1, 1]$ are known as the exponent, location, scale and symmetry parameters respectively.

- If $\alpha = 2$ then Z has a Gaussian distribution, $N(\beta, 2c)$.

- If $\theta = 0$ then the distribution of Z is symmetric about β. The symmetric stable distributions (i.e. those which are symmetric about 0) have characteristic functions of the form

$$\phi(u) = e^{-c|u|^{\alpha}}, \quad u \in \mathbb{R}. \tag{2.5.13}$$

- If $\alpha = 1$, $\theta = 0$, $c = 1$ and $\beta = 0$ then Z has the Cauchy distribution with probability density

$$f(z) = \frac{1}{\pi(1 + z^2)}, \quad z \in \mathbb{R}.$$

- The symmetric stable distributions satisfy the defining property in Definition 2.5.10 with $a_n = n^{1/\alpha}$ and $b_n = 0$, since if $Z, Z_1, \dots, Z_n$ all have the characteristic function (2.5.13) and $Z_1, \dots, Z_n$ are independent, then

$$Ee^{iu(Z_1+\cdots+Z_n)} = e^{-nc|u|^\alpha} = Ee^{iuZn^{1/\alpha}}.$$

Definition 2.5.11. (Laplace-Stieltjes transform) [21, p. 176]
The Laplace-Stieltjes transform of a random variable $Y \geq 0$ *is the function* $l_Y : \mathbb{R}_+ \to \mathbb{R}_+$ *given by* $l_Y(t) = Ee^{-tY}$.

Laplace-Stieltjes transforms exist for all positive random variables.

Properties of the Laplace-Stieltjes transform for a random variable Y

- l_Y is uniformly continuous and $0 < l_Y(t) < l_Y(0) = 1$ for all t.
- If Y and U are positive and independent, then $l_{Z+U} = l_Z l_U$.
- If Y and U are positive and $l_Y(t) = l_U(t)$ for all t belonging to an open interval in $\mathbb{R}_+$, then $Y \stackrel{d}{=} U$.
- If $Y \geq 0$ and $EY^k < \infty$, then the derivative $l_Y^{(k)}$ exists and $EY^k = (-1)^k l_Y^{(k)}(0)$.
- Given positive random variables $Y, Y_1, \dots, Y_n, \dots$, the relation $Y_n \stackrel{d}{\to} Y$ holds if and only if $l_{Y_n}(t) \to l_Y(t)$ for all $t \in \mathbb{R}_+$.

In what follows, we will study the joint asymptotic behavior of $(\overline{X}_n, \widehat{s}_n^2)$ under the assumption that (Z_t) is iid symmetric α-stable (sαs) for some $\alpha \in (0, 2)$. This means that Z has characteristic function

$$Ee^{i\lambda Z} = e^{-c|\lambda|^\alpha}, \quad \lambda \in \mathbb{R}, \tag{2.5.14}$$

for some constant $c > 0$. It is well known that

$$\begin{aligned} E|Z|^p &< \infty, \quad 0 < p < \alpha, \\ E|Z|^p &= \infty, \quad p \geq \alpha, \end{aligned}$$

see [37, p. 18]. Hence $var(Z) = \infty$ for $0 < \alpha < 2$ and $E|Z| < \infty$ only if $1 < \alpha < 2$.

We also recall the notion of a positive α-stable random variable Y. It has Laplace-Stieltjes transform

$$Ee^{-sY} = e^{-c's^\alpha}, \quad s \geq 0, \tag{2.5.15}$$

for some constant $c' > 0$, see [37, p. 15]. Here it is necessary that $\alpha \in (0,1)$. It is immediate from (2.5.14) and (2.5.15) that for iid copies Z_i of Z and Y_i of Y, respectively, and weights $c_1, \ldots, c_n \in \mathbb{R}$ and $d_1, \ldots, d_n > 0$, any $n \geq 1$

$$\begin{aligned} c_1 Z_1 + \cdots + c_n Z_n &\stackrel{d}{=} Z_1 (\sum_{i=1}^{n} |c_i|^\alpha)^{\frac{1}{\alpha}}, \\ d_1 Y_1 + \cdots + d_n Y_n &\stackrel{d}{=} Y_1 (\sum_{i=1}^{n} d_i^\alpha)^{\frac{1}{\alpha}}. \end{aligned} \tag{2.5.16}$$

In what follows, we will often make use of the joint mixed characteristic function-Laplace-Stieltjes transformation of a pair of random variables (A, B), where $B \geq 0$, given by

$$f(\lambda, s) = E e^{i\lambda A - sB}, \quad \lambda \in \mathbb{R}, s \geq 0.$$

This transform determines the distribution of (A, B) by virtue of the Stone-Weierstrass theorem, see [36, p. 115]. Moreover if $(A_n, B_n) \stackrel{d}{\rightarrow} (A, B)$ for a sequence $((A_n, B_n))$ with $B_n \geq 0$ *a.s.*, $n \geq 1$, this is equivalent to

$$f_n(\lambda, s) = E e^{i\lambda A_n - sB_n} \rightarrow f(\lambda, s), \quad \lambda \in \mathbb{R}, s \geq 0. \tag{2.5.17}$$

Proposition 2.5.12. *Assume $E\sigma^\alpha < \infty$ and $\alpha \in (0,2)$. Then*

$$(n^{1-\frac{1}{\alpha}} \overline{X}_n, n^{1-\frac{2}{\alpha}} \widehat{s}_n^2) \stackrel{d}{\rightarrow} (S_\alpha, Y_{\frac{\alpha}{2}}),$$

where S_α is sαs with characteristic function

$$E e^{i\lambda S_\alpha} = e^{-E\sigma^\alpha c |\lambda|^\alpha}, \quad \lambda \in \mathbb{R},$$

$Y_{\frac{\alpha}{2}}$ has Laplace-Stieltjes transformation

$$E e^{-sY_{\frac{\alpha}{2}}} = e^{-E\sigma^\alpha E|N|^\alpha c (2s)^{\frac{\alpha}{2}}}, \quad s \geq 0,$$

and N is $N(0,1)$ distributed. The joint mixed characteristic function-Laplace-Stieltjes transform of $(S_\alpha, Y_{\frac{\alpha}{2}})$ is given by

$$f(\lambda, t) = e^{-E\sigma^\alpha E|\lambda + \sqrt{2t} N|^\alpha c}, \quad \lambda \in \mathbb{R}, t \geq 0. \tag{2.5.18}$$

Remark 2.5.13. Note that

$$f(0, t) = E e^{-E\sigma^\alpha E|N|^\alpha (2t)^{\frac{\alpha}{2}}}, \quad c \in \mathbb{R}, t \geq 0.$$

is indeed the Laplace-Stieltjes transformation of a positive $\frac{\alpha}{2}$-stable random variable for $\alpha \in (0,2)$, see (2.5.15). Similarly,

$$f(\lambda, 0) = e^{-cE\sigma^\alpha |\lambda|^\alpha}, \quad \lambda \in \mathbb{R}.$$

is the characteristic function of a sαs random variable.

Proof. (Proof of Proposition 2.5.12.) The joint mixed Laplace-Stieltjes transform-characteristic function of $(n^{1-\frac{1}{\alpha}}\overline{X}_n, n^{-\frac{2}{\alpha}}(X_1^2+\cdots+X_n^2))$ is given by

$$f_n(\lambda,\frac{s^2}{2}) = Ee^{i\lambda n^{-\frac{1}{\alpha}}(X_1+\cdots+X_n)-\frac{s^2}{2}(X_1^2+\cdots+X_n^2)n^{-\frac{2}{\alpha}}},\quad \lambda\in\mathbb{R},\ s\in\mathbb{R}.$$

Let (N_t) be iid $N(0,1)$ independent of (σ_t) and (Z_t). Conditioning on (X_t), we get

$$\begin{aligned} f_n(\lambda,\frac{s^2}{2}) &= E(Ee^{i\lambda n^{-\frac{1}{\alpha}}(X_1+\cdots+X_n)+is(N_1X_1+\cdots+N_nX_n)n^{-\frac{1}{\alpha}}}|(X_t)) \\ &= Ee^{in^{-\frac{1}{\alpha}}\sum_{t=1}^n X_t(\lambda+N_ts)}. \end{aligned}$$

Here we used the fact that $Ee^{isN}=e^{-\frac{s^2}{2}}, s\in\mathbb{R}$. Now condition on (N_t) and (σ_t) to obtain

$$\begin{aligned} f_n(\lambda,\frac{s^2}{2}) &= EE(e^{in^{-\frac{1}{\alpha}}\sum_{t=1}^n Z_t\sigma_t(\lambda+N_ts)}|(N_t),(\sigma_t)) \\ &= Ee^{-Z_1(n^{-1}\sum_{t=1}^n\sigma_t^\alpha|\lambda+N_ts|^\alpha)^{1/\alpha}}. \end{aligned}$$

Here we used (2.5.16). Thus, using the characteristic function of Z_1,

$$f(\lambda,\frac{s^2}{2}) = Ee^{-n^{-1}\sum_{t=1}^n\sigma_t^\alpha|\lambda+sN_t|^\alpha c}.$$

Notice that (σ_t) is ergodic and independent of (N_t) which is an iid sequence. Therefore $((\sigma_t,N_t)_{t\in\mathbb{Z}})$ is ergodic and in consequence $(g(\sigma_t,N_t))_{t\in\mathbb{Z}}$. Therefore $(\sigma_t^\alpha|\lambda+sN_t|^\alpha)$ is ergodic. Hence the SLLN applies to this sequence and we get

$$f_n(\lambda,s^2/2) \quad\to\quad e^{-E(\sigma^\alpha|\lambda+sN|^\alpha)c},$$

by Lebesgue dominated convergence. Writing $t=s^2/2$ or $s=\sqrt{2t}$, we obtain

$$f_n(\lambda,t)\to e^{-E\sigma^\alpha E|\lambda+\sqrt{2t}N|^\alpha c} = f(\lambda,t).$$

Notice that

$$f(\lambda,0)=e^{-E\sigma^\alpha c|\lambda|\alpha},\quad \lambda\in\mathbb{R}.$$

This is the characteristic function of

$$S_\alpha \stackrel{d}{=} Z_1(E\sigma^\alpha)^{\frac{1}{\alpha}}.$$

On the other hand,

$$f(0,t)=e^{-E\sigma^\alpha(2t)^{\alpha/2}E|N|^\alpha c}.$$

By (2.5.15) this is the Laplace-Stieltjes transform of a positive $\alpha/2$-stable random variable $Y_{\alpha/2}$. Since $f(\lambda,t)\neq f(\lambda,0)f(0,t)$ in general, S_α and $Y_{\alpha/2}$ are dependent.

The result is proved by observing that

$$\begin{aligned} n^{1-2/\alpha}\widehat{s}_n^2 &= n^{1-2/\alpha}(\frac{1}{n}\sum_{t=1}^{n} X_t^2 - (\overline{X}_n)^2) \\ &= n^{-2/\alpha}\sum_{t=1}^{n} X_t^2 - (n^{-\frac{1}{\alpha}-\frac{1}{2}}(X_1 + \cdots + X_n))^2 \\ &= n^{-2/\alpha}\sum_{t=1}^{n} X_t^2 + o_p(1). \end{aligned}$$

Hence we proved that

$$n(n^{\frac{-1}{\alpha}}\overline{X}_n, n^{-\frac{2}{\alpha}}\widehat{s}_n^2) = (n^{1-\frac{1}{\alpha}}\overline{X}_n, n^{-\frac{2}{\alpha}}(X_1^2 + \cdots + X_n^2)) + o_p(1) \xrightarrow{d} (S_\alpha, Y_{\frac{\alpha}{2}}),$$

as desired. □

An immediate consequence of the continuous mapping theorem is the following relation.

Lemma 2.5.14. *Under the assumptions in Proposition 2.5.12, the following relation holds.*

$$\sqrt{n}\,\frac{\overline{X}_n}{\widehat{s}_n} \xrightarrow{d} \frac{S_\alpha}{\sqrt{Y_{\alpha/2}}} \tag{2.5.19}$$

Proof. We observe by the continuous mapping theorem that

$$\begin{aligned} \frac{n\overline{X}_n}{\sqrt{n}\widehat{s}_n} &= \frac{X_1 + \cdots + X_n}{(X_1^2 + \cdots + X_n^2)^{\frac{1}{2}}} \\ &= \frac{n^{-\frac{1}{\alpha}}(X_1 + \cdots + X_n)}{(n^{-\frac{2}{\alpha}}(X_1^2 + \cdots + X_n^2))^{\frac{1}{2}}} \xrightarrow{d} \frac{S_\alpha}{(Y_{\frac{\alpha}{2}})^{\frac{1}{2}}}. \end{aligned}$$

□

The relation (2.5.19) means that the standardized sample mean satisfies a central limit theorem. In contrast to the case when $var(X) < \infty$, the limit is not Gaussian, but is a rather unfamiliar ratio of two dependent infinite variance variables. For comparison, we consider the case where (X_i) is an iid mean zero unit variance sequence. Then the central limit theorem and the SLLN imply that $\sqrt{n}\ \overline{X}_n \xrightarrow{d} N(0,1)$ and $\widehat{s}_n^2 \to 1$ *a.s.* Hence the central limit theorem for the standardized sample mean holds:

$$\sqrt{n}\,\frac{\overline{X}_n}{\widehat{s}_n} \xrightarrow{d} N(0,1). \tag{2.5.20}$$

The next example illustrates this result.

Example 2.5.15. Let (Q_t) be a stochastic volatility model given by $Q_t = Z_t\sigma_t$ where $\log(\sigma_t) = V_t$ is an ARMA$(1,1)$ process given by

$$V_t = 0.7V_t + 0.7\delta_t + \delta_t,$$

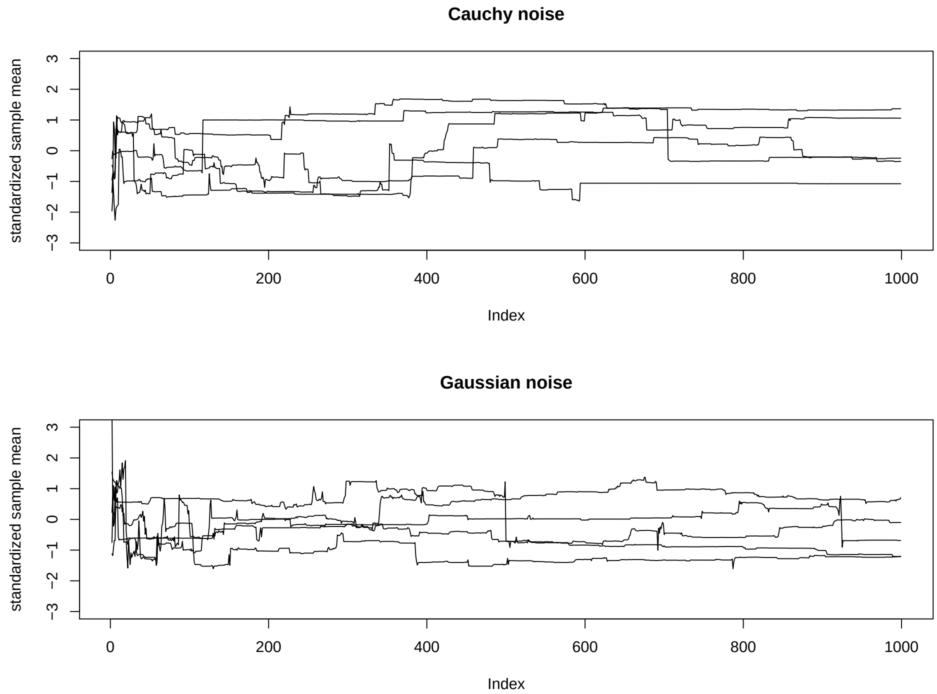

Figure 2.9: (Example 2.5.15) Top: 5 sample mean paths for a stochastic volatility model when Z comes from the Cauchy distribution ($\alpha = 1$). Bottom: 5 sample mean paths for a stochastic volatility model when Z comes from the standard normal distribution ($\alpha = 2$).

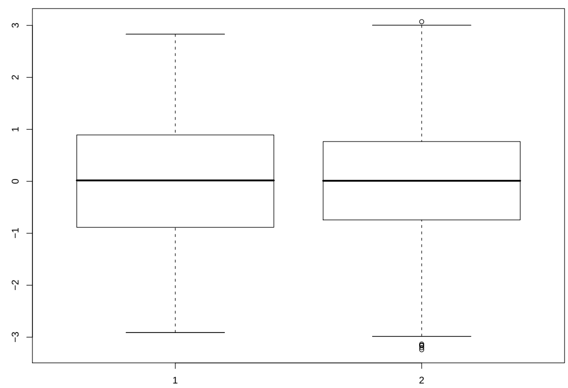

Figure 2.10: (Example 2.5.15) Boxplot for 10000 standardized sample means for a stochastic volatility model (1) when Z comes from the Cauchy distribution ($\alpha = 1$) and (2) the boxplot for 10000 standardized sample means when Z comes from the standard normal distribution ($\alpha = 2$).

for iid (δ_t) with common distribution $N(0, 1)$. We simulated 10 samples each of size 1000 observations from the stochastic volatility model. Five of these samples have Z from a Cauchy distribution and the other five samples are for the case when Z has a normal distribution, i.e., when $\alpha = 1$ and 2 respectively.

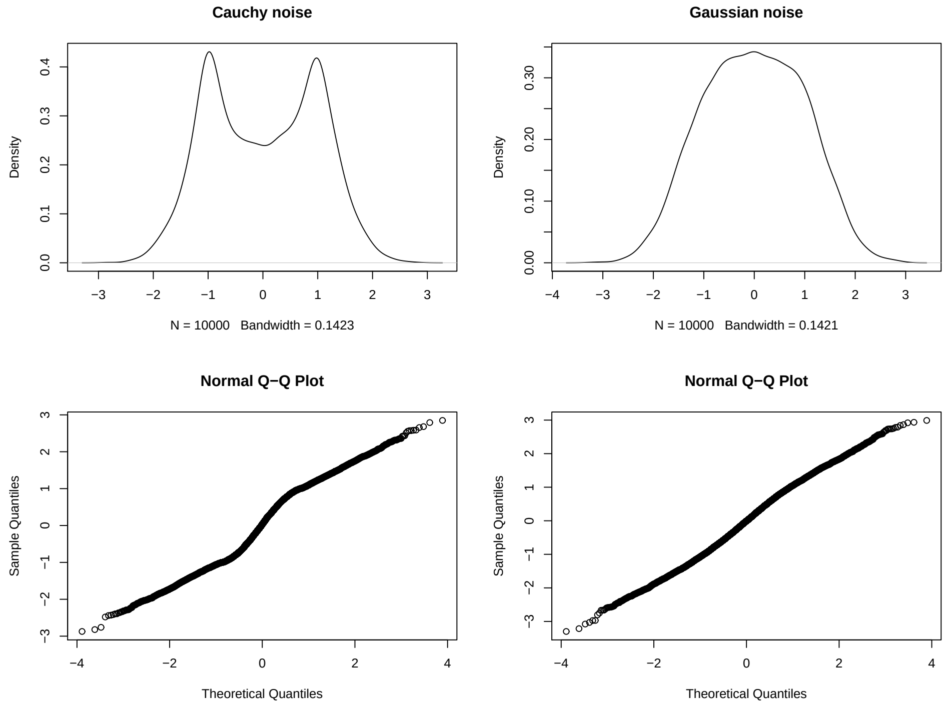

Figure 2.11: (Example 2.5.15) Let (Q_t) be a stochastic volatility model given by $Q_t = Z_t\sigma_t$ where $\log(\sigma_t) = V_t$ is an ARMA$(1,1)$ process given by $V_t = 0.7V_{t-1} + 0.7\delta_{t-1} + \delta_t$, for iid (δ_t) with common distribution $N(0,1)$. Top: Left: The density from 10000 standardized sample means with sample size 1000 for a stochastic volatility model when Z comes from the Cauchy distribution ($\alpha = 1$). Right: The density of 10000 standardized sample means with sample size 1000 for a stochastic volatility model when Z comes from the standard Gaussian distribution ($\alpha = 2$). Bottom: Left: The QQ–plot for 10000 standardized sample means for a stochastic volatility model when Z comes from the Cauchy distribution ($\alpha = 1$). Right: The QQ–plot for 10000 standardized sample means for a stochastic volatility model when Z comes from the standard Gaussian distribution ($\alpha = 2$).

We normalized (standardized) the means using the Relation (2.5.19). The result for the standardized sample mean paths are given in Figure 2.9. The behavior of the two sequences is close. The results in this figure show that the normalization is reasonable to get close to the limiting distribution. To discover if both cases have the same limiting distribution we simulated 10000 samples each of size 1000 observations for the Cauchy and normal cases. The boxplot for the statistic in equation (2.5.19) is given in Figure 2.10. The boxplot figure suggests that both limiting distributions are symmetric and have some similarity in their behavior but not the same distributions. Figure 2.11 introduces the densities for both distributions. In the case of the Gaussian distribution, the limiting density is close to the normal distribution. For the Cauchy case, the limiting density is symmetric around zero but not a Gaussian distribution. The QQ-plot emphasizes this idea. The following table summarizes the results for these samples:

Item	Cauchy	Gaussian
number of samples	10000	10000
sample size	1000	1000
mean	0.005887241	0.009034064
s.d.	0.9910633	0.9998287
median	0.01660911	0.009490259
1st. quantile	-0.88545755	-0.743342986
3rd quantile	0.89173822	0.764400447
minimum	-2.911422	-3.241606
maximum	2.830357	3.070718

Lemma 2.5.16. *The limit $S_\alpha/\sqrt{Y_{\frac{\alpha}{2}}}$ is symmetric and has unit variance. If (X_t) is strongly mixing, $E(S_\alpha/\sqrt{Y_{\frac{\alpha}{2}}})^4 < 3$, hence $S_\alpha/\sqrt{Y_{\frac{\alpha}{2}}}$ is not standard normal.*

Proof. As a matter of fact the limiting variable in (2.5.19) is symmetric as in (2.5.20). This follows from the fact that $X_i = \varepsilon_i|X_i|, i = 1, \ldots, n$ are independent symmetric, given $|X_i|,\ i = 1, \ldots, n$. Then

$$\frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}} = \frac{\varepsilon_1|X_1| + \cdots + \varepsilon_n|X_n|}{\sqrt{X_1^2 + \cdots + X_n^2}} \tag{2.5.21}$$

is a sum of independent random variables, conditionally on $(|X_i|)$. In turn, (2.5.21) is symmetric itself. We conclude that (2.5.21) has all odd moments zero, provided these moments exists. The same remains valid for the distributional limit (2.5.19) of (2.5.21) since the weak limits of symmetric probability measures are also symmetric.

Next we show that (2.5.21) has variance one:

$$\begin{aligned} E\frac{(X_1 + \cdots + X_n)^2}{X_1^2 + \cdots + X_n^2} &= E(\frac{X_1^2 + \cdots + X_n^2}{X_1^2 + \cdots + X_n^2}) + E(\frac{\sum_{1\le i\ne j\le n} \varepsilon_i|X_i|\varepsilon_j|X_j|}{X_1^2 + \cdots + X_n^2}) \\ &= 1 + E\sum_{1\le i\ne j\le n} E(\frac{\varepsilon_i\varepsilon_j|X_i||X_j|}{X_1^2 + \cdots + X_n^2}). \end{aligned}$$

The expectations on the right hand side exist and are finite since, by $2ab \le a^2 + b^2,\ a, b \in \mathbb{R}$,

$$\frac{1}{2}E\frac{2|X_i||X_j|}{X_1^2 + \cdots + X_n^2} \le \frac{1}{2}E\frac{X_i^2 + X_j^2}{X_1^2 + \cdots + X_n^2} \le \frac{1}{2}.$$

The random variables $\varepsilon_i\varepsilon_j$ are symmetric for $i \ne j$:

$$P(\varepsilon_j = \pm 1) = EP(\varepsilon_i\varepsilon_j = \pm 1|\varepsilon_i = \pm 1) = \frac{1}{2}. \tag{2.5.22}$$

Hence $\frac{\varepsilon_i\varepsilon_j|X_i||X_j|}{X_1^2+\cdots+X_n^2}$ has finite mean and is symmetric. This implies that

$$E\frac{(X_1 + \cdots + X_n)^2}{X_1^2 + \cdots + X_n^2} = 1.$$

An application of the Lebesgue dominated convergence theorem (switching from convergence in distribution to a.s. convergence on a suitable probability space) implies that

$$1 = E\frac{(X_1 + \cdots + X_n)^2}{X_1^2 + \cdots + X_n^2} \to E(\frac{S_\alpha^2}{Y_{\frac{\alpha}{2}}}) = 1.$$

We conclude that $S_\alpha/\sqrt{Y_{\frac{\alpha}{2}}}$ is a symmetric unit variance random variable.

However, $S_\alpha/\sqrt{Y_{\frac{\alpha}{2}}}$ is not standard normal. Indeed, if it were standard normal it had 4th moment 3. We show that $E(S_\alpha/\sqrt{Y_{\alpha/2}})^4 < 3$.

We have

$$\begin{aligned} E\frac{(X_1+\cdots+X_n)^4}{(X_1^2+\cdots+X_n^2)^2} &= E\left(\frac{X_1^2+\cdots+X_n^2+\sum_{1\le i\neq j\le n}X_iX_j}{X_1^2+\cdots+X_n^2}\right)^2 \\ &= 1+E\left(\frac{\sum_{1\le i\neq j\le n}X_iX_j}{X_1^2+\cdots+X_n^2}\right)^2+2E\frac{\sum_{1\le i\neq j\le n}X_iX_j}{X_1^2+\cdots+X_n^2}. \end{aligned} \quad (2.5.23)$$

The expectation of $E\left(\frac{\sum_{1\le i\neq j\le n}X_iX_j}{X_1^2+\cdots+X_n^2}\right)^2$ exists because

$$\frac{|X_iX_j|}{X_1^2+\cdots+X_n^2} \le \frac{1}{2} \quad \text{for } \textit{every } i\neq j,$$

by the same reasons as given before. The expectation of

$$E\Big(\frac{\sum_{1\le i\neq j\le n}X_iX_j}{X_1^2+\cdots+X_n^2}\Big)=0,$$

as shown above. We have

$$\begin{aligned} \Big(\sum_{1\le i\neq j\le n}X_iX_j\Big)^2 &= \Big(2\sum_{i=2}^{n}\sum_{j=1}^{i-1}X_iX_j\Big)^2 = 4\sum_{i=2}^{n}\sum_{j=1}^{i-1}X_iX_j\sum_{i'=2}^{n}\sum_{j'=1}^{i'-1}X_{i'}X_{j'} \\ &= 4\sum_{i=2}^{n}\sum_{j=1}^{i-1}(X_iX_j)^2+4\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{i'=2}^{n}\sum_{j'=1,(i,j)\neq(i',j')}^{i'-1}X_iX_jX_{i'}X_{j'}. \end{aligned} \quad (2.5.24)$$

We observe that the random variables

$$\frac{X_iX_jX_{i'}X_{j'}}{(X_1^2+\cdots+X_n^2)^2} = \frac{\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'}|X_i||X_j||X_{i'}||X_j'|}{(X_1^2+\cdots+X_n^2)^2} \quad (2.5.25)$$

are symmetric if $2\le i\le n,\ \ 1\le j\le i-1$ and $(i,j)\neq(i',j')$. Indeed, then it is excluded that $\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'}=\varepsilon_l^2\varepsilon_k^2$ for certain l,k and the random variable $\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'}$ is symmetric. Assume $i\neq i'$ then

$$EP(\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'}=\pm1|\varepsilon_{i'}=\pm1)=P(\varepsilon_i\varepsilon_j\varepsilon_{j'}=\pm1)$$

If $j=j'$, then $\varepsilon_j\varepsilon_{j'}=1$, hence the right hand side is $\frac{1}{2}$. If $j\neq j'$, then since $i\neq j$

$$EP(\varepsilon_i\varepsilon_j\varepsilon_{j'}=\pm1|\varepsilon_i=\pm1)=P(\varepsilon_j\varepsilon_{j'}=\pm1)=\frac{1}{2}$$

as proved in (2.5.22). We conclude that the random variables in (2.5.25) have mean zero.

We conclude from (2.5.23) and (2.5.24) that

$$E\frac{(X_1+\cdots+X_n)^2}{(X_1^2+\cdots+X_n^2)^2} = 1+4\sum_{i=1}^{n}\sum_{j=1}^{i-1}E\frac{(X_iX_j)^2}{(X_1^2+\cdots+X_n^2)^2}$$
$$= 1+2E\frac{(\sum_{i=1}^n X_i^2)^2-\sum_{i=1}^n X_i^4}{(X_1^2+\cdots+X_n^2)^2}$$
$$= 3-2E\frac{\sum_{i=1}^n X_i^4}{(X_1^2+\cdots+X_n^2)^2}. \quad (2.5.26)$$

We observe that

$$\frac{\sum_{i=1}^n X_i^4}{(X_1^2+\cdots+X_n^2)^2}\le 1. \quad (2.5.27)$$

Hence we are finished if we can show that

$$E\frac{\sum_{i=1}^n X_i^4}{(X_1^2+\cdots+X_n^2)^2}\to EA>0, \quad (2.5.28)$$

for some non-degenerate random variable A. Indeed, then

$$E\frac{(X_1+\cdots+X_n)^2}{(X_1^2+\cdots+X_n^2)^2}\to 3-2EA<3,$$

because $A\le 1$ $a.s.$ by (2.5.27). We observe that

$$E\frac{\sum_{i=1}^n X_i^4}{(X_1^2+\cdots+X_n^2)^2}=E\left(\frac{\sum_{i=1}^n N_iX_i^2}{\sum_{i=1}^n X_i^2}\right)^2, \quad (2.5.29)$$

where (N_i) is an iid $N(0,1)$ sequence independent of (X_i). Indeed, conditioning on (X_i) yields

$$EE\left(\left(\frac{\sum_{i=1}^n N_iX_i^2}{\sum_{i=1}^n X_i^2}\right)^2|(X_i)\right) = EE\left(\left(\frac{N_1(\sum_{i=1}^n X_i^4)^{\frac{1}{2}}}{\sum_{i=1}^n X_i^2}\right)^2|(X_i)\right)$$
$$= E(N_1^2)E\frac{\sum_{i=1}^n X_i^4}{(\sum_{i=1}^n X_i^2)^2}=E\frac{\sum_{i=1}^n X_i^4}{(\sum_{i=1}^n X_i^2)^2}$$

as desired.

On the other hand, the random variables $s_1N_iX_i^2+s_2X_i^2=(s_1N_i+s_2)X_i^2$ constitute a strictly stationary ergodic martingale sequence for every $s_1,s_2\in\mathbb{R}$. By Breiman's theorem and the fact that X_i is regularly varying with index $\alpha\in(0,2)$ (for more details about Breiman's theorem see Section 3.1.4), we have

$$P((s_1N_1+s_2)X_1^2>x)\sim E(s_1N_1+s_2)_+^{\alpha/2}P(X_1^2>x),$$

and

$$P((s_1N_1+s_2)X_1^2\le -x)\sim E(s_1N_1+s_2)_-^{\alpha/2}P(X_1^2>x),$$

where $x_\pm = \max(0, \pm x)$. Since X_1^2 is regularly varying with index $\alpha/2$, we conclude that $(s_1 N_1 + s_2) X_1^2$ is regularly varying with index $\alpha/2$. According to [19] and using the strong mixing property of (X_t), hence of $((s_1 N_t + s_2) X_t)$, the limit of

$$n^{-\frac{2}{\alpha}} \sum_{i=1}^{n} (s_1 N_i + s_2) X_i^2 \xrightarrow{d} Y_{\frac{\alpha}{2}}(s_1, s_2),$$

is an $\frac{\alpha}{2}-$stable random variable $Y_{\frac{\alpha}{2}}(s_1, s_2)$. Thus we proved by the Cramèr-Wold device that

$$\begin{pmatrix} s_1 & , & s_2 \end{pmatrix} n^{-\frac{2}{\alpha}} \sum_{i=1}^{n} \begin{pmatrix} N_i X_i^2 \\ X_i^2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} s_1 & , & s_2 \end{pmatrix} \underline{Y_{\frac{\alpha}{2}}},$$

for an $\frac{\alpha}{2}-$stable random vector $\underline{Y_{\frac{\alpha}{2}}}$ in $\mathbb{R}^2$. According to Samorodnitsky and Taqqu [37], $\underline{Y_{\frac{\alpha}{2}}}$ is $\frac{\alpha}{2}-$stable in $\mathbb{R}^2$, since a vector is $\frac{\alpha}{2}-$stable if and only if all its linear combinations are $\alpha-$stable. Now an application of the continuous mapping theorem yields

$$\frac{n^{-\frac{2}{\alpha}} \sum_{i=1}^{n} N_i X_i^2}{n^{-\frac{2}{\alpha}} \sum_{i=1}^{n} X_i^2} \xrightarrow{d} \frac{Y_{\frac{\alpha}{2}}(1)}{Y_{\frac{\alpha}{2}}(2)},$$

and by (2.5.29)

$$E\left(\frac{\sum_{i=1}^{n} X_i^4}{(X_1^2 + \cdots + X_n^2)^2}\right) \to E\left(\frac{Y_{\frac{\alpha}{2}}(1)}{Y_{\frac{\alpha}{2}}(2)}\right)^2.$$

Thus we proved that $E(S_\alpha / \sqrt{Y_{\frac{\alpha}{2}}})^4 < 3$. □

2.6 Asymptotic theory for the periodogram of a stochastic volatility sequence

The techniques used in analyzing stationary time series may be divided into two categories: time domain analysis and frequency domain analysis. The former deals with the observed data directly, as in conventional statistical analysis with independent observations. The frequency domain analysis, also called spectral analysis, applies the Fourier transform to the data (or ACVF) first, and the analysis proceeds with the transformed data only. The spectral analysis is in principle equivalent to the time domain analysis based on the ACVF. However, it provides an alternative way of viewing a process via decomposing it into a sum of uncorrelated periodic components with different frequencies, which for some applications may be more illuminating.

Theorem 2.6.1. (Herglotz) [9, p. 117–118]

The function $\gamma : \mathbb{Z} \to \mathbb{C}$ with $\gamma(-h) = \overline{\gamma(h)}$ is the autocovariance function of a stationary process if and only if there exists a right-continuous, non-decreasing, bounded function F on $[-\pi, \pi]$ such that $F(-\pi) = 0$ and

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ih\lambda} dF(\lambda) \quad h \in \mathbb{Z}. \tag{2.6.1}$$

The function F satisfying (2.6.1) is unique. However, notice that F is in general not a probability distribution function since $F(\pi) \neq 1$ is possible.

Definition 2.6.2. (Spectral distribution function of a stationary process)
Suppose that the stationary process (X_t) has an autocovariance function with representation

$$\gamma_X(h) = \int_{(-\pi,\pi]} e^{ih\lambda} dF_X(\lambda), \quad h \in \mathbb{Z} \tag{2.6.2}$$

where F_X is the right-continuous, non-decreasing, bounded function F_X on $[-\pi,\pi]$ with $F_X(-\pi) = 0$ corresponding to γ_X in Herglotz's theorem. The function F_X is called the spectral distribution function of the process (X_t), the corresponding measure the spectral distribution, and (2.6.2) *is the spectral representation of the autocovariance function γ_X. Moreover, if F_X is absolutely continuous with respect to Lebesgue measure then the corresponding density function f_X, i.e.,*

$$F_X(\lambda) = \int_{(-\pi,\pi]} f_X(x)dx, \quad \lambda \in (-\pi,\pi],$$

is called the spectral density of (X_t).

Corollary 2.6.3. [9, p. 119] *A complex-valued function $\gamma(\cdot)$ defined on the integers is the autocovariance function of a stationary process $(X_t)_{t\in\mathbb{Z}}$ if and only if either of the following conditions holds:*

- $\gamma(h) = \int_{(-\pi,\pi]} e^{ih\nu} dF(\nu)$ *for all $h = 0, \pm 1, \ldots$, where F is right-continuous, non-decreasing, bounded function on $[-\pi,\pi]$ with $F(-\pi) = 0$, or*
- $\sum_{i,j=1}^{n} a_i \gamma(i-j) \overline{a}_j \geq 0$ *for all positive integers n and for any $a = (a_1, \ldots, a_n)' \in \mathbb{C}^n$, $n \geq 1$.*

The spectral distribution function $F(\cdot)$ (and the corresponding spectral density if there is one) will be referred to as the spectral distribution function (and the spectral density) of both $\gamma(\cdot)$ and of (X_t).

Corollary 2.6.4. [9, p. 120] *An absolutely summable complex-valued function $\gamma(\cdot)$ defined on the integers is the autocovariance function of a stationary process if and only if*

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \geq 0 \quad \text{for all } \lambda \in [-\pi,\pi], \tag{2.6.3}$$

in which case $f(\cdot)$ is the spectral density of $\gamma(\cdot)$.

2.6.1 Estimation of the spectral density

Let (X_t) be a stationary sequence with spectral density f. Recall from Corollary 2.6.4 that the spectral density of (X_t) has representation

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma_X(n),$$

provided the autocovariances $\gamma_X(h)$ are absolutely summable. For a stochastic volatility model

$$X_t = \sigma_t Z_t,$$

where (σ_t) is a strictly stationary ergodic sequence, $\sigma_t > 0$, independent of the iid sequence (Z_t), the spectral density exists if and only if $var(X_0)$ is finite. If $EZ_t = 0$ then

$$\gamma_X(h) = cov(X_0, X_h) = 0, \quad h \neq 0.$$

From Equation (2.6.3), we have that (X_t) has spectral density

$$f_X(\lambda) = var(X_0)/2\pi, \quad \lambda \in [0, \pi].$$

It is natural to replace the autocovariances $\gamma_X(h)$ by their sample versions and to get an estimator of f_X in this way. Thus a natural (method of moment) estimator of $f_X(\lambda)$ is the periodogram given by

$$I_{n,X}(\lambda) = |a_n^{-1} \sum_{t=1}^{n} X_t e^{i\lambda t}|^2,$$

for a fixed frequency $\lambda \in (0, \pi)$. We study two cases

1. (Z_t) is iid $N(0, \sqrt{2})$.

2. (Z_t) is iid symmetric $\alpha-$stable ($S\alpha S$) for $\alpha < 2$, hence has infinite variance, see Section 2.5.2.

The two cases can be considered from a unified point of view since

$$Ee^{itZ_1} = e^{-|t|^\alpha}, \quad t \in \mathbb{R}, \quad \alpha \in (0, 2],$$

where we assumed in both cases that the scaling is in this standard way. The case $\alpha = 2$ corresponds to $N(0, \sqrt{2})$.

We choose the normalization $a_n = n^{\frac{1}{\alpha}}$, $n = 1, 2, \ldots$. Then

$$I_{n,X}(\pi\omega) = S_n^2(\omega) + C_n^2(\omega), \quad \omega \in (0, 1),$$

where

$$\begin{aligned} S_n(\omega) &= a_n^{-1} \sum_{t=1}^{n} X_t \cos(\pi\omega t), \\ C_n(\omega) &= a_n^{-1} \sum_{t=1}^{n} X_t \sin(\pi\omega t). \end{aligned}$$

The joint characteristic function of $S_n(\omega)$ and $C_n(\omega)$ is given by

$$f_n(s_1, s_2; \omega) = Ee^{is_1 S_n(\omega) + is_2 C_n(\omega)}, \quad (s_1, s_2) \in \mathbb{R}^2.$$

Then by stability of (Z_t):

$$\begin{aligned} f_n(s_1, s_2; \omega) &= E e^{i a_n^{-1} \sum_{t=1}^n Z_t \sigma_t (s_1 \cos(\pi\omega t) + s_2 \sin(\pi\omega t))} \\ &= E e^{i a_n^{-1} Z_1 (\sum_{t=1}^n \sigma_t^\alpha |s_1 \cos(\pi\omega t) + s_2 \sin(\pi\omega t)|^\alpha)^{\frac{1}{\alpha}}} \\ &= E e^{-\frac{1}{n} \sum_{t=1}^n \sigma_t^\alpha |s_1 \cos(\pi\omega t) + s_2 \sin(\pi\omega t)|^\alpha}. \end{aligned} \tag{2.6.4}$$

In what follows, we study the limit of

$$A_n = \frac{1}{n} \sum_{t=1}^n \sigma_t^\alpha |s_1 \cos(\pi\omega t) + s_2 \sin(\pi\omega t)|^\alpha.$$

We have

$$E A_n = \frac{E\sigma_0^\alpha}{n} \sum_{t=1}^n |s_1 \cos(\pi\omega t) + s_2 \sin(\pi\omega t)|^\alpha.$$

Write

$$\nu(x) = |s_1 \cos(\pi x) + s_2 \sin(\pi x)|^\alpha, \tag{2.6.5}$$

and

$$(s_1, s_2) = r(\cos(\pi\varphi), \sin(\pi\varphi)), \quad \varphi \in (0, 2].$$

Then

$$\begin{aligned} \nu(x) &= r^\alpha |\cos(\pi\varphi)\cos(\pi x) + \sin(\pi\varphi)\sin(\pi x)|^\alpha \\ &= r^\alpha |\cos(\pi(\varphi - x))|^\alpha. \end{aligned}$$

The function ν has period 1.

1. Assume $\omega \in (0,1)$ is rational, i.e., $\omega = q_1/q_2$ for positive relatively prime integers q_1, q_2. Then

$$\frac{1}{n}\sum_{t=1}^n \nu(\omega t) = \frac{1}{n}\sum_{t=1}^n \nu(\{\omega t\}) = \frac{1}{n}\sum_{t=1}^n \nu(\{\frac{q_1}{q_2} t\}), \tag{2.6.6}$$

   where $\{x\} = x - [x]$ denotes the fractional part of x.

   Let

$$n_k = \#\{t \le n : \frac{q_1}{q_2} k \equiv \frac{q_1}{q_2} t \mod 1\}, \quad n \to \infty.$$

   Then there exist r_k such that

$$\frac{n_k}{n} \to r_k, \quad k = 1, 2, 3, \ldots, \quad n \to \infty.$$

Let $N(\omega)$ denote those integers such that for any $k, k' \in N(\omega)$, $\frac{q_1}{q_2}k \not\equiv \frac{q_1}{q_2}k' \mod 1$. For $k \in N(\omega)$, $r_k = \frac{1}{\# N(\omega)}$. Hence

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\nu(\{\frac{q_1}{q_2}t\}) &\rightarrow \sum_{k\in N(\omega)} r_k \cos\nu(\{\frac{q_1}{q_2}k\}) \\ &= \frac{1}{\# N(\omega)}\sum_{k\in N(\omega)}\nu(\{\frac{q_1}{q_2}k\}) = E\nu(\{\frac{q_1}{q_2}U(\omega)\}), \qquad (2.6.7)\end{aligned}$$

where $U(\omega)$ is discretely uniform on $N(\omega)$.

2. Assume ω irrational. The same calculation as in (2.6.6) yields

$$\frac{1}{n}\sum_{t=1}^{n}\nu(\omega t) = \frac{1}{n}\sum_{t=1}^{n}\nu(\{\omega t\}).$$

By Weyl's theorem [40] the right hand side converges to

$$\int_0^1 \nu(u)du = E\nu(U),$$

for a $U(0,1)$ random variable U.

Thus we obtained:

$$\begin{aligned} EA_n &\rightarrow E\sigma_0^\alpha r^\alpha \begin{cases} E\nu(U), & \text{if } \omega \text{ is irrational;} \\ E\nu(\omega U(\omega)), & \text{if } \omega \text{ is rational.}\end{cases} \\ &= g(s_1, s_2; \omega),\end{aligned}$$

where $\nu(x)$ is defined in (2.6.5).

Next we prove that $var(A_n) \rightarrow 0$ under suitable conditions on (σ_t).

We observe that

$$\begin{aligned} var(A_n) &= \frac{r^{2\alpha}}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} cov(\sigma_t^\alpha, \sigma_s^\alpha)\nu(\omega t)\nu(\omega s) \\ &\leq \frac{c}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}|\gamma_{\sigma^\alpha}(t-s)| \\ &= \frac{c}{n} var(\sigma_0^\alpha) + \frac{2c}{n}\sum_{h=1}^{n-1}(1-\frac{h}{n})|\gamma_{\sigma^\alpha}(h)|,\end{aligned}$$

for some constant $c > 0$. The right hand side converges to zero if $\gamma_{\sigma^\alpha}(h) \rightarrow 0$ as $h \rightarrow \infty$. Thus we proved

$$A_n \stackrel{P}{\rightarrow} g(s_1, s_2; \omega), \quad \text{if } \gamma_{\sigma^\alpha}(h) \rightarrow 0 \text{ as } h \rightarrow \infty.$$

This, (2.6.4) and Lebesgue dominated convergence prove that

$$f_n(s_1, s_2; \omega) \rightarrow e^{-g(s_1, s_2; \omega)}.$$

The right hand side is the characteristic function of a vector (T_1, T_2) which is jointly $\alpha-$stable and symmetric. Such a characteristic function has representation

$$Ee^{i(s_1T_1+s_2T_2)} = e^{-\int_{\mathbb{S}} |s_1x_1+s_2x_2|^\alpha} d\Gamma(x_1, x_2), \tag{2.6.8}$$

for a measure Γ on the unit circle

$$\mathbb{S} = \{\underline{x} \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}.$$

See [37]. We have for irrational ω and $\alpha < 2$

$$\begin{aligned} e^{-g(s_1,s_2;\omega)} &= e^{-E\sigma_0^\alpha r^\alpha E|\cos\pi(\varphi-U)|^\alpha} \\ &= e^{-E\sigma_0^\alpha r^\alpha E|\cos\pi U|^\alpha} \\ &= e^{-E\sigma_0^\alpha(\sqrt{s_1+s_2})^\alpha \int_0^1 |\cos\pi u|^\alpha du} \\ &= Ee^{-(E\sigma_0^\alpha)^{\frac{2}{\alpha}}(\int_0^1|\cos\pi u|^\alpha du)^{\frac{2}{\alpha}} A(s_1^2+s_2^2)} \\ &= Ee^{-(E\sigma_0^\alpha)^{\frac{1}{\alpha}}(\int_0^1|\cos\pi u|^\alpha du)^{\frac{1}{\alpha}} A^{\frac{1}{2}}(N_1s_1+N_2s_2)}, \end{aligned} \tag{2.6.9}$$

where $A > 0$ is $\alpha/2-$stable (see Equation (2.5.15)), independent of the iid $N(0, \sqrt{2})$ random variables N_1, N_2. A random vector with representation

$$A^{\frac{1}{2}}(N_1, N_2),$$

is known to be a stable sub-Gaussian $S\alpha S$ vector, see [37].

We have for a rational ω

$$e^{-g(s_1,s_2;\omega)} = e^{-E\sigma_0^\alpha E|s_1\cos(\pi\omega U(\omega))+s_2\sin(\pi\omega U(\omega))|^\alpha},$$

which again has the form in Equation (2.6.8).

The case $\alpha = 2$ deserves special attention. In this case, for ω irrational in (2.6.9)

$$e^{-g(s_1,s_2;\omega)} = e^{-E\sigma_0^2(s_1^2+s_2^2)\int_0^1\cos^2(\pi u)du} = e^{-\frac{E\sigma_0^2}{2}(s_1^2+s_2^2)}.$$

Hence the limit is of the form (N_1, N_2) for iid $N(0, \sigma_0^2)$ random variables N_1, N_2. For rational ω the same result applies since

$$E(s_1\cos(\pi\omega U(\omega)) + s_2\sin(\pi\omega U(\omega)))^2 = \frac{s_1^2}{2} + \frac{s_2^2}{2},$$

where we used the orthogonality of $\cos(\pi\omega U(\omega))$, $\sin(\pi\omega U(\omega))$.

We summarize our results.

Proposition 2.6.5. *Assume that (X_t) is a strictly stationary ergodic stochastic volatility model with iid $S\alpha S$ noise (Z_t) for some $\alpha < 2$ and finite variance (σ_t). Then*

$$I_{n,X}(\lambda) \xrightarrow{d} T_1^2 + T_2^2, \tag{2.6.10}$$

where (T_1, T_2) is jointly $S\alpha S$ and the distribution of (T_1, T_2) depends on whether $\omega = \lambda/\pi$ is rational or irrational. If ω is irrational, $T_1^2 + T_2^2 \stackrel{d}{=} A(N_1^2 + N_2^2)$, where A is $\frac{\alpha}{2}$-stable positive, independent of the iid $N(0,1)$ random variables $N_1,\ N_2$.

2.6.2 Self-normalization of the periodogram for $\alpha < 2$

Since the normalization a_n in the definition of the periodogram is in general unknown it is reasonable to replace it by a known normalization. We have observed in Section 2.6 that

$$n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} X_t^2 \xrightarrow{d} Y_{\alpha/2},$$

for a positive $\alpha/2-$stable random variable. Since $a_n^2 = n^{2/\alpha}$, it seems reasonable to replace a_n^2 by $\sum_{t=1}^{n} X_t^2$. In order to show this, one has to prove that there is joint convergence:

$$\left(a_n^{-2} \sum_{t=1}^{n} X_t^2 \quad , \quad I_{n,\alpha}(\lambda)\right) \xrightarrow{d} \left(R_1 \quad , \quad R_2\right),$$

for non-degenerate random variables R_1, R_2. This is achieved if we can show

$$\left(a_n^{-2} \sum_{t=1}^{n} X_t^2 \quad , \quad S_n(\omega) \quad , \quad C_n(\omega)\right) \xrightarrow{d} \left(R_1 \quad , \quad R_{2S} \quad , \quad R_{2C}\right),$$

for non-degenerate random variables R_1, R_{2S}, R_{2C}, such that $R_2 = R_{2S}^2 + R_{2C}^2$.

We consider the joint Laplace-Stieltjes transform-characteristic function

$$\begin{aligned}
f_n(s_1, s_2, s_3; \omega) &= E\exp(-\frac{s_1^2}{2} a_n^{-2} \sum_{t=1}^{n} X_t^2 + is_2 S_n(\omega) + is_3 C_n(\omega)) \\
&= E\exp(is_1 a_n^{-1} \sum_{t=1}^{n} N_t X_t + is_2 S_n(\omega) + is_3 C_n(\omega)) \\
&= E\exp(in^{-\frac{1}{\alpha}} \sum_{t=1}^{n} Z_t \sigma_t (s_1 N_t + \cos(\pi\omega t) s_2 + \sin(\pi\omega t) s_3)) \\
&= E\exp\left(in^{-\frac{1}{\alpha}} Z_1 (\sum_{t=1}^{n} \sigma_t^\alpha |s_1 N_t + \cos(\pi\omega t) s_2 + \sin(\pi\omega t) s_3|^\alpha)^{\frac{1}{\alpha}}\right) \\
&= E\exp(-n^{-1} \sum_{t=1}^{n} \sigma_t^\alpha |s_1 N_t + \cos(\pi\omega t) s_2 + s_3 \sin(\pi\omega t)|^\alpha),
\end{aligned} \tag{2.6.11}$$

where (N_t) is an iid $N(0,1)$ sequence independent of $(Z_t), (\sigma_t)$.

Write

$$\beta_n = \frac{1}{n} \sum_{t=1}^{n} \sigma_t^\alpha |s_1 N_t + \cos(\pi\omega t) s_2 + s_3 \sin(\pi\omega t)|^\alpha.$$

Then

$$E\beta_n = \frac{1}{n}\sum_{t=1}^{n} E|s_1 N_1 + s_2\cos(\pi\omega t) + s_3\sin(\pi\omega t)|^\alpha.$$

Write

$$w(x) = E|s_1 N_1 + s_2\cos(\pi x) + s_3\sin(\pi x)|^\alpha.$$

Notice that w has period 1. Therefore

$$w(x) = E|s_1 N_1 + s_2\cos(\pi\{x\}) + s_3\sin(\pi\{x\})|^\alpha.$$

The same reasons as for EA_n above yield that

$$\begin{aligned} E\beta_n \quad &\to \quad \begin{cases} E|s_1 N_1 + s_2\cos(\pi U) + s_3\sin(\pi U)|^\alpha, & \text{for } \omega\in(0,1) \text{ irrational;} \\ E|s_1 N_1 + s_2\cos(\pi\omega U(\omega)) + s_3\sin(\pi\omega U(\omega))|^\alpha, & \text{for } \omega\in(0,1) \text{ rational.}\end{cases} \\ &= g(s_1,s_2,s_3;\omega) \end{aligned} \tag{2.6.12}$$

Here U is $U(0,1)$ independent of N_1 and $U(\omega)$ is defined in Equation (2.6.7) and independent of N_1.

We also have for some constant $c>0$

$$\begin{aligned} var(\beta_n) \quad &= \quad n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} w(\{\omega t\})w(\{\omega s\})cov(\sigma_t^\alpha,\sigma_s^\alpha) \\ &\le \quad cn^{-1}var(\sigma_0^\alpha) + \frac{1}{n}c2\sum_{h=1}^{n-1}(1-\frac{h}{n})|\gamma_{\sigma^\alpha}(h)| \\ &\quad\ \to 0 \quad \text{as } n\to\infty, \text{ if } \gamma_{\sigma^\alpha}(h)\to 0 \text{ as } h\to\infty. \end{aligned}$$

Hence we have

$$\beta_n \xrightarrow{P} g(s_1,s_2,s_3;\omega), \quad \text{if } \gamma_{\sigma^\alpha}(h)\to 0 \text{ as } h\to\infty.$$

By Lebesgue dominated convergence and relation (2.6.11),

$$f_n(s_1,s_2,s_3;\omega)\to e^{-g(s_1,s_2,s_3;\omega)},$$

for any $(s_1,s_2,s_3)\in\mathbb{R}^3$. This shows the joint convergence of $\big(n^{-\frac{2}{\alpha}}\sum_{t=1}^n X_t^2 \quad, S_n(\omega) \quad, C_n(\omega)\big)$, to $\big(R_1 \quad, R_{2S} \quad, R_{2C}\big)$ and hence the continuous mapping theorem implies that

$$\widetilde{I}_{n,X}(\lambda) = \frac{I_{n,X}(\lambda)}{n^{-\frac{2}{\alpha}}\sum_{t=1}^n X_t^2} = \frac{|\sum_{t=1}^n X_t e^{i\lambda t}|^2}{\sum_{t=1}^n X_t^2} = \frac{S_n^2(\omega)+C_n^2(\omega)}{n^{-\frac{2}{\alpha}}\sum_{t=1}^n X_t^2} \xrightarrow{d} \frac{R_{2S}^2+R_{2C}^2}{R_1^2}.$$

For $\omega=\lambda/\pi$ irrational we know from Section 2.6.1 that

$$R_{2S}^2+R_{2C}^2 \stackrel{d}{=} cA(N_1^2+N_2^2),$$

for some constant $c > 0$, A positive $\alpha/2-$stable, independent of N_1, N_2. Up to a constant multiple, R_1^2 has the same distribution as A, but A and R_1^2 are dependent. Therefore, we have for ω irrational

$$\widetilde{I}_{n,X}(\lambda) = \frac{|\sum_{t=1}^{n} X_t e^{i\lambda t}|^2}{\sum_{t=1}^{n} X_t^2} \xrightarrow{d} \frac{cA}{R_1^2}(N_1^2 + N_2^2).$$

In the case $\alpha = 2$, i.e., for iid $N(0,1)$ noise variables Z_t, both A and R_1^2 degenerate to constants. For A this was explained above and for R_1^2 this is a consequence of the strong law of large numbers and the ergodicity of (X_t). We also conclude that $\frac{N_1^2+N_2^2}{2}$ is exponentially $Exp(1)$ distributed. Therefore, the periodogram of a stochastic volatility model at irrational $\omega = \lambda/\pi$ has the same limits as in the case of iid X_t, in the case $\alpha < 2$ see [23] and for $\alpha = 2$ see [9].

As in the case of linear processes (X_t) it seems feasible to show that smoothed versions of the standardized periodogram $\widetilde{I}_{n,X}(\lambda)$ are consistent estimators of the constant spectral density provided $var(X_t) < \infty$ and the same result might apply for regularly varying X_t with $\alpha < 2$. In order to show this one needs to consider the limit distribution of $(\widetilde{I}_{n,X}(\lambda_h))_{h=1,\ldots,m}$ at frequencies $0 < \lambda_1 < \ldots < \lambda_k < \pi$, where $\lambda_k = \lambda_k(m) \to \lambda$ as $m \to \infty$ and $k = k(m) \to \infty$ as $m \to \infty$. This was shown in [23] for infinite variance linear processes (X_t) and the case of finite variance linear processes (X_t) can be found in [9].

Example 2.6.6. We simulated 1000 observations from two different samples of stochastic volatility models in the same way as in Example 2.5.15. We estimated the spectral density by using the raw periodogram for the two samples. The results are shown in Figure 2.12. In a later step, we standardized the observations in the two samples using the transformation $\sqrt{n}X/\widehat{\gamma}_X(0)$. We again estimated the spectral density for these two samples of standardized samples. The result of the two estimated spectral densities are close to each other, see Figure 2.12.

2.7 Asymptotic theory for the sample ACVF and ACF of a stochastic volatility process

2.7.1 Asymptotic theory for the sample ACVF and ACF for (X_t)

This section will deal with the asymptotic behavior of the sample ACVFs and ACFs for the sequence (X_t). Recall that the sample ACVF and ACF of any stationary process (A_t) are given by

$$\hat{\gamma}_A(h) = \frac{1}{n}\sum_{t=1}^{n-h}(A_t - \bar{A}_n)(A_{t+h} - \bar{A}_n), \quad \hat{\rho}_A(h) = \frac{\hat{\gamma}_A(h)}{\hat{\gamma}_A(0)}, \quad 0 \le h < n, \tag{2.7.1}$$

respectively, where $\bar{A}_n = \frac{1}{n}\sum_{t=1}^{n} A_t$ is the sample mean.

We will always assume (X_t) is a stochastic volatility process with specification given in Section 2.4.3, in particular $EZ = 0$ and $EZ^2 = 1$, and (σ_t) is defined through equation (2.4.1), where (η_t) is an iid sequence

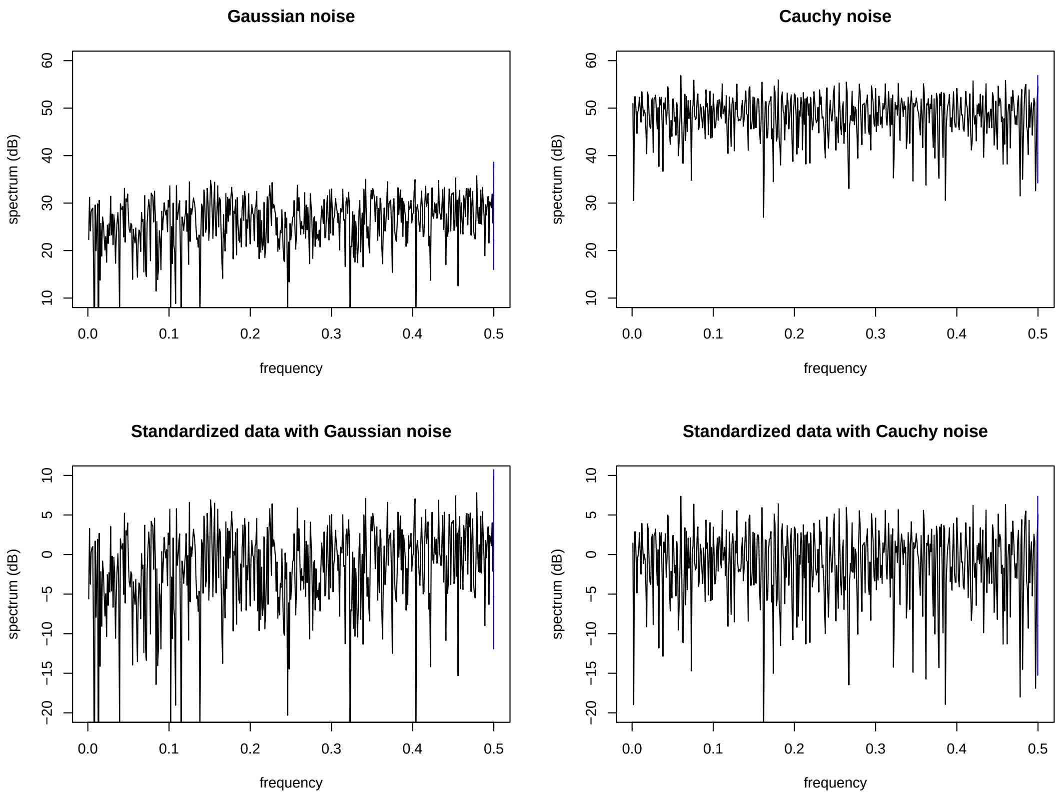

Figure 2.12: (Example 2.6.6) Let (Q_t) be a stochastic volatility model given by $Q_t = Z_t\sigma_t$ where $\log(\sigma_t) = V_t$ is an ARMA(1,1) process given by $V_t = 0.7V_{t-1} + 0.7\delta_{t-1} + \delta_t$, for iid (δ_t) with common distribution $N(0,1)$. Top: Left: Estimated spectral log-density of 1000 observations from a stochastic volatility model when Z comes from the Gaussian distribution ($\alpha = 2$). Right: Estimated spectral log-density of 1000 observations from a stochastic volatility model when Z comes from the Cauchy distribution ($\alpha = 1$). Bottom: Left: Estimated spectral log-density of 1000 standardized observations from a stochastic volatility model when Z comes from the Gaussian distribution ($\alpha = 2$). Right: Estimated spectral log-density of 1000 standardized observations from a stochastic volatility model when Z comes from the Cauchy distribution ($\alpha = 1$). Note that the λ's on the x-axis correspond to the frequency $2\pi\lambda$.

with $E(\eta) = 0, var(\eta) = E(\eta_t^2) = \tau^2 < \infty$ and the moment generating function of η, $m_\eta(s) = Ee^{s\eta}$, is finite for all $s \in \mathbb{R}$, $\sum_{i=0}^{\infty}\psi_i^2 < \infty$ and $\psi_0 = 1$.

Lemma 2.7.1. *If (X_t) is a stochastic volatility sequence as specified above. Then $\widehat{\gamma}_X(h) \to 0$ a.s. for any $h > 0$ and*

$$\sqrt{n}\,\widehat{\gamma}_X(h) \xrightarrow{d} N(0, E(\sigma_0^2\sigma_h^2)), \qquad h > 0, \tag{2.7.2}$$

where

$$E(\sigma_0^2\sigma_h^2) = \prod_{i=0}^{\infty} m_\eta(2(\psi_i + \psi_{i+h})) \prod_{i=0}^{h-1} m_\eta(2\psi_i).$$

Proof. We start with the decomposition

$$\begin{aligned}\widehat{\gamma}_X(h) &= (\frac{n-h}{n})(\frac{1}{n-h})\sum_{t=1}^{n-h}(X_t-\overline{X}_n)(X_{t+h}-\overline{X}_n), \\ &= (1-\frac{h}{n})\left(\frac{1}{n-h}\sum_{t=1}^{n-h}X_tX_{t+h}-(\overline{X}_n)^2\right). \end{aligned} \tag{2.7.3}$$

Moreover, (X_t) and (X_tX_{t+h}) are strictly stationary ergodic sequences. and

$$\begin{aligned} E|X| &= E\sigma E|Z| < \infty, \\ E|X_0X_h| &= E(\sigma_0\sigma_h)(E|Z|)^2 < \infty. \end{aligned}$$

The SLLN and (2.7.3) imply that

$$\widehat{\gamma}_X(h) \to EX_0X_h - (EX)^2 = 0 \ a.s.$$

Similarly,

$$\begin{aligned} &\sqrt{n-h}(\frac{1}{n-h}\sum_{t=1}^{n-h}X_tX_{t+h}-(\overline{X}_n)^2) \\ &= \frac{1}{\sqrt{n-h}}\sum_{t=1}^{n-h}X_tX_{t+h}-\frac{\sqrt{n-h}}{n}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_t)^2. \end{aligned}$$

The sequences (X_t) and (X_tX_{t+h}) constitute mean zero finite variance strictly stationary ergodic martingale difference sequences. This was proved in Lemma 2.4.3. Hence the CLT in Theorem 2.1.15 applies to $\frac{1}{n}\sum_{t=1}^{n}X_t$ and therefore

$$\frac{\sqrt{n-h}}{n}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_t)^2 \xrightarrow{P} 0.$$

An application of the CLT (see Theorem 2.1.15) yields that

$$\frac{1}{\sqrt{n-h}}\sum_{t=1}^{n-h}X_tX_{t+h} \xrightarrow{d} N(0, var(X_0X_h)),$$

where $var(X_0X_h) = E(\sigma_0^2\sigma_h^2)$. This proves the statement. □

Using the multivariate central limit theorem for a strictly stationary ergodic martingale difference in Lemma 2.1.18, we can obtain a multivariate version of Lemma 2.7.1. As in the proof of Lemma 2.7.1, we have

$$\begin{pmatrix} \frac{1}{\sqrt{n-1}}\hat{\gamma}_X(1) \\ \frac{1}{\sqrt{n-2}}\hat{\gamma}_X(2) \\ \vdots \\ \frac{1}{\sqrt{n-m}}\hat{\gamma}_X(m) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n-1}}\sum_{t=1}^{n-1}X_tX_{t+1} \\ \frac{1}{\sqrt{n-2}}\sum_{t=1}^{n-2}X_tX_{t+2} \\ \vdots \\ \frac{1}{\sqrt{n-m}}\sum_{t=1}^{n-m}X_tX_{t+m} \end{pmatrix} + o_p(1). \tag{2.7.4}$$

Now apply the multivariate central limit theorem (Lemma 2.1.18) to the strictly stationary ergodic martingale difference sequence $((X_tX_{t+1}, X_tX_{t+2}, \ldots, X_tX_{t+m})')_{t\in\mathbb{Z}}$. According to Lemma 2.1.18 the asymptotic variance in the central limit theorem is given by

$$\Sigma = cov(((X_0X_i),(X_0X_j)))_{i,j=1,\ldots,m}.$$

By the construction of a stochastic volatility sequence, the covariances $cov((X_0X_i),(X_0X_j)),\ \ i \neq j$ vanish. This simplifies the structure of Σ significantly to

$$\Sigma = diag(var(X_0X_1),\cdots,var(X_0X_m)) = diag(E(\sigma_0^2\sigma_1^2),\ldots,E(\sigma_0^2\sigma_m^2)). \tag{2.7.5}$$

Proposition 2.7.2. *If (X_t) is a stochastic volatility sequence satisfying the conditions given at the beginning of the section, we have*

$$\sqrt{n}\ (\widehat{\gamma}_X(h))_{h=1,\ldots,m} \xrightarrow{d} N(\underline{0},\Sigma),$$

where Σ is given in (2.7.5).

Proposition 2.7.2 helps one to prove the central limit theorem for the sample ACF $\hat{\rho}_X(h)$. We start with the result for one lag $h > 0$.

Lemma 2.7.3. *If (X_t) is a stochastic volatility sequence satisfying the conditions given at the beginning of the section then*

$$\widehat{\rho}_X(h) \to 0 \quad a.s., \quad h > 0,$$

and the central limit theorem for the sample ACF at lag h of (X_t) holds:

$$\sqrt{n}\ \hat{\rho}_X(h) \xrightarrow{d} N(0,\nu^2), \tag{2.7.6}$$

where

$$\nu^2 = \frac{E(\sigma_0^2\sigma_h^2)}{\gamma_X^2(0)} = \frac{\prod_{i=0}^{\infty} m_\eta(2(\psi_i+\psi_{i+h}))\prod_{i=0}^{h-1} m_\eta(2\psi_i)}{(\prod_{i=0}^{\infty} m_\eta(2\psi_i))^2}.$$

In particular if η has a Gaussian $N(0,\tau^2)$ distribution then ν^2 has the form

$$\nu^2 = e^{4\tau^2\sum_{i=0}^{\infty}\psi_i\psi_{i+h}} = e^{4\gamma_Y(h)}. \tag{2.7.7}$$

Proof. Lemma 2.7.1 yields that $\widehat{\gamma}_X(h) \to 0$ *a.s.* Moreover, the ergodic theorem applied to (X_i^2) yields

$$\widehat{\gamma}_X(0) = \frac{1}{n}\sum_{t=1}^{n} X_t^2 - (\overline{X}_n)^2 \to EX^2 - (EX)^2 = \gamma_X(0)\ a.s.$$

Hence we have

$$\widehat{\rho}_X(h) \to 0 \;\; a.s., \quad h > 0,$$

and

$$\sqrt{n}\;\hat{\rho}_X(h) \;\;=\;\; \frac{\sqrt{n}\hat{\gamma}_X(h)}{\gamma_X(0)(1+o_p(1))}. \tag{2.7.8}$$

We know from Lemma 2.7.1 that

$$\sqrt{n}\widehat{\gamma}_X(h) \xrightarrow{d} N(0, E(\sigma_0^2\sigma_h^2)).$$

Hence the continuous mapping theorem applies to (2.7.8) and (2.7.6) follows. If η is $N(0,\tau^2)$ we apply Lemma 2.4.7. This yields that ν^2 is given by (2.7.7)

$$\nu^2 = \frac{e^{4\tau^2\sum_{i=0}^{\infty}(\psi_i^2+\psi_i\psi_{i+h})}}{e^{4\tau^2\sum_{i=0}^{\infty}\psi_i^2}}.$$

□

Following the lines of the proof of Lemma 2.7.3 and using Proposition 2.7.2, we conclude that

$$\sqrt{n}(\widehat{\rho}_X(h))_{h=1,\dots,m} = \frac{\sqrt{n}(\widehat{\gamma}_X(h))_{h=1,\dots,m}}{\gamma_X(0)}(1+o_p(1)).$$

Now the following proposition gives the central limit theorem for the sample ACF of a stochastic volatility sequence as a consequence of Proposition 2.7.2.

Proposition 2.7.4. *If* (X_t) *is a stochastic volatility sequence satisfying the conditions given at the beginning of the section, we have*

$$\sqrt{n}(\widehat{\rho}_X(h))_{h=1,\dots,m} \xrightarrow{d} N(\underline{0}, \frac{1}{(E(\sigma^2))^2}\Sigma),$$

where Σ *is given in* (2.7.5).

2.7.2 Asymptotic theory for the sample ACVF and ACF for $(|X_t|^p)$, $0 < p < \infty$

In this section we will consider the sample ACVF and ACF for $(|X|^p)$ for any $p > 0$. We assume that the sequence (σ_t) is a strongly mixing sequence with rate function (α_t). Then the sequence $(|X_t|^p)$ inherits the strong mixing property with the same rate function, see page 17. Under conditions on the rate of decay of (α_t), the central limit theorem for strictly stationary strongly mixing sequences can be applied to $(|X_t|^p)$. The sequence $(\sigma_t^p\sigma_{t+h}^p)$ constitutes a strictly stationary strongly mixing sequence with the same rate function as (σ_t) and $(|X_t|^p)$. The central limit theorem (see Theorem 2.1.16) for this sequence is given by:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\sigma_t^p\sigma_{t+h}^p - E(\sigma_0^p\sigma_h^p)) \xrightarrow{d} N(0, var(\sigma_0^p\sigma_h^p) + 2\sum_{i=1}^{\infty}\gamma_{\sigma_0^p\sigma_h^p}(i)), \quad h > 0. \tag{2.7.9}$$

Here we assumed that $E(\sigma_0\sigma_h)^{2p+\delta} < \infty$ for some $\delta > 0$ and $\sum_{h=0}^{\infty}\alpha_h^{\delta/(2+\delta)} < \infty$.

In addition, the sequence $(|X_tX_{t+h}|^p - E(\sigma_0\sigma_h)^p(E|Z|^p)^2)$ is also a strictly stationary strongly mixing sequence with the same rate function as (σ_t). Again the central limit theorem of Theorem 2.1.16 applies:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|X_tX_{t+h}|^p - E(\sigma_0\sigma_h)^p(E|Z|^p)^2) \xrightarrow{d} N(0, \upsilon^2), \quad h > 0, \tag{2.7.10}$$

where

$$\upsilon^2 = var(|X_0X_h|^p) + 2\sum_{i=1}^{\infty}\gamma_{|X_0X_h|^p}(i).$$

Here we again assumed that $E(\sigma_0\sigma_h)^{2p+\delta} < \infty$ for some $\delta > 0$ and $\sum_{h=0}^{\infty}\alpha_h^{\delta/(2+\delta)} < \infty$.

In what follows, we make precise under which conditions these central limit theorems hold. The following lemma will be useful.

Lemma 2.7.5. *Let (A_t) be a strictly stationary sequence such that (A_t) satisfies the CLT $\sqrt{n}(\overline{A}_n - EA) \xrightarrow{d} N(0, \nu^2)$ for some $\nu > 0$ and $var(A) < \infty$. Then*

$$\Delta_n = \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n-h}(A_t - EA)(A_{t+h} - EA) - \sum_{t=1}^{n-h}(A_t - \overline{A}_n)(A_{t+h} - \overline{A}_n)\right) \xrightarrow{P} 0 \tag{2.7.11}$$

Proof. We assume without loss of generality that $EA = 0$. We can decompose Δ_n as follows

$$\begin{aligned}\Delta_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n-h}A_tA_{t+h} - \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n-h}A_tA_{t+h} - \overline{A}_n\sum_{t=1}^{n-h}A_{t+h} - \overline{A}_n\sum_{t=1}^{n-h}A_t + (n-h)(\overline{A}_n)^2\right)\\ &= \sqrt{n}(\overline{A}_n)^2 + o_p(1) \xrightarrow{P} 0.\end{aligned}$$

□

Proposition 2.7.6. *Let $p > 0$, $E|Z|^{2p+\delta} < \infty$ for some $\delta > 0$. Assume that (X_t) is a stochastic volatility sequence with specification at the beginning of Section 2.7.1. Moreover, assume that the volatility sequence (σ_t) is strongly mixing with rate function (α_t) satisfying $\sum_{t=1}^{\infty}\alpha_t^{\frac{\delta}{2+\delta}} < \infty$ for some $\delta > 0$. Then the central limit theorem for the sample ACVF for the sequence $(|X_t|^p)$ is given by*

$$\sqrt{n}(\hat{\gamma}_{|X|^p}(h) - \gamma_{|X|^p}(h)) \xrightarrow{d} N(0, \nu^2(h)), \tag{2.7.12}$$

where

$$\begin{aligned}\nu^2(h) &= var(|X_0X_h|^p) + 4(E|X|^p)^2 var(|X|^p) - 4E|X|^p cov(|X_0X_h|^p, |X_h|^p) \\ &+8(E|X|^p)^2 \sum_{t=1}^{\infty} \gamma_{|X|^p}(t) - 4E|X|^p \sum_{t=1}^{\infty} cov(|X_0X_h|^p, |X_{t+h}|^p) \\ &+2\sum_{t=1}^{\infty} \gamma_{|X_0X_h|^p}(t) - 4E|X|^p \sum_{t=1}^{\infty} cov(|X_tX_{t+h}|^p, |X_h|^p) \qquad (2.7.13)\end{aligned}$$

Proof. By assumption, (σ_t), hence $(|X_t|^p)$ is strongly mixing with rate function (α_t) satisfying $\sum_{t=1}^{\infty} \alpha_t^{\frac{\delta}{2+\delta}} < \infty$ for some $\delta > 0$ and so is $((|X_t|^p - E|X|^p)(|X_{t+h}|^p - E|X|^p))$. The central limit theorem for strongly mixing sequences in Theorem 2.1.16 implies that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n-h} [(|X_t|^p - E|X|^p)(|X_{t+h}|^p - E|X|^p) - \gamma_{|X|^p}(h)] \xrightarrow{d} N(0, \nu^2(h)). \qquad (2.7.14)$$

By an application of Lemma 2.7.5 one may replace $E|X_t|^p$ in (2.7.14) by $\frac{1}{n}\sum_{t=1}^{n} |X_t|^p$. This concludes the proof for one lag $h > 0$. □

A multivariate central limit theorem for the vector $(\gamma_{|X|^p}(h))_{h=1,\dots,m}$ also holds.

Proposition 2.7.7. *Under the same assumptions in Proposition 2.7.6 a central limit theorem for*

$$(\widehat{\gamma}_{|X|^p}(h))_{h=1,\cdots,m},$$

is given by

$$\sqrt{n}\begin{pmatrix} (\widehat{\gamma}_{|X|^p}(1) - \gamma_{|X|^p}(1)) \\ (\widehat{\gamma}_{|X|^p}(2) - \gamma_{|X|^p}(2)) \\ \vdots \\ (\widehat{\gamma}_{|X|^p}(m) - \gamma_{|X|^p}(m)) \end{pmatrix} \xrightarrow{d} N(\underline{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \nu^2(1) & \Lambda(1,2) & \cdots & \Lambda(1,m) \\ \Lambda(1,2) & \nu^2(2) & \cdots & \Lambda(2,m) \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda(1,m) & \Lambda(2,m) & \cdots & \nu^2(m) \end{pmatrix},$$

$$\begin{aligned}\Lambda(0,h) &= cov((|X_0X_h|^p - 2E|X|^p|X_h|^p), (|X_0|^{2p} - 2E|X|^p|X_0|^p)) \\ &+2\sum_{t=1}^{\infty} cov((|X_tX_{t+h}|^p - 2E|X|^p|X_{t+h}|^p), (|X_0|^{2p} - 2E|X|^p|X_0|^p)) \\ &= cov(|X_0X_h|^p, |X_0|^{2p}) - 2E|X|^p cov(|X_h|^p, |X_0|^{2p}) - 2E|X|^p cov(|X_0|^p, |X_0X_h|^p) \\ &+4(E|X|^p)^2\gamma_{|X|^p}(h) + 2\sum_{t=1}^{\infty} cov(|X_tX_{t+h}|^p, |X_0|^{2p}) - 4E|X|^p \sum_{t=1}^{\infty} cov(|X_{t+h}|^p, |X_0|^{2p}) \\ &-4E|X|^p \sum_{t=1}^{\infty} cov(|X_0|^p, |X_tX_{t+h}|^p) + 8(E|X|^p)^2 \sum_{t=1}^{\infty} cov(|X_{t+h}|^p, |X_0|^p),\end{aligned}$$

and, slightly abusing notation,

$$\Lambda(h-l) = \Lambda(0, |h-l|) = \Lambda(h, l).$$

The variance $\nu^2(h)$ *is given by* (2.7.13).

Proof. The central limit theorem for $(\widehat{\gamma}_{|X|^p}(h))_{h=1,\cdots,m}$ is analogous to the proof of Proposition 2.7.6 by observing that

$$\Big(\big((|X_t|^p - E|X|^p)(|X^p_{t+h} - E|X|^p)\big)_{h=1,\cdots,m}\Big)_{t\in\mathbb{Z}},$$

is strongly mixing sequence with rate functions (α_h). An application of Lemma 2.1.19 yields the central limit theorem for

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}(|X_t|^p - E|X|^p)(|X_{t+1}|^p - E|X|^p)\\ (|X_t|^p - E|X|^p)(|X_{t+2}|^p - E|X|^p)\\ \vdots\\ (|X_t|^p - E|X|^p)(|X_{t+m}|^p - E|X|^p)\end{pmatrix}.$$

Lemma 2.7.5 again allows one to replace $E|X|^p$ by $\frac{1}{n}\sum_{t=1}^{n}|X_t|^p$. This concludes the proof. □

Let us have a look at the sample ACF for the sequence $(|X_t|^p)$. This sequence as mentioned before is a strictly stationary strongly mixing sequence. The following lemma gives a central limit theorem for the sample ACF of the stochastic volatility model.

Lemma 2.7.8. *Assume the conditions in Proposition 2.7.6 and* $E(\sigma_0\sigma_h)^{2p+\delta} < \infty$ *and* $E|Z|^{2p+\delta} < \infty$ *for some* $\delta > 0$*. Let* $\nu^2(h)$ *be the variance of the limiting distribution in Proposition 2.7.6 given by formula* (2.7.13)*,* λ^2 *the variance of the limiting distribution in Lemma 2.5.8, i.e.*

$$\sqrt{n}(\widehat{\gamma}_{|X|^p}(0) - \gamma_{|X|^p}(0)) \xrightarrow{d} N(0, \lambda^2),$$

and $\Lambda = \Lambda(0, h)$ *is given in Proposition 2.7.7. Then*

$$\sqrt{n}(\hat{\rho}_{|X|^p}(h) - \rho_{|X|^p}(h)) \xrightarrow{d} N\left(0, \frac{1}{(\gamma_{|X|^p}(0))^2}\nu^2(h) + 2\frac{\gamma_{|X|^p}(h)}{(\gamma_{|X|^p}(0))^3}\Lambda + \frac{(\gamma_{|X|^p}(h))^2}{(\gamma_{|X|^p}(0))^4}\lambda^2\right). \tag{2.7.15}$$

Proof. The difference between the ACF and the estimated ACF at lag h can be decomposed as:

$$\begin{aligned}&\sqrt{n}(\hat{\rho}_{|X|^p}(h) - \rho_{|X|^p}(h))\\ &= \sqrt{n}\frac{(\hat{\gamma}_{|X|^p}(h) - \gamma_{|X|^p}(h))\gamma_{|X|^p}(0) - \gamma_{|X|^p}(h)(\hat{\gamma}_{|X|^p}(0) - \gamma_{|X|^p}(0))}{(\gamma_{|X|^p}(0))^2}(1 + o_p(1)), \qquad (2.7.16)\\ &= \sqrt{n}\left(\tfrac{1}{\gamma_{|X|^p}(0)} \quad , \quad \tfrac{\gamma_{|X|^p}(h)}{(\gamma_{|X|^p}(0))^2}\right)\begin{pmatrix}\hat{\gamma}_{|X|^p}(h) - \gamma_{|X|^p}(h)\\ \hat{\gamma}_{|X|^p}(0) - \gamma_{|X|^p}(0)\end{pmatrix} + o_p(1),\end{aligned}$$

where we used that

$$\hat{\gamma}_{|X|^p}(0) = \gamma_{|X|^p}(0) + o_p(1).$$

Another application of Theorem 2.1.19 yields,

$$\sqrt{n}\begin{pmatrix}\hat\gamma_{|X|^p}(h)-\gamma_{|X|^p}(h)\\ \hat\gamma_{|X|^p}(0)-\gamma_{|X|^p}(0)\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}|X_tX_{t+h}|^p - E|X_0X_h|^p - 2E|X|^p(|X_{t+h}|^p - E|X|^p)\\ |X_t|^{2p} - E|X|^{2p} - 2E|X|^p(|X_t|^p - E|X|^p)\end{pmatrix} + o_p(1)$$
$$\stackrel{d}{\to} N\left(\underline{0},\Sigma\right), \tag{2.7.17}$$

where

$$\Sigma = \begin{pmatrix}\nu^2 & \Lambda\\ \Lambda & \lambda^2\end{pmatrix}.$$

Then it follows from (2.7.16) and (2.7.17) that

$$\sqrt{n}(\hat\rho_{|X|^p}(h)-\rho_{|X|^p}(h)) \stackrel{d}{\to} N(0, a'\Sigma a).$$

where

$$a' = \begin{pmatrix}\frac{1}{\gamma_{|X|^p}(0)} & , & \frac{\gamma_{|X|^p}(h)}{(\gamma_{|X|^p}(0))^2}\end{pmatrix}.$$

Then

$$a'\Sigma a = \begin{pmatrix}\frac{1}{\gamma_{|X|^p}(0)} & , & \frac{\gamma_{|X|^p}(h)}{(\gamma_{|X|^p}(0))^2}\end{pmatrix}\begin{pmatrix}\nu^2 & \Lambda\\ \Lambda & \lambda^2\end{pmatrix}\begin{pmatrix}\frac{1}{\gamma_{|X|^p}(0)}\\ \frac{\gamma_{|X|^p}(h)}{(\gamma_{|X|^p}(0))^2}\end{pmatrix}.$$

□

Remark 2.7.9. The proofs of Lemmas 2.7.8 and 2.1.19 are helpful in constructing a multivariate central limit theorem for the sample ACF $(\widehat\rho_{|X|^p}(h))_{h=1,\cdots,m}$. Let

$$A = \frac{1}{(\gamma_{|X|^p}(0))^2}\begin{pmatrix}-\gamma_{|X|^p}(1) & \gamma_{|X|^p}(0) & 0 & 0 & \cdots & 0\\ -\gamma_{|X|^p}(2) & 0 & \gamma_{|X|^p}(0) & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -\gamma_{|X|^p}(m) & 0 & 0 & 0 & \cdots & \gamma_{|X|^p}(0)\end{pmatrix},$$

$$S = \begin{pmatrix}\sqrt{n}(\widehat\gamma_{|X|^p}(0)-\gamma_{|X|^p}(0))\\ \sqrt{n}(\widehat\gamma_{|X|^p}(1)-\gamma_{|X|^p}(1))\\ \sqrt{n}(\widehat\gamma_{|X|^p}(2)-\gamma_{|X|^p}(2))\\ \vdots\\ \sqrt{n}(\widehat\gamma_{|X|^p}(m)-\gamma_{|X|^p}(m))\end{pmatrix}.$$

Then the differences between the ACF's and their estimators can be written as

$$\begin{pmatrix}\sqrt{n}(\hat\rho_{|X|^p}(1)-\rho_{|X|^p}(1))\\ \sqrt{n}(\hat\rho_{|X|^p}(2)-\rho_{|X|^p}(2))\\ \vdots\\ \sqrt{n}(\hat\rho_{|X|^p}(m)-\rho_{|X|^p}(m))\end{pmatrix} = AS + o_p(1) \stackrel{d}{\to} N(0, A'\Sigma A),$$

by the multivariate central limit theorem given in Theorem 2.1.19. Here

$$\Sigma = \begin{pmatrix}\lambda^2 & \Lambda(0,1) & \cdots & \Lambda(0,m)\\ \Lambda(0,1) & \nu^2(1) & \cdots & \Lambda(1,m)\\ \vdots & \vdots & \vdots & \vdots\\ \Lambda(0,m) & \Lambda(1,m) & \cdots & \nu^2(m)\end{pmatrix}.$$

Chapter 3

Some extreme value theory for stochastic volatility models

3.1 Preliminaries on extreme value theory

A risk manger is responsible for a portfolio consisting of a few up to hundred or thousands of financial assets and contracts. Management or investors have also imposed risk preferences that the risk manager is trying to meet. To evaluate the position the risk manager tries to assess the loss distribution to make sure that the current positions is in accordance with imposed risk preferences. If it is not, then the risk manager must rebalance the portfolio until a desirable loss distribution is obtained. We may view a financial investor as a player participating in the game at financial market and the loss distribution must be evaluated in order to know which game the investor is participating in.

There are some standard methods for computing value at risk (VaR) and expected shortfall (ES) for a portfolio of risky assets like empirical VaR and ES, using bootstrap to obtain confidence intervals, historical simulation, Monte-Carlo methods, etc. One of these methods is extreme value theory. The main advantage of the methods of extreme value theory is that these methods are designed to be able to say as much as possible about the tail of the underling distribution.

In a risk management context one is typically interested in the loss distribution from an investment in financial assets or the distribution of the claims arriving to an insurance company. Extreme events are particularly frightening because although they are by definition rare, they may cause severe losses to a financial institution or insurance company. Extreme value theory is the proper tool for estimating the probability of such events.

3.1.1 Extreme value theory for iid random variables

Let $X, X_1, X_2, ...$ be a sequence of iid non-degenerate random variables (rv). We define the sample maxima M_n as

$$M_1 = X_1,\ M_n = \max(X_1, X_2, ...X_n), \quad n \geq 2.$$

A non-degenerate rv X is called max-stable if for all $n \geq 1$ there exist $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$M_n \stackrel{d}{=} c_n X + d_n. \tag{3.1.1}$$

We are interested in the limiting distribution of M_n as $n \to \infty$ under affine transformations.

If there exists a random variable Y with a non-degenerate distribution H such that for some $c_n > 0$ and $d_n \in \mathbb{R}$

$$c_n^{-1}(M_n - d_n) \stackrel{d}{\to} Y, \tag{3.1.2}$$

then by the Fisher-Tippett theorem [11, p. 121] Y belongs to the type of one of the distributions Fréchet, Weibull or Gumbel. The Fréchet distribution function is given by

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \qquad x > 0, \qquad \alpha > 0,$$

the Weibull distribution function by

$$\Psi_\alpha = \exp\left(-(-x)^\alpha\right), \qquad x \leq 0, \qquad \alpha > 0,$$

and the Gumbel distribution function by

$$\Lambda(x) = \exp\left(-e^{-x}\right), \qquad x \in \mathbb{R}.$$

Maximum domain of attraction (MDA)

Definition 3.1.1. *A random variable X (and its distribution F) belongs to the maximum domain of attraction of the extreme value distribution H if* (3.1.2) *holds. We write $X \in MDA(H)$ or $F \in MDA(H)$.*

The distribution function F belongs to the maximum domain of attraction of the extreme value distribution H if and only if

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\log H(x), \qquad x \in \mathbb{R}. \tag{3.1.3}$$

where $\overline{F}(x) = 1 - F(x)$, see [11, p. 128].

By the above definition and the relation (3.1.2) the maximum domain of attraction of the three distributions can be characterized as follows :

- A random variable X belongs to the maximum domain of attraction of Φ_α if and only if $\overline{F}(x) = x^{-\alpha}L(x),\ x > 0$, L is slowly varying function, i.e. $\lim_{x\to\infty}\frac{L(tx)}{L(x)} = 1$ for all $t > 0$. In this case, one can choose

$$c_n = \inf(x \in \mathbb{R} : F(x) \geq 1 - \frac{1}{n}) = F^{\leftarrow}(1 - \frac{1}{n}),$$

and $d_n = 0$. Examples of distributions in this class are the Pareto, Cauchy, Burr, loggamma distributions.

- A random variable X belongs to the maximum domain of attraction of Ψ_α if and only if the right end point

$$x_F = \sup\left(x \in \mathbb{R} : F(x) < 1\right),$$

is finite and $\overline{F}(x_F - \frac{1}{x}) = x^{-\alpha}L(x),\ x > 0$, where L is slowly varying function. The corresponding norming constants are $c_n = x_F - F^{\leftarrow}(1 - \frac{1}{n})$ and $d_n = x_F$. Examples of this class are beta and uniform distributions.

- A random variable X belongs to the maximum domain of attraction of Λ if and only if there exists $z < x_F$ such that

$$\overline{F}(x) = c(x)\exp\left(-\int_z^x \frac{g(t)}{a(t)}dt\right), \qquad z < x < x_F, \tag{3.1.4}$$

where $c(\cdot)$ and $g(\cdot)$ are measurable functions, $\lim_{x\to x_F} c(x) = c > 0$, $\lim_{x\to x_F} g(x) = 1$ and $a(x)$ is a positive, absolutely continuous function with $\lim_{x\to x_F} a'(x) = 0$. The norming constants are given by $d_n = F^{\leftarrow}(1 - \frac{1}{n})$ and $c_n = a(d_n)$. Examples of distributions in $MDA(\Lambda)$ are the normal, lognormal, exponential, Weibull and gamma distributions.

3.1.2 The extremal index

In contrast to Section 3.1.1, in reality extremal events often tend to occur in clusters caused by local dependence in the data. For instance, large claims in insurance are mainly due to hurricanes, storms, floods, earthquakes, etc. Claims are then linked with these events and do not occur independently. The same can be observed with financial data such as exchange rates and asset prices. If one large value in such a time series occurs we can usually observe a cluster of large values over short periods afterwards, see Section 2.2.

The extremal index is a quantity which, in an intuitive way, allows one to characterize the relationship between the dependence structure of the data and their extremal behavior.

Definition 3.1.2. (Extremal index [11, p. 416])

Let (X_n) *be a strictly stationary sequence with marginal distribution function* F *and* θ *a non-negative number. Assume that for every* $\tau > 0$ *there exists a sequence* (u_n) *such that*

$$\lim_{n\to\infty} n\overline{F}(u_n) = \tau, \tag{3.1.5}$$

$$\lim_{n\to\infty} P(M_n \le u_n) = e^{-\theta\tau}. \tag{3.1.6}$$

Then θ *is called the extremal index of the sequence* (X_n)*.*

The extremal index θ always belongs to the interval $[0,1]$, see Section 8.1 in [11].

3.1.3 Univariate regular variation

Definition 3.1.3. (Slowly varying function [28, p. 105])

A measurable positive function $L(x)$ *on* $(0,\infty)$ *is slowly varying if it satisfies the asymptotic relation*

$$\frac{L(cx)}{L(x)} \to 1, \quad x \to \infty, \tag{3.1.7}$$

for all $c > 0$*.*

The class of slowly varying functions includes constants, logarithms, iterated logarithm, powers of logarithms. Every slowly varying function has the representation

$$L(x) = c_0(x) \exp \int_{x_0}^{x} \frac{\varepsilon(t)}{t} dt, \quad 0 < x_0 < x, \tag{3.1.8}$$

where $\varepsilon(t) \to 0$ as $t \to \infty$ and $c_0(t)$ is a positive function satisfying $c_0(t) \to c_0$ for some positive constant c_0. Using the representation in Equation (3.1.8), one can show that for every $\delta > 0$,

$$\lim_{x\to\infty} \frac{L(x)}{x^\delta} = 0, \quad \lim_{x\to\infty} x^\delta L(x) = \infty,$$

i.e., L is negligible compared to any power function x^δ, $x^{-\delta}$.

Definition 3.1.4. (Regularly varying function and regularly varying random variable [28, p. 106])

Let L *be a slowly varying function in the sense of* (3.1.7)*.*

1. *For any* $\delta \in \mathbb{R}$*, the function* $f(x) = x^\delta L(x), x > 0$*, is said to be regularly varying with index* δ*.*

2. *A positive random variable* X *and its distribution are said to be regularly varying with (tail) index* $\alpha \ge 0$ *if the right tail of the distribution has the representation* $P(X > x) = L(x)x^{-\alpha}, x > 0$*.*

An alternative way of defining regular variation with index δ is to require

$$\lim_{x\to\infty} \frac{f(cx)}{f(x)} = c^{\delta}, \text{ for all } \quad c > 0.$$

Hence a distribution F of a positive random variable X is regularly varying with index $\alpha \geq 0$ if and only if

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}, \text{ for each } x > 0. \tag{3.1.9}$$

We call a random variable X and its distribution regularly varying if its distribution is regularly varying in a "balanced manner" in both tails, i.e. if for some $p \in [0, 1]$ the following two limits exist

$$\lim_{t\to\infty} \frac{P(X > tx)}{P(|X| > t)} = px^{-\alpha} \quad \text{and} \quad \lim_{t\to\infty} \frac{P(X \leq -tx)}{P(|X| > t)} = qx^{-\alpha}, \tag{3.1.10}$$

for every $x > 0$ and $q = 1-p$. The word "balance" is imprecise. Apart from distributions which have both tails regularly varying of a "similar shape", this definition includes distributions with one tail regularly varying with index α and the other tail of any other shape as long as this tail is lighter.

From the definition of the regularly varying distributions we conclude that they have very heavy tails, in particular for small α. Examples of regularly varying distributions include the Pareto and Burr distributions which are standard models for large claims in (re)insurance applications.

Example 3.1.5. Some univariate regularly varying distributions

Here we give some well known regularly varying distributions. We introduce the right tail $\overline{F}$ or density f for these distributions.

Pareto $\overline{F}(x) = \frac{x^{\alpha}}{(x+k)^{\alpha}}, \quad x \geq 0, \quad k > 0, \quad \alpha > 0,$

student with n degrees of freedom

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\Pi n}}(1 + x^2/n)^{(n+1)/2}, \quad x \in \mathbb{R},$$

log-gamma $f(x) = \frac{\gamma^{\beta}}{\Gamma(\beta)}(\log x)^{\beta-1}x^{-\alpha-1}, \quad x \geq 1, \quad \alpha, \beta > 0,$

Burr $\overline{F}(x) = (\frac{k}{k+x^{\tau}})^{\alpha}, \quad \alpha, k, \tau > 0,$

Cauchy $f(x) = (\pi(1 + x^2))^{-1}, \quad x \in \mathbb{R},$

Fréchet $\overline{F}(x) = e^{-x^{-\alpha}}, \quad x > 0, \ \alpha > 0,$

Generalized Pareto distribution with positive shape parameter ξ. The distribution function is given by

$$G_{\xi}(x) = 1 - (1 + \xi x)^{-1/\xi}, \quad \xi > 0, \ x \geq 0.$$

In various cases the analysis of the moments of a random variable X can be refined by a study of the asymptotic tail behavior of the distribution of X. The close relation between the moments and the tails can be seen e.g. from the fact that for any non-negative random variable X,

$$EX = \int_0^\infty P(X > x)dx. \tag{3.1.11}$$

Assume X has a power law tail of the form

$$P(X > x) = x^{-\alpha}L(x), \quad x > 0, \tag{3.1.12}$$

where $\alpha > 0$ and L is a slowly varying function. Since for every $\delta > 0$ there exist positive constants x_0 and c_1, c_2 such that

$$c_1 x^{-\delta} \le L(x) \le c_2 x^{-\delta}, \quad x \ge x_0, \tag{3.1.13}$$

the contribution of L to the tail in Equation (3.1.12) is negligible compared to the power law $x^{-\alpha}$. From Equations (3.1.11) – (3.1.13) we conclude that

$$E(X^{\alpha+\delta}) \begin{cases} < \infty, & \delta < 0, \\ = \infty, & \delta > 0, \end{cases}$$

whereas EX^α may be finite or infinite, depending on the slowly varying function L.

There exists empirical evidence that the distribution of log-returns is well approximated in its left and right tails by a regularly varying function (possibly with different tail indices on the left and on the right). Also teletraffic data (file lengths, transmission durations, throughput rates, etc.) and insurance claims are often found to have power law tails.

Regular variation conditions are important for limit theorems in probability theory. It is well known (Feller [16]) that for an iid sequence (X_n) with distribution F and $S_n = X_1 + \cdots + X_n$, $n \ge 1$, there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that

$$c_n^{-1}(S_n - d_n) \xrightarrow{d} Y, \tag{3.1.14}$$

for some non-degenerate Y if and only if

$$\sigma^2(x) = EX^2 I_{\{|X| \le x\}}, \quad x > 0,$$

is regularly varying. In particular if $EX^2 < \infty$, Y is normal, and if X is regularly varying with index $\alpha < 2$, then

$$x^{-2}\sigma^2(x) = x^{-2}\int_{-x}^{x} y^2 dF(y) \sim \frac{\alpha}{2-\alpha} P(|X| > x).$$

In the case $\alpha < 2$, regular variation of X is equivalent to the fact that Y in (3.1.14) has an infinite variance $\alpha-$stable distribution. Its distribution is given by the characteristic function in Equation (2.5.14). Stable distributions have the stability property. This means that for iid random variables Y_i with an $\alpha-$stable distribution there exists $b_n > 0$ such that

$$Y_1 + \cdots + Y_n \stackrel{d}{=} n^{-\frac{1}{\alpha}} Y_1 + b_n, \quad n \geq 1.$$

Examples of stable distributions are the normal distribution with $\alpha = 2$ and the Cauchy distribution with $\alpha = 1$. The latter distribution is symmetric. The Cauchy distribution coincides with the $t-$distribution with one degree of freedom.

Regular variation is also a necessary and sufficient condition for convergence of maxima of iid random variables X_i in the maximum domains of attraction of the Fréchet and Weibull distributions, see Section 3.1.1.

3.1.4 Multivariate regular variation

Definition 3.1.6. (Multivariate regular variation [1])
The $d-$dimensional random vector $\underline{X} = (X_1, \ldots, X_d)'$ and its distribution are said to be regularly varying with index $\alpha > 0$ if there exists a random vector Θ with values in $\mathbb{S}^{d-1}$, where $\mathbb{S}^{d-1} = \{\underline{x} \in \mathbb{R}^d : |\underline{x}| = 1\}$ denotes the unit sphere in $\mathbb{R}^d$ with respect to the norm $|.|$, such that for all $t > 0$,

$$\frac{P(|\underline{X}| > ty, \underline{X}/|\underline{X}| \in \cdot)}{P(|\underline{X}| > y)} \xrightarrow{w} t^{-\alpha} P(\Theta \in \cdot), \quad \text{as } y \to \infty. \tag{3.1.15}$$

The symbol $\xrightarrow{w}$ stands for weak convergence on the Borel $\sigma-$algebra of $\mathbb{S}^{d-1}$ and the probability measure $P(\Theta \in \cdot)$ on $\mathbb{S}^{d-1}$ is the spectral measure of regular variation of $\underline{X}$.

Observe that, in the case $d = 1$, Equation (3.1.15) coincides with condition (3.1.10), which was the definition of regular variation of X in $\mathbb{R}$. The spectral measure in this case is given by $P(\Theta = 1) = p$ and $P(\Theta = -1) = q$. Definition 3.1.6 is often used as a definition of multivariate regular variation. The following definition is equivalent to it.

Definition 3.1.7. (Equivalent definition for multivariate regular variation [34])
The $d-$dimensional random vector $\underline{X} = (X_1, \ldots, X_d)'$ and its distribution are said to be regularly varying with index $\alpha > 0$ if there exists a random vector $\Theta \in \mathbb{S}^{d-1}$ and a sequence (a_n), $a_n \to \infty$, such that for any $t > 0$,

$$nP(|\underline{X}| > ta_n, \underline{X}/|\underline{X}| \in \cdot) \xrightarrow{w} t^{-\alpha} P(\Theta \in \cdot), \qquad n \to \infty. \tag{3.1.16}$$

Here $\xrightarrow{w}$ denotes weak convergence.

As shown in [35] Section 5.4.2., Definitions 3.1.6 and 3.1.7 are equivalent to

$$\frac{P(x^{-1}\underline{X} \in \cdot)}{P(|\underline{X}| > x)} \xrightarrow{v} \mu(\cdot), \tag{3.1.17}$$

where $\xrightarrow{v}$ denote vague convergence in $\overline{\mathbb{R}}^d \setminus \{\underline{0}\}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ and μ is a Radon measure (finite on compact sets) on $\overline{\mathbb{R}}^d \setminus \{\underline{0}\}$ satisfying

$$\mu(tA) = \mu(A)t^{-\alpha}, \quad t > 0,$$

for any Borel set $A \subset \overline{\mathbb{R}}^d \setminus \{\underline{0}\}$, bounded away from zero, i.e. there exists ϵ such that $A \subset \{\underline{x} : |\underline{x}| > \epsilon\}$.

The upper tail coefficient

In the literature on risk management, one often finds the upper tail coefficient λ_u of a two-dimensional random vector $\underline{X} = (X_1, X_2)$ as a measure of extremal risk in $\underline{X}$, see [25]. For such a random vector with $X_1 \stackrel{d}{=} X_2$ and jointly regularly varying distribution we have

$$\lambda_u = \lim_{x\to\infty} P(X_1 > x | X_2 > x) = \lim_{x\to\infty} \frac{P(x^{-1}(X_1, X_2) \in (1,\infty)^2)}{P((X_1, X_2) \in \mathbb{R} \times (1,\infty))}.$$

Notice that $(1,\infty)^2$ and $\mathbb{R} \times (1,\infty)$ are bounded away from zero. Therefore,

$$\lambda_u = \lim_{x\to\infty} \frac{P(x^{-1}(X_1, X_2) \in (1,\infty)^2)/P(|\underline{X}| > x)}{P(x^{-1}(X_1, X_2) \in \mathbb{R} \times (1,\infty))/P(|\underline{X}| > x)} = \frac{\mu((1,\infty)^2)}{\mu(\mathbb{R} \times (1,\infty))}.$$

The definition of the upper tail index for a regularly varying vector $\underline{X} \in \mathbb{R}^d$ can be generalized in many different ways. For example, let $A \subset \overline{\mathbb{R}}^{d-1}$ be a set bounded away from zero. Then

$$\begin{aligned} P(x^{-1}(X_2, \dots, X_d) \in A | X_1 > x) &= \frac{P(x^{-1}\underline{X} \in (1,\infty) \times A)}{P(X_1 > x)} \\ &\to \frac{\mu((1,\infty) \times A)}{\mu((1,\infty) \times \mathbb{R}^{d-1})}, \end{aligned}$$

where we assumed that $\mu((1,\infty) \times \mathbb{R}^{d-1}) > 0$.

For example, let $A_1 = (1,\infty)^{d-1}$. Then

$$\begin{aligned} P(x^{-1}(X_2, \dots, X_d) \in A_1 | X_1 > x) &= P(X_i > x, i = 2, \dots, d | X_1 > x) \\ &\to \frac{\mu((1,\infty)^d)}{\mu((1,\infty) \times \mathbb{R}^{d-1})}, \end{aligned}$$

where we assumed that $\mu((1,\infty) \times \mathbb{R}^{d-1}) > 0$. Or let $A_2 = ([0,1]^{d-1})^c$. Then

$$\begin{aligned} P(x^{-1}(X_2, \dots, X_d) \in A_2 | X_1 > x) &= P(X_i > x \text{ for some } i = 2, \dots, d | X_1 > x) \\ &\to \frac{\mu((1,\infty) \times A_2)}{\mu((1,\infty) \times \mathbb{R}^{d-1})}, \end{aligned}$$

where we assumed $\mu((1,\infty) \times \mathbb{R}^{d-1}) > 0$.

Examples of spectral measures

We consider some examples of regularly varying random vectors and try to determine their spectral measures. For the purpose of illustration we focus on two-dimensional vectors with positive components in the first two examples.

Example 3.1.8. (Total dependence)
We assume that $\underline{X} = (X, X)'$ for some regularly varying X with index $\alpha > 0$ and choose the max-norm $|\underline{x}| = \max(x_1, x_2)$. In this case, with $nP(|\underline{X}| > x_n) \sim 1$,

$$nP(|\underline{X}| > x_n, \underline{X}/|\underline{X}| \in S) = nP(|\underline{X}| > x_n, (1,1) \in S) \sim I_S((1,1)),$$

where we assume that $(1,1)$ is not at the boundary of S. The spectral measure is degenerate and concentrated at the intersection of the unit sphere with the line $x = y$. The same remark applies to any norm in $\mathbb{R}^2$.

For an iid sequence $\underline{X_1}, \ldots, \underline{X_n}$ with the same distribution as $\underline{X}$, it is clear from the dependence structure of the components that an $\underline{X_i}$ far away from the origin occurs if both components in the vector are large at the same time.

In contrast to the previous artificial example, for a real-life time series $\underline{X_1}, \ldots, \underline{X_n}$ with dependent non-identical components we do not expect that all these vectors lie on the line $x = y$. However, when the $\underline{X_i}$s of large modulus stay away from the axes, we have an indication of asymptotic dependence. The observed features can be generalized to vectors with values in $\mathbb{R}^d$.

Example 3.1.9. (Independence between components)
We assume that $\underline{X} = (X_1, X_2)$ has independent components and $X_1 \stackrel{d}{=} X_2$ for a regularly varying X_1 with index $\alpha > 0$. Choose the max-norm $|X| = \max(|X_1|, |X_2|)$. Then

$$\begin{aligned} P(|\underline{X}| > x) &= P(\max(X_1, X_2) > x) = 1 - P(\max(X_1, X_2) \le x) \\ &= 1 - (P(X_1 \le x))^2 = (1 - P(X_1 \le x))(1 + P(X_1 \le x)) \sim 2P(X_1 > x). \end{aligned}$$

Choose c_n such that $P(|X| > c_n) \sim \frac{1}{n}$. Then for $y > 0$,

$$\begin{aligned} nP(c_n^{-1}\underline{X} \in (y, \infty)^2) &= nP(c_n^{-1} X_i > y; i = 1, 2) = n[P(X_1 > c_n y)]^2 \\ &\sim n[P(X_1 > c_n)]^2 y^{-2\alpha} \to 0. \end{aligned}$$

Since $y > 0$ is arbitrary, we can choose y so small that $(y, \infty)^2 \supset A$ for any set A which is bounded away

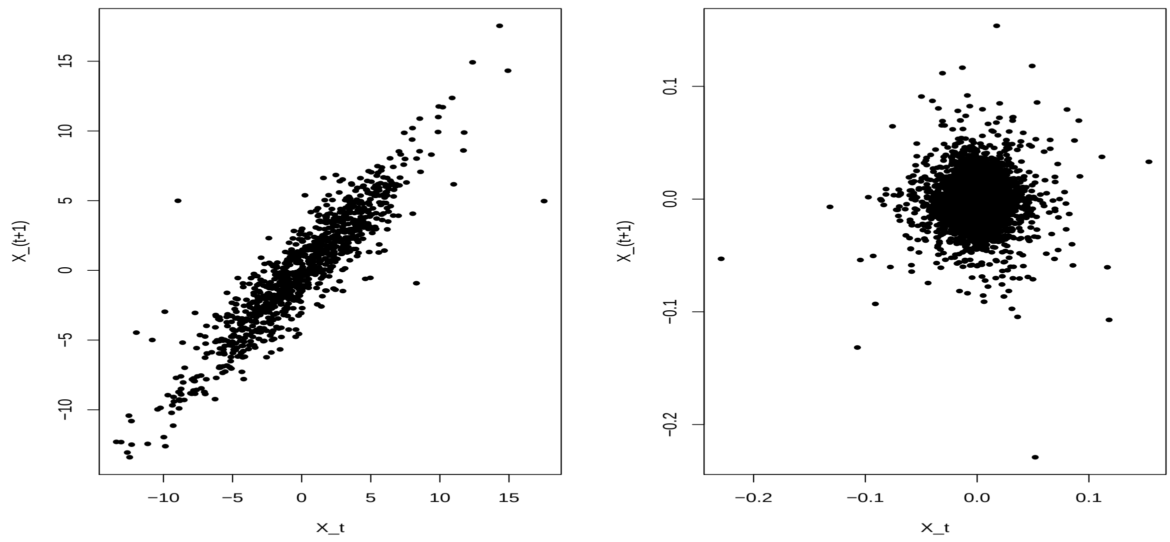

Figure 3.1: (Example 3.1.9) Left: Plot of 1000 lagged vectors $\underline{X}_t = (X_t, X_{t+1})$ for the AR(1) process $X_{t+1} = 0.9X_t + Z_t$ for iid symmetric regularly varying noise (Z_t) with tail index 1.8. The vectors $\underline{X}_t$ with large norm are typically concentrated along the line $y = 0.9x$. Right: Scatterplot of the pairs (X_t, X_{t+1}) of the daily log-return X_t of S&P500 series. The extremes in the series do not tend to cluster around the axes. This is an indication of dependence in the tails.

from the axes. Therefore $nP(c_n^{-1}\underline{X} \in A) \to 0$ for such sets. In particular, for sets

$$A(r, S) = \{\underline{x} \in \mathbb{R}_t^2 : |\underline{X}| > r, \frac{\underline{X}}{|\underline{X}|} \in S\},$$

where $S \subset \mathbb{S}$ and S is bounded away from $(0,1)$ and $(1,0)$,

$$nP(c_n^{-1}\underline{X} \in A(r, S)) = nP(|\underline{X}| > c_n r, \frac{\underline{X}}{|\underline{X}|} \in S) \to 0, \quad r > 0.$$

Therefore

$$\mu(A(r, S)) = r^{-\alpha}\mu(A(1, S)) = 0,$$

but $\mu(A(1, S))$ for any Borel sets $S \subset \mathbb{S}$ is the spectral measure of $\underline{X}$. Hence the spectral measure is concentrated at $(0,1)$ and $(1,0)$.

Example 3.1.10. (A toy model)

This example helps one to understand the spectral measure. Let

$$\underline{X} = R(\cos\Phi, \sin\Phi), \tag{3.1.18}$$

where the radius R has distribution $P(R > r) = r^{-\alpha}, r \geq 1$, for some $\alpha > 0$ and is independent of the random angle Φ with distribution on $(-\pi, \pi]$. Choosing the Euclidean norm $|\cdot|$ and exploiting the independence of R and Φ, this vector is immediately seen to be regularly varying with $\Theta = (\cos\Phi, \sin\Phi)$:

$$\frac{P(|\underline{X}| > tx, \underline{X}/|\underline{X}| \in S)}{P(|\underline{X}| > x)} = \frac{P(R > tx, \Theta \in S)}{P(R > x)} = \frac{P(R > tx)}{P(R > x)} P(\Theta \in S) = t^{-\alpha} P(\Theta \in S),$$

provided $\min(tx, x) \geq 1$. The knowledge of the distribution of Φ allows for some straightforward interpretation of the two-dimensional dependence in the tails.

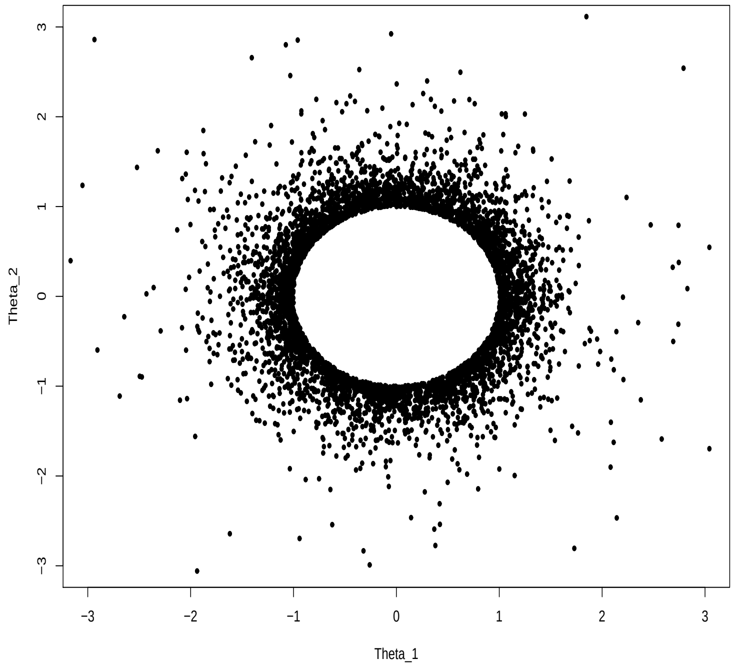

Figure 3.2: (Example 3.1.10) IID vectors $\underline{X_i}$ from model (3.1.18) with tail index $\alpha = 5$ and Φ is uniform on $(-\pi, \pi]$.

Functions of regularly varying vectors

Regularly variation remains valid under various useful operations. We consider some of them.

Products

Let X, Y be two independent random variables such that $X, Y > 0$ a.s. and X is regularly varying with index $\alpha > 0$ and $EY^{\alpha+\delta} < \infty$, for some $\delta > 0$. It follows from Breiman [7] that

$$P(XY > x) \sim E(Y^{\alpha}) P(X > x), \quad x \to \infty. \tag{3.1.19}$$

In particular, XY is regularly varying with index α. Thus the product inherits the heavier tail of the two factors. The condition $E(Y^{\alpha+\delta}) < \infty$ can be relaxed if more information about the tail $P(Y > x)$ is available. For example, if $P(Y > x) \sim cx^{-\alpha}$ for some $c > 0$, then (3.1.19) remains valid although $EY^\alpha = \infty$.

It is also well-known that XY is regularly varying with index $\alpha > 0$ if both X, Y are regularly varying with index $\alpha > 0$.

The following result is a multivariate extension of Breiman's result to random vectors. It can be found in [1].

Lemma 3.1.11. *Let $\underline{A}$ be a random $(r \times d)-$matrix such that $E\|\underline{A}\|^{\alpha+\delta} < \infty$ for some $\delta > 0$ ($\|.\|$ is an appropriate matrix norm). Assume that $\underline{A}$ is independent of the $d-$dimensional regularly varying vector $\underline{X}$ with index $\alpha > 0$ and limit measure μ in Equation (3.1.17). Then*

$$\frac{P(x^{-1}\underline{AX} \in \cdot)}{P(|\underline{X}| > x)} \xrightarrow{v} E\mu(\{\underline{x} \in \mathbb{R}^d : \underline{A}\ \underline{x} \in \cdot\}) = v(\cdot).$$

Here $\xrightarrow{v}$ denotes vague convergence. In particular, if v is a full measure in $\mathbb{R}^d$, (i.e.: not concentrated on a lower-dimensional subspace of $\mathbb{R}^d$) $\underline{A}\ \underline{X}$ is regularly varying with index $\alpha > 0$.

Summation

Let $\underline{X} = (X_1 \ldots, X_d)$ be a regularly varying vector with index $\alpha > 0$. Then $\sum_{i=1}^d c_i X_i$ is regularly varying with index $\alpha > 0$ for any choice of $c_i \in \mathbb{R},\ i = 1, \ldots, d$, such that $c_i \neq 0$ for at least one i.

Indeed, define $\underline{c} = (c_1, \ldots, c_d)'$ and

$$A_{\underline{c}} = \{\underline{x} \in \mathbb{R}^d : \underline{c}'\underline{x} > 1\},$$

then it follows from (3.1.17) that

$$\frac{P(\sum_{i=1}^d c_i X_i > tx)}{P(|\underline{X}| > x)} = \frac{P(x^{-1}\underline{X} \in A_{\underline{c}})}{P(|\underline{X}| > x)} \to \mu(tA_{\underline{c}}) = t^{-\alpha}\mu(A_{\underline{c}}) \quad \text{as } x \to \infty,$$

and

$$\frac{P(\sum_{i=1}^d c_i X_i < -tx)}{P(|\underline{X}| > x)} \to t^{-\alpha}\mu(\{\underline{x} : \underline{c}'\underline{x} < -1\}) = t^{-\alpha}\mu(A_{-\underline{c}}).$$

In particular,

$$\begin{aligned}
\frac{P(X_1 + \cdots + X_d > x)}{P(|\underline{X}| > x)} &\to \mu(\{\underline{x} : x_1 + \cdots + x_d > 1\}), \\
\frac{P(X_1 + \cdots + X_d > x)}{P(|\underline{X}| \leq -x)} &\to \mu(\{\underline{x} : x_1 + \cdots + x_d \leq -1\}).
\end{aligned}$$

As a special case, let X be a regularly varying variable and $P(|Y| > x) = o(P(|X| > x))$. Then, $X + Y$ is a regularly varying variable such that

$$\begin{aligned} P(X+Y > x) &\sim P(X > x), \quad x \to \infty, \\ P(X+Y \leq -x) &\sim P(X \leq -x), \quad x \to \infty. \end{aligned}$$

3.2 Regular variation of GARCH and stochastic volatility models

3.2.1 The GARCH model

Recall the GARCH(p, q) process (X_n) from Section 2.3.1. In this case it is well known that (σ_n) and (X_n) have regularly varying finite dimensional distributions, see [1]. This holds under general condition on the iid noise (Z_t). A sufficient condition for regularly variation of σ_t and X_t is that Z_t has a density with unbounded support. This includes $t-$ and normally distributed Z_t. The index α of regular variation of σ_t is given by

$$E(\alpha_1 Z_1^2 + \beta_1)^{\frac{\alpha}{2}} = 1, \tag{3.2.1}$$

in the GARCH$(1, 1)$ case, provided this solution exists. Then a result by Kesten [22] implies that

$$P(\sigma > x) \sim cx^{-\alpha}. \tag{3.2.2}$$

Breiman's result in Equation (3.1.19) shows that

$$P(X_t > x) = P(\sigma_t Z_t > x) \sim E(Z_+^\alpha) P(\sigma > x) \sim E(Z_+^\alpha) cx^{-\alpha}.$$

In the GARCH$(1, 1)$ case it is possible to determine the spectral distribution of the finite-dimensional distributions of (σ_t^2). To see this, we write

$$\sigma_t^2 = M_t, \quad A_t = \alpha_1 Z_{t-1}^2 + \beta_1, \quad B_0 = \alpha_0.$$

Then

$$\begin{aligned} \sigma_t^2 = M_t &= \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2 \\ &= B_0 + A_t M_{t-1}. \end{aligned}$$

Iteration yields

$$\begin{aligned} M_t &= A_t \cdots A_1 M_0 + \sum_{i=1}^{t} A_t \cdots A_{i+1} B_0 \\ &= A_t \cdots A_1 M_0 + B_0 + \sum_{i=1}^{t-1} A_t \cdots A_{i+1} B_0. \end{aligned}$$

Under the conditions of Kesten's [22] result and with $E|Z|^{\alpha+\delta} < \infty$ for some $\delta > 0$ (this condition is trivially satisfied for normally and $t-$distributed Z with $\nu > \alpha$ degrees of freedom), and for $h \geq 0$,

$$\begin{aligned}\underline{M}_{t,t+h} &= (M_t, \cdots, M_{t+h}) \\ &= (A_t \cdots A_1, \ldots, A_{t+h} \cdots A_1)M_0 + (\sum_{i=1}^{t} A_t \cdots A_{i+1}, \ldots, \sum_{i=1}^{t+h} A_{t+h} \cdots A_{i+1})B_0 \\ &= A_t \cdots A_1(1, A_{t+1}, \ldots, A_{t+h} \cdots A_{t+1})M_0 + R_t, \end{aligned} \tag{3.2.3}$$

where $E|R_t|^{\alpha/2+\delta} < \infty$.

The following lemma is helpful in this case.

Lemma 3.2.1. *(Lemma 3.12 in [20])*

Assume that $\underline{X}_1 \in \mathbb{R}^d$ is regularly varying with index $\alpha > 0$ and limit measure μ and $\underline{X}_2 \in \mathbb{R}^d$ is such that $P(|\underline{X}_2| > x) = o(P(|\underline{X}_1| > x))$ as $x \to \infty$. Then $\underline{X}_1 + \underline{X}_2$ is regularly varying with index $\alpha > 0$ and limit measure μ.

An application of Lemma 3.2.1 shows that regular variation of $\underline{M}_{t,t+h}$ is solely determined by the regular variation of $A_t \cdots A_1(1, A_{t+1}, \ldots, A_{t+h} \cdots A_{t+1})M_0$. Indeed, since $E|R_t|^{\alpha/2+\delta} < \infty$ for some $\delta > 0$,

$$nP(|n^{-\frac{2}{\alpha}} R_t| > \epsilon) \to 0, \quad \text{for all } \epsilon > 0. \tag{3.2.4}$$

Therefore the spectral distribution of $\underline{M}_{t,t+h} \stackrel{d}{=} \underline{M}_h$ is given by

$$\begin{aligned} & nP(|\underline{M}_h| > rn^{\frac{2}{\alpha}}, \frac{\underline{M}_h}{|\underline{M}_h|} \in S) \\ = \quad & nP(M_0 A_t \cdots A_1 |(1, A_{t+1}, \ldots, A_{t+h} \cdots A_{t+1})| > rn^{\frac{2}{\alpha}}, \frac{(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})}{|(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|} \in S) \\ = \quad & nP\left(M_0 A_t \cdots A_1 |(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})| I_{\{\frac{(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})}{|(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|} \in S\}} > rn^{\frac{2}{\alpha}}\right) \\ \sim \quad & nP(M_0 > rn^{\frac{2}{\alpha}}) E\left(A_t \cdots A_1 |(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})| I_{\{\frac{(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})}{|(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|} \in S\}}\right)^{\frac{\alpha}{2}} \\ & \sim r^{-\alpha} c E\left(A_t \cdots A_1 |(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})| I_{\{\frac{(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})}{|(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|} \in S\}}\right)^{\frac{\alpha}{2}}, \end{aligned}$$

where we used Breiman's result, see (3.1.19). By virtue of (3.2.2) we thus have

$$nP(\sigma_0^2 > rn^{\frac{2}{\alpha}}) \quad \sim \quad cr^{-\frac{\alpha}{2}}.$$

It is immediate that the spectral measure of $\underline{M}_0$ is then given by the probability measure on $\mathbb{S}^{h+1}$

$$F(S) = \frac{E\left[(A_t \cdots A_1)^{\frac{\alpha}{2}} |(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|^{\frac{\alpha}{2}} I_{\{\frac{(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})}{|(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|} \in S\}}\right]}{E\left[(A_t \cdots A_1)^{\frac{\alpha}{2}} |(1, A_{t+1}, \ldots, A_{t+1} \cdots A_{t+h})|^{\frac{\alpha}{2}}\right]},$$

for any Borel set $S \subset \mathbb{S}^h$. By virtue of (3.2.1) we have

$$E(A_t \cdots A_1)^{\frac{\alpha}{2}} = 1.$$

Hence

$$F(S) = \frac{E\left[(1, A_1, \ldots, A_1 \cdots A_h)^{\frac{\alpha}{2}} I_{\{\frac{(1,A_1,\ldots,A_1\cdots A_h)}{|(1,A_1,\ldots,A_1\cdots A_h)|} \in S\}}\right]}{E|(1, A_1, \ldots, A_1 \cdots A_h)|^{\frac{\alpha}{2}}},$$

is the spectral measure of the distribution of $\underline{M}_h$.

The spectral measure of the finite-dimensional distributions of (X_t^2) follows in a similar way. Using the recursion in (3.2.3), we see that

$$\begin{aligned}(X_t^2, \ldots, X_{t+h}^2) &= (Z_t^2\sigma_t^2, \ldots, Z_{t+h}^2\sigma_{t+h}^2) \\ &= (Z_t^2 A_t \cdots A_1 M_0, \ldots, Z_{t+h}^2 A_{t+h} \cdots A_1 M_0) + S_t \qquad (3.2.5) \\ &= \underline{N}_{t,t+h} + S_t,\end{aligned}$$

where $E|S_t|^{\frac{\alpha}{2}+\delta} < \infty$ for some $\delta > 0$ if $E|Z|^{\alpha+\delta} < \infty$.

Using the same argument as for the finite-dimensional distributions of (σ_t^2), we see that the regular variation of $(X_t^2, \ldots, X_{t+h}^2)$ is completely determined by the regular variation of M_0 in (3.2.5). Proceed as before for $\underline{N}_{t,t+h} \stackrel{d}{=} \underline{N}_h$:

$$\begin{aligned}
&nP(|\underline{N}_h| > rn^{\frac{2}{\alpha}}, \frac{\underline{N}_h}{|\underline{N}_h|} \in S) \\
&= nP\Big(M_0 A_t \cdots A_1 |(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})| > rn^{\frac{2}{\alpha}}, \\
&\qquad \frac{(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})}{|(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})|} \in S\Big) \\
&= nP\big(M_0 A_t \cdots A_1 |(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})| \\
&\qquad I_{\{\frac{(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})}{|(Z_t^2, Z_{t+1}^2 A_{t+1}, \ldots, Z_{t+h}^2 A_{t+1} \cdots A_{t+h})|} \in S\}} > rn^{\frac{2}{\alpha}}\Big) \\
&\sim nP(M_0 > rn^{\frac{2}{\alpha}}) E\left(|(Z_1^2, Z_2^2 A_2, \ldots, Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}} I_{\{\frac{(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})}{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_{h+1} \cdots A_{h+1})|} \in S\}}\right) \\
&\sim cr^{-\alpha} E\left(|(Z_1^2, Z_2^2 A_2, \ldots, Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}} I_{\{\frac{(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})}{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_{h+1} \cdots A_{h+1})|} \in S\}}\right).
\end{aligned}$$

Here we again used Breiman's result (see (3.1.19)) and Kesten's result (see [22]).

Now it is immediate that the spectral measure of $\underline{N}_h$ is given by

$$E\left(\frac{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}} I_{\{\frac{(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})}{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})|} \in S\}}}{E|(Z_1^2, Z_2^2 A_2, \ldots, Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}}}\right),$$

for any Borel set $S \subset \mathbb{S}^h$.

For the ARCH(1) case, i.e., $\beta_1 = 0$, the spectral measure therefore simplifies to

$$\frac{E\left(|(\alpha_1 Z_1^2, \alpha_2 Z_2^2 A_2, \ldots, \alpha_1 Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}} I_{\{\frac{(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})}{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})|} \in S\}}\right)}{E|(\alpha_1 Z_1^2, \alpha_2 Z_2^2 A_2, \ldots, \alpha_1 Z_{1+h}^2 A_2 \cdots A_{1+h})|^{\frac{\alpha}{2}}}$$
$$= \frac{E\left(|(A_2, A_3 A_2, \ldots, A_{2+h} \cdots A_2)|^{\frac{\alpha}{2}} I_{\{\frac{(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})}{|(Z_1^2, Z_2^2 A_2, \ldots, Z_{h+1}^2 A_2 \cdots A_{h+1})|} \in S\}}\right)}{E|(A_2, A_3 A_2, \ldots, A_{2+h} \cdots A_2)|^{\frac{\alpha}{2}}}$$
$$= \frac{E\left(|(1, A_1, A_1 A_2, \ldots, A_1 \cdots A_h)|^{\frac{\alpha}{2}} I_{\{\frac{(1, A_1, A_1 A_2, \ldots, A_1 \cdots A_h)}{|(1, A_1, A_1 A_2, \ldots, A_1 \cdots A_h)|} \in S\}}\right)}{E|(1, A_1, A_1 A_2, \ldots, A_1 \cdots A_h)|^{\frac{\alpha}{2}}},$$

where we used that $EA_1^{\alpha/2} = 1$, see (3.2.1).

This spectral measure is not tractable and can only be calculated by using numerical or simulation methods.

The spectral measures of the finite-dimensional distribution of $(|X_t|^{2p})$ for any $p > 2$ can be calculated from the spectral measure of X_t^2. Indeed, for any set $B \subset \overline{\mathbb{R}}^{h+1} \setminus \{\underline{0}\}$, bounded away from zero whose boundary has limiting measure zero,

$$nP(n^{-\frac{2p}{\alpha}}(|X_0|^{2p}, \ldots, |X_h|^{2p}) \in B) = nP(n^{-\frac{2}{\alpha}}(X_0^2, \ldots, X_h^2) \in B^{\frac{1}{p}}),$$

where $B^{\frac{1}{p}} = \{(x_0^{\frac{1}{p}}, \ldots, x_h^{\frac{1}{p}}) : (x_0, \ldots, x_h) \in \mathbb{R}_+^{h+1}, (x_0, \ldots, x_h) \in B\}$. In particular, if $B = \{\underline{x} \in \overline{\mathbb{R}}^{h+1} : |\underline{x}| > 1, \frac{\underline{x}}{|\underline{x}|} \in S\}$ for some $S \in \mathbb{S}^h$, we obtain

$$nP(n^{-\frac{2p}{\alpha}}(|X_0|^{2p}, \ldots, |X_h|^{2p}) \in B) \to P(\underline{\theta}_p \in S).$$

In the following lemma we study some consequences of the regular variation of the finite-dimensional distributions of (X_t).

Lemma 3.2.2. *1. For* GARCH$(1,1)$

$$P(X_2^2 > x, \ldots, X_n^2 > x \mid X_1^2 > x) \to \frac{E\left(\min\left(Z_1^2, Z_2^2 A_2, \ldots, Z_n^2 \prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}\right)}{E|Z|^\alpha}. \tag{3.2.6}$$

For ARCH(1)

$$P(X_2^2 > x, \ldots, X_n^2 > x \mid X_1^2 > x) \to E\left(\min\left(1, A_1, \ldots, \prod_{j=1}^{n-1} A_j\right)^{\frac{\alpha}{2}}\right). \tag{3.2.7}$$

2. *For* GARCH(1,1)

$$P(X_i^2 \le x \ \ \textit{for all } 2 \le i \le n \mid X_1^2 > x) \to 1 - \frac{E\left(Z_1^2 \wedge \max(Z_2^2 A_2, \ldots, Z_n^2 \prod_{j=2}^n A_j)\right)^{\frac{\alpha}{2}}}{E|Z|^\alpha}\,. \tag{3.2.8}$$

For ARCH(1) the right-hand side in (3.2.8) *is*

$$\frac{\alpha}{2}\int_1^\infty P\left(\max(A_1,\ldots,\prod_{j=1}^{n-1} A_j) \le v^{-1}\right) v^{-1-\frac{\alpha}{2}}\,dv\,.$$

The extremal index of the sequence (X_t^2) *is the limit as* $n \to \infty$*:*

$$\theta = \frac{\alpha}{2}\int_1^\infty P(\sup_{n\ge 1}\prod_{j=1}^n A_j \le \nu^{-1})\nu^{-1-\frac{\alpha}{2}}\,d\nu.$$

3. *For* $n \ge 2$ *and* $0 < a < b < \infty$, *for* GARCH(1,1)

$$\begin{aligned} &P(x^{-1}X_n^2 \in (a,b] \mid X_1^2 > x) \\ \to\quad &(E|Z|^\alpha)^{-1}\Big(E\left(\min\left(Z_1^2\,, a^{-1}Z_n^2\prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}\right) - E\left(\min\left(Z_1^2\,, b^{-1}Z_n^2\prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}\right)\Big). \end{aligned} \tag{3.2.9}$$

For ARCH(1)

$$\begin{aligned} &P(x^{-1}X_n^2 \in (a,b] \mid X_1^2 > x) \\ \to\quad &E\left(\min\left(1\,, a^{-1}\prod_{j=1}^{n-1} A_j\right)^{\frac{\alpha}{2}}\right) - E\left(\min\left(1\,, b^{-1}\prod_{j=1}^{n-1} A_j\right)^{\frac{\alpha}{2}}\right). \end{aligned} \tag{3.2.10}$$

4. *For* GARCH(1,1)

$$P(X_2 > x, \ldots, X_n > x \mid X_1 > x) \quad \to \quad \frac{E\left(\min((Z_1)_+^2, (Z_2)_+^2 A_1\,, \ldots, (Z_n)_+^2 \prod_{j=1}^{n-1} A_j)\right)^{\frac{\alpha}{2}}}{EZ_+^\alpha}\,. \tag{3.2.11}$$

5. *For* GARCH(1,1)

$$\begin{aligned} &P(X_2 \le x, \ldots, X_n \le x \mid X_1 > x) \\ \to\quad &1 - \frac{E\left(\max(0, Z_2 A_1^{1/2}, \ldots, Z_n \prod_{j=1}^{n-1} A_j^{1/2}) \wedge (Z_1)_+\right)^\alpha}{EZ_+^\alpha} + P^{n-1}(Z_1 < 0). \end{aligned}$$

The extremal index of (X_t) *is the limit as* $n \to \infty$ *of the right-hand side.*

For ARCH(1) with symmetric Z *the right-hand side becomes with* $r_i = \mathrm{sign}(Z_i)$

$$\alpha\int_1^\infty P\left(\max(0, r_2 A_1, \ldots, r_n \prod_{j=1}^{n-1} A_j) \le v^{-1}\right) v^{-1-\alpha}\,dv + 2^{-n+1}\,.$$

The extremal index of (X_t) is given by

$$\theta = \alpha \int_1^\infty P\left(\sup(0, r_2 A_1, \ldots, r_n \prod_{j=1}^{n-1} A_j, \ldots) \le \nu^{-1}\right) \nu^{-1-\alpha} d\nu.$$

6. *For* GARCH$(1,1)$ *and* $\epsilon > 0$

$$P(|X_n^2 - X_1^2| \le \epsilon x \mid X_1^2 > x) \quad \to \quad 1 - \frac{E\left(\epsilon^{-1}|Z_n^2 \prod_{j=2}^n A_j - Z_1^2| \wedge Z_1^2\right)^{\frac{\alpha}{2}}}{E|Z|^\alpha}$$

For ARCH(1),

$$P(|X_n^2 - X_1^2| \le \epsilon x \mid X_1^2 > x) \quad \sim \quad 1 - E\left(\epsilon^{-1}|\prod_{j=2}^n A_j - 1| \wedge 1\right)^{\frac{\alpha}{2}}.$$

7. *For* GARCH$(1,1)$ *and* $c > 1$,

$$P(X_1^2 + \cdots + X_n^2 > c\, x \mid X_1^2 > x) \sim \frac{E(c^{-1}(Z_1^2 + Z_2^2 A_1 + \cdots + Z_n^2 \prod_{j=1}^{n-1} A_j) \wedge Z_1^2)^{\frac{\alpha}{2}}}{E|Z|^\alpha}.$$

For ARCH(1) and $c > 1$,

$$P(X_1^2 + \cdots + X_n^2 > c\, x \mid X_1^2 > x) \sim E(c^{-1}(1 + Z_1^2 + \cdots + \prod_{j=1}^{n-1} A_j) \wedge 1)^{\frac{\alpha}{2}}.$$

8. *For* GARCH$(1,1)$

$$\begin{aligned} P(\min(X_1^2, \ldots, X_n^2) > x \mid \max(X_1^2, \ldots, X_n^2)) \quad &\sim \quad \frac{P(\sigma_1^2 \min(Z_1^2, Z_2^2 A_1, \ldots, Z_n^2 \prod_{j=1}^{n-1} A_j) > x)}{P(\sigma_1^2 \max(Z_1^2, Z_2^2 A_1, \ldots, Z_n^2 \prod_{j=1}^{n-1} A_j) > x)} \\ &\sim \quad \frac{E(\min(Z_1^2, Z_2^2 A_1, \ldots, Z_n^2 \prod_{j=1}^{n-1} A_j))^{\frac{\alpha}{2}}}{E(\max(Z_1^2, Z_2^2 A_1, \ldots, Z_n^2 \prod_{j=1}^{n-1} A_j))^{\frac{\alpha}{2}}}. \end{aligned}$$

For ARCH(1)

$$P(\min(X_1^2, \ldots, X_n^2) > x \mid \max(X_1^2, \ldots, X_n^2) > x) \quad \sim \quad \frac{E(\min(1, A_1, \ldots, \prod_{j=1}^{n-1} A_j))^{\frac{\alpha}{2}}}{E(\max(1, A_1, \ldots, \prod_{j=1}^{n-1} A_j))^{\frac{\alpha}{2}}}.$$

Proof. (1) By Breiman's result

$$P(X_1^2 > x) \sim E|Z|^\alpha\, P(\sigma^2 > x)\,. \tag{3.2.12}$$

By Lemma 3.2.1 and Breiman, see also (3.2.3) and (3.2.4),

$$\begin{aligned} P(X_1^2 > x, \ldots, X_n^2 > x) \quad &\sim \quad P\left(\sigma_1^2\, Z_1^2 > x, \sigma_1^2 Z_2^2 A_2 > x, \ldots, \sigma_1^2 Z_n^2 \prod_{j=2}^n A_j > x\right) \\ &\sim \quad P(\sigma^2 > x)\, E\left(\min\left(Z_1^2, Z_2^2\, A_2, \ldots, Z_n^2 \prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}\right). \end{aligned} \tag{3.2.13}$$

For the ARCH(1) process $EA_1^{\frac{\alpha}{2}} = E(\alpha_1 Z^2)^{\frac{\alpha}{2}} = 1$ and therefore the right-hand side reads as

$$\begin{aligned} & P(\sigma^2 > x)\,\alpha_1^{-\alpha/2} E\left(\min\left(A_2, A_3A_2, \ldots, \prod_{j=2}^{n} A_j\right)^{\frac{\alpha}{2}}\right) \\ = \quad & P(\sigma^2 > x)\,\alpha_1^{-\alpha/2} E\left(\min\left(1, A_1, \ldots, \prod_{j=1}^{n-1} A_j\right)^{\frac{\alpha}{2}}\right). \end{aligned}$$

Taking the ratio of (3.2.13) and (3.2.12) and letting $x \to \infty$, we obtain (3.2.6) and (3.2.7).

(2) Let $\mu(x,\infty) = x^{-\frac{\alpha}{2}}$, $x > 0$. By the joint regular variation of $(X_1^2, \ldots, X_n^2)$ and Lemma 3.2.1,

$$\begin{aligned} & P(X_i^2 \le x \ \text{ for } 2 \le i \le n \mid X_1^2 > x) \\ \sim \quad & \frac{E\mu\{s : s\,Z_1^2 > 1\,, Z_2^2 A_1 s \le 1, \ldots, Z_n^2 \prod_{j=2}^n A_j s \le 1\}}{E|Z|^\alpha} \\ = \quad & \frac{E\mu\{s : s\,Z_1^2 > 1\,, s\,\max(Z_2^2 A_2, \ldots, Z_n^2 \prod_{j=2}^n A_j) \le 1\}}{E|Z|^\alpha} \\ = \quad & 1 - \frac{E\left(Z_1^2 \wedge \max(Z_2^2 A_2, \ldots, Z_n^2 \prod_{j=2}^n A_j)\right)^{\frac{\alpha}{2}}}{E|Z|^\alpha}\,. \end{aligned}$$

For ARCH(1), observing that $E(\alpha_1 Z_1)^{\frac{\alpha}{2}} = 1$, the right-hand side becomes

$$\begin{aligned} & 1 - E\left(A_2 \wedge A_2 \max(A_3, \ldots, \prod_{j=3}^{n+1} A_j)\right)^{\frac{\alpha}{2}} \\ = \quad & 1 - E\left(1 \wedge \max(A_1, \ldots, \prod_{j=1}^{n-1} A_j)\right)^{\frac{\alpha}{2}} \\ = \quad & \int_0^1 P(y^{2/\alpha} > \max(A_1, \ldots, \prod_{j=1}^{n-1} A_j))\,dy \\ = \quad & \frac{\alpha}{2} \int_1^\infty P\left(\max(A_1, \ldots, \prod_{j=1}^{n-1} A_j) \le v^{-1}\right) v^{-1-\frac{\alpha}{2}}\,dv\,. \qquad (3.2.14) \end{aligned}$$

It follows from [11, p. 422], Section 8.1, that the extremal index θ of the sequence (X_t^2) is given as the limit as $n \to \infty$ of (3.2.14).

(3) For $c > 0$ and $n \geq 2$, by Breiman's result,

$$\begin{aligned} P(X_n^2 > xc, X_1^2 > x) &\sim P\left(\sigma_1^2 c^{-1} Z_n^2 \prod_{j=2}^n A_j > x \,, Z_1^2\sigma_1^2 > x\right) \\ &\sim P(\sigma_1^2 > x)\, E\left(\min\left(Z_1^2 \,, c^{-1} Z_n^2 \prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}\right). \end{aligned}$$

For ARCH(1),

$$P(X_n^2 > xc, X_1^2 > x) \sim P(\sigma_1^2 > x)\, \alpha_1^{-\frac{\alpha}{2}} E\left(\min\left(1 \,, c^{-1} \prod_{j=1}^{n-1} A_j\right)^{\frac{\alpha}{2}}\right).$$

Then (3.2.9) and (3.2.10) follow.

(4) By Breiman's result

$$P(X_1 > x) \sim EZ_+^\alpha P(\sigma > x), \tag{3.2.15}$$

and by Breiman's result and Lemma 3.2.1,

$$\begin{aligned} P(X_1 > x, \ldots, X_n > x) &\sim P(Z_1\sigma_1 > x, Z_2 A_2^{1/2}\sigma_1 \,, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}\sigma_1 > x) \\ &\sim P(\sigma > x)\, E\left(\min((Z_1)_+^2, (Z_2)_+^2 A_1 \,, \ldots, (Z_n)_+^2 \prod_{j=2}^n A_j\right)^{\frac{\alpha}{2}}. \end{aligned} \tag{3.2.16}$$

Combining (3.2.15) and (3.2.16), we get (3.2.11).

(5) With $\nu(x,\infty) = x^{-\alpha}$.

$$\begin{aligned} & P(X_2 \leq x, \ldots, X_n \leq x \mid X_1 > x) \\ \to\ & \frac{E\left(\nu\{s : s\max(Z_2 A_2^{1/2}, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}) \leq 1, sZ_1 > 1\}\right)}{EZ_+^\alpha} \\ =\ & 1 - \frac{E\left(\max(0, Z_2 A_2^{1/2}, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}) \wedge (Z_1)_+\right)^\alpha}{EZ_+^\alpha} \\ & + \frac{E[(Z_1)_+^\alpha I_{(-\infty,0)}(\max(Z_2 A_2^{1/2}, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}))]}{EZ_+^\alpha} \end{aligned} \tag{3.2.17}$$

$$\begin{aligned} =\ & 1 - \frac{E\left(\max(0, Z_2 A_2^{1/2}, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}) \wedge (Z_1)_+\right)^\alpha}{EZ_+^\alpha} + \frac{E(Z_1)_+^\alpha I_{\{Z_2<0,\ldots,Z_n<0\}}}{EZ_+^\alpha} \\ =\ & 1 - \frac{E\left(\max(0, Z_2 A_2^{1/2}, \ldots, Z_n \prod_{j=2}^n A_j^{1/2}) \wedge (Z_1)_+\right)^\alpha}{EZ_+^\alpha} + P^{n-1}(Z_1 < 0). \end{aligned} \tag{3.2.18}$$

For ARCH(1) with symmetric Z, with $r_i = \text{sign}(Z_i)$, observing that $(|Z_t|)$ and (r_t) are independent,

$$
\begin{aligned}
& 1 - \frac{E\left(\max(0, Z_2 A_2^{1/2}, \dots, Z_n \prod_{j=2}^n A_j^{1/2}) \wedge (Z_1)_+\right)^{2\alpha}}{E Z_+^\alpha} \\
= & \; 1 - \frac{0.5E\left(\max(0, r_2 A_2 A_3, , \dots, r_n \prod_{j=2}^{n+1} A_j) \wedge A_2\right)^\alpha}{0.5E(\alpha_1 Z^2)^\alpha} \\
= & \; 1 - E\left(\max(0, r_2 A_2, , \dots, r_n \prod_{j=2}^{n} A_j) \wedge 1\right)^\alpha \\
= & \int_0^1 P\left(y^{1/\alpha} > \max(0, r_2 A_1, , \dots, r_n \prod_{j=1}^{n-1} A_j)\right) dy
\end{aligned}
$$

and

$$P^{n-1}(Z_1 < 0) = 2^{-n+1}$$

In view of [11, p. 422], Section 8.1, the limit of (3.2.18) as $n \to \infty$ yields the extremal index of (X_t).

(6) The set $\{(x, y) : |x-y| \le \epsilon, x > 1\}$ is bounded away from zero. Therefore with $\mu(x, \infty) = x^{-\frac{\alpha}{2}}, \;\; x > 0$,

$$
\begin{aligned}
P(|X_n^2 - X_1^2| \le \epsilon x \mid X_1^2 > x) \;\; & \sim \;\; P(\sigma_1^2 |Z_n^2 \prod_{j=1}^{n-1} A_j - Z_1^2| \le \epsilon x \mid \sigma_1^2 Z_1^2 > x) \\
& \to \;\; \frac{E\mu(\{s : s\,|Z_n^2 \prod_{j=1}^{n-1} A_j - Z_1^2| \le \epsilon, s Z_1^2 > 1\})}{E|Z|^\alpha} \\
& = \;\; 1 - \frac{E\left(\epsilon^{-1}|Z_n^2 \prod_{j=1}^{n-1} A_j - Z_1^2| \wedge Z_1^2\right)^\alpha}{E|Z|^{\frac{\alpha}{2}}}
\end{aligned}
$$

For ARCH(1) the right-hand side becomes

$$1 - E\left(\epsilon^{-1}|\prod_{j=2}^{n} A_j - 1| \wedge 1\right)^{\frac{\alpha}{2}}.$$

The relations (7) and (8) are obtained by similar arguments. □

3.2.2 The stochastic volatility model

We start by studying the one-dimensional tails. Let the sequence $(X_t) = (Z_t \sigma_t)$ be a stochastic volatility process given by Definition 2.3.2. Since σ_t and Z_t are independent it follows that X_t is regularly varying with index $\alpha > 0$, if Z_t is regularly varying with index α and $E(\sigma_t^{\alpha+\delta}) < \infty$ or if σ_t is regularly varying with index $\alpha > 0$ and $E(|Z|^{\alpha+\delta}) < \infty$ for some $\delta > 0$. This follows from Breiman's result in Equation (3.1.19).

Example 3.2.3. (Breiman's result for stochastic volatility model)

If $\sigma_t = e^{Y_t}$ for a Gaussian linear process (Y_t) then σ_t is log-normal, hence $E\sigma_t^\alpha < \infty$ for any $\alpha > 0$. Hence regular variation of X_t is due to the regular variation of Z_t. Moreover, if Z_t is regularly varying with index $\alpha > 0$, so is $Z_t Z_{t+h}$, see Remark after Relation (3.1.19). Then by Breiman's result

$$\begin{aligned} P(X_t X_{t+h} > x) &\sim E(\sigma_0\sigma_h)^\alpha P(Z_0 Z_h > x), \\ P(X_t X_{t+h} \leq -x) &\sim E(\sigma_0\sigma_h)^\alpha P(Z_0 Z_h < -x). \end{aligned}$$

Next we show regular variation of the finite-dimensional distributions of (X_t). We assume that $E\sigma_0^{\alpha+\delta} < \infty$ for some $\delta > 0$ and that (Z_t) is iid and regularly varying with index $\alpha > 0$. This follows by an application of a multivariate version of Breiman's result, see Lemma 3.1.11. Write

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{pmatrix} \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \underline{A}.\underline{Z}.$$

Since the Z_t's are iid, $\underline{Z}$ is regularly varying with index α and spectral measure concentrated at the intersection of the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ and the axes (see Example 3.1.9). By assumption, $E\|A\|^{\alpha+\delta} < \infty$, hence by the multivariate Breiman result

$$\begin{aligned} \frac{P(x^{-1}\underline{X} \in \cdot)}{P(|\underline{Z}| > x)} &\stackrel{v}{\rightarrow} E\mu(\underline{x} \in \mathbb{R}^d : \underline{A}.\underline{x} \in \cdot) \\ &= E\mu(\underline{x} \in \mathbb{R}^d : \underline{x} \in \underline{A}^{-1}\cdot) \end{aligned}$$

Since the spectral measure is concentrated at the axes, it follows that $\mu([\underline{x}, \underline{y}]) = 0$ for any $\underline{x} \leq \underline{y}$ such that $\underline{0} \notin [\underline{x}, \underline{y}]$.

We conclude that $\mu(A) = 0$ for any set $A \subset \overline{\mathbb{R}}^d \setminus \{\underline{0}\}$ which does not intersect the axes. Therefore the spectral measure of $\underline{X}$ is concentrated at the intersection of the unit sphere $\mathbb{S}^{d-1}$ and the axes. In other words, the vector $\underline{X}$ has very much the same extremal behavior (this means for large $|\underline{X}|$) as the iid vector $\underline{Z}$. See Example 3.1.9.

3.3 The extremal behavior of a stochastic volatility model with regularly varying noise

We know that there exist only three types of different limit laws for affinely transformed maxima of iid random variables: the Fréchet distribution Φ_α, the Weibull distribution Ψ_α and the Gumbel distribution Λ

for affinely transformed maxima of iid random variables. It is impossible to build a general extreme value theory for the class of all strictly stationary sequences. For a (strictly) stationary sequence (X_n), there are two conditions (see below) which ensure that its sample maxima (M_n) and the corresponding maxima $(\widetilde{M}_n)$ of an iid sequence $(\widetilde{X}_n)$ with common distribution function $F(x) = P(\widetilde{X}_1 \leq x) = P(X_1 \leq x)$ exhibit similar limit behavior. We call $(\widetilde{X}_n)$ an iid sequence associated with (X_n) or simply an associated iid sequence. We write $F \in$MDA(H) for any of the extreme value distributions H if there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that $c_n^{-1}(\widetilde{M}_n - d_n) \xrightarrow{d} Y$, for a random variable Y with distribution H. For the derivation of the limit probability of $P(\widetilde{M}_n \leq u_n)$ for a sequence of thresholds (u_n) the following factorization property is used:

$$P(\widetilde{M}_n \leq u_n) = P^n(\widetilde{X} \leq u_n) \approx e^{-n\overline{F}(u_n)}. \tag{3.3.1}$$

In general, for any $\tau \in [0, \infty]$,

$$P(\widetilde{M}_n \leq u_n) \rightarrow e^{-\tau},$$

if and only if

$$n\overline{F}(u_n) \rightarrow \tau = 1.$$

In what follows, we assume this condition for some $\tau = 1$.

It is clear that we cannot directly apply (3.3.1) to maxima of a stochastic volatility sequence. However, to overcome this problem we assume that there is a specific type of asymptotic independence. Recall the conditions D and D' from [11], Section 4.4.

Definition 3.3.1. Condition $D(u_n)$

For any integers $p, q \geq 1$ and $n \geq 1$

$$1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_q \leq n$$

such that $j_1 - i_p \geq l$ we have

$$|P(\max_{i \in A_1 \cup A_2} X_i \leq u_n) - P(\max_{i \in A_1} X_i \leq u_n) P(\max_{i \in A_2} X_i \leq u_n)| \leq \alpha_{n,l}, \tag{3.3.2}$$

where $A_1 = i_1, \ldots, i_p$, $A_2 = j_1, \ldots, j_q$ and $\alpha_{n,l} \rightarrow 0$ as $n \rightarrow \infty$ for some sequence $l = l_n = o(n)$.

Condition $D(u_n)$ is a distributional mixing condition, weaker than most of the classical forms of dependence restrictions. Condition $D(u_n)$ implies, for example, that

$$P(M_n \leq u_n) = (P(M_{[n/k]} \leq u_n))^k + o(1),$$

for constant or slowly increasing k. This relation already indicates that the limit behavior of (M_n) and its associated sequence $(\widetilde{M_n})$ must be closely related.

If (X_n) is strongly mixing with rate function (α_h) (see 2.1.10), then it is immediate that (3.3.2) holds. Indeed, for a sequence (X_n),

$$\begin{aligned} &|P(\max_{i\in A_1\cup A_2} X_i \le u_n) - P(\max_{i\in A_1} X_i \le u_n)P(\max_{i\in A_2} X_i \le u_n)| \\ &= \ |P(X_{i_1-i_p} \le u_n, \dots, X_0 \le u_n, X_{j_1-i_p} \le u_n, \dots, X_{j_q-i_p} \le u_n) - P(\max_{i\in A_1} X_i \le u_n)P(\max_{i\in A_2} X_i \le u_n)| \\ &\le \ \alpha_{j_1-i_p}, \end{aligned}$$

since

$$\{X_{i_1-i_p} \le u_n, \dots, X_0 \le u_n\} \in \sigma(\dots, X_{-1}, X_0),$$

and

$$\{X_{j_1-i_p} \le u_n, \dots, X_{j_q-i_p} \le u_n\} \in \sigma(X_{j_1-i_p}, X_{j_1-i_p+1}, \dots).$$

Definition 3.3.2. Condition $D'(u_n)$

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = 0.$$

The condition $D'(u_n)$ is an "anti-clustering condition" on the strictly stationary sequence (X_n). Indeed, notice that $D'(u_n)$ implies

$$\lim_{k\to\infty} \limsup_{n\to\infty} E(\sum_{1\le i<j\le [n/k]} I_{\{X_i>u_n, X_j>u_n\}}) \le \lim_{k\to\infty} \limsup_{n\to\infty} [n/k] \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) = 0,$$

so that, on average, joint exceedances of u_n by pairs (X_i, X_j) become very unlikely for large n.

We assume that $X_t = \sigma_t Z_t$ constitutes a stochastic volatility process with volatility sequence (σ_t), (Z_t) is iid independent of (σ_t). We want to show that D and D' are satisfied for the stochastic volatility model (X_t) under mild conditions on (σ_t).

Lemma 3.3.3. *Let (X_t) be a stochastic volatility model with a strongly mixing sequence (σ_n) with rate function $(\alpha_h(\sigma))$. Then (X_t) satisfies condition $D(u_n)$ with u_n satisfying $P(X_1 > u_n) \sim \frac{1}{n}$.*

Proof. We have for $B \subset \{1, 2, \dots\}$

$$\begin{aligned} P(\max_{i\in B} X_i \le u_n) &= E\left(P(\max_{i\in B} \sigma_i Z_i \le u_n|(\sigma_t))\right) \\ &= E\left(\prod_{i\in B} P(Z \le \frac{u_n}{\sigma_i}|\sigma_i)\right). \end{aligned}$$

Write

$$f(z) = P(Z \leq z), \quad z \in \mathbb{R}.$$

Then $f(Z) \leq 1$ and condition (3.3.2) reads as follows:

$$|E(\prod_{i \in A_1 \cup A_2} f(\frac{u_n}{\sigma_i})) - E(\prod_{i \in A_1} f(\frac{u_n}{\sigma_i}))E(\prod_{i \in A_2} f(\frac{u_n}{\sigma_i}))| \leq \alpha_{n,l}. \tag{3.3.3}$$

If (σ_t) is strongly mixing, it follows by the proof of Lemma 2.4.2 that (3.3.3) holds with $\alpha_{n,l} \leq 4\alpha_l(\sigma)$. □

Proposition 3.3.4. *Let (X_t) be a stochastic volatility model with iid regularly varying noise (Z_t) with index $\alpha > 0$. Assume that the following conditions hold*

1. *There exist integers $r_n \to \infty$ such that $r_n/n^{1-\varepsilon'} \to 0$ for some $\varepsilon' \in (0,1)$.*
2. *The sequence (X_n) is strongly mixing with rate function (α_n) such that $n\sum_{i=r_n+1}^{\infty} \alpha_i \to 0$ as $n \to \infty$.*
3. *$E\sigma_1^{2\alpha} < \infty$.*

Then (X_t) satisfies condition $D'(u_n)$ for (u_n) satisfying $P(X_1 > u_n) \sim \frac{1}{n}$.

Proof. Let (r_n) be as in condition (1). Then

$$n \sum_{i=2}^{[n/k]} P(X_1 > u_n, X_i > u_n) = n \sum_{i=2}^{r_n} P(X_1 > u_n, X_i > u_n) + n \sum_{i=r_n+1}^{[n/k]} P(X_1 > u_n, X_i > u_n).$$

By Markov's inequality and the definition of (u_n), for $\delta \in (0, \alpha)$,

$$\begin{aligned} n \sum_{i=2}^{r_n} P(X_1 > u_n, X_i > u_n) &= n \sum_{i=2}^{r_n} E(P(Z > \frac{u_n}{\sigma_1}|\sigma_1)P(Z > \frac{u_n}{\sigma_i}|\sigma_i)) \\ &\leq n u_n^{-2(\alpha-\delta)} \sum_{i=1}^{r_n} E(\sigma_1\sigma_i)^{\alpha-\delta} \\ &\leq n u_n^{-2(\alpha-\delta)} r_n const (E\sigma_1^{2(\alpha-\delta)})^{\frac{1}{2}} \\ &\leq n n^{-(\frac{1}{\alpha}-\varepsilon)(2(\alpha-\delta))} r_n const, \end{aligned} \tag{3.3.4}$$

for any $\varepsilon > 0$. Here we used that $E|X_1|^{\alpha-\delta} < \infty$ and

$$E(\sigma_1\sigma_i)^{\alpha-\delta} \leq (E\sigma_1^{2(\alpha-\delta)})^{\frac{1}{2}} < \infty,$$

by the Cauchy-Schwarz inequality and by condition (3). We also exploited the fact that $u_n = n^{\frac{1}{\alpha}} l(n)$ for some slowly varying function l such that $u_n \geq n^{\frac{1}{\alpha}-\epsilon}$ for small $\epsilon > 0$ and large n.

Hence

$$n \sum_{i=2}^{r_n} P(X_1 > u_n, X_i > u_n) \leq n^{-1+\varepsilon'} r_n const \to 0,$$

for some $\varepsilon' > 0$ as in condition (1).

We have

$$\begin{aligned} n\sum_{i=r_n+1}^{[n/k]} P(X_1 > u_n, X_i > u_n) &= n\sum_{i=r_n+1}^{[n/k]} [P(X_1 > u_n, X_i > u_n) - (P(X_1 > u_n))^2] \\ &\quad + n([\frac{n}{k}] - r_n)(P(X_1 > u_n))^2. \end{aligned}$$

By definition of u_n,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \frac{(nP(X_1 > u_n))^2}{k} = 0. \tag{3.3.5}$$

By definition of strong mixing,

$$n\sum_{i=r_n+1}^{[n/k]} |P(X_1 > u_n, X_i > u_n) - (P(X_1 > u_n))^2| \leq n \sum_{i=r_n+1}^{[\frac{n}{k}]} \alpha_{i-1} \leq n\sum_{i=r_n+1}^{\infty} \alpha_{i-1} \to 0,$$

by assumption (2). □

Remark 3.3.5. The above three conditions in Proposition 3.3.4 are satisfied if (σ_t) is strongly mixing with geometric rate $(\alpha_h(\sigma))$ and if $E\sigma_1^{2\alpha} < \infty$. Indeed, then we can choose $r_n = n^\gamma$ for any $\gamma < 1 - \epsilon'$ and $n\sum_{i=r_n}^{\infty} \leq na^{r_n} \to 0$ for some $a \in (0,1)$.

The following two standard theorems are helpful in constructing a result for the limiting distribution of the maxima of a stochastic volatility sequence.

Theorem 3.3.6. (Limit distribution of maxima of a stationary sequence [11, p. 215–216])
Let (X_n) be a strictly stationary sequence with common distribution function $F \in MDA(H)$ for some extreme value distribution H, i.e. there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\log H(x), \quad x \in \mathbb{R}. \tag{3.3.6}$$

Assume that for $x \in \mathbb{R}$ the sequence $(u_n) = (c_n x + d_n)$ satisfies the condition $D(u_n)$ and $D'(u_n)$. Then (3.3.6) *is equivalent to each of the following relations:*

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} Y, \tag{3.3.7}$$

$$c_n^{-1}(\widetilde{M}_n - d_n) \xrightarrow{d} Y, \tag{3.3.8}$$

where Y has distribution H.

Theorem 3.3.7. (Weak convergence of point processes of exceedances, stationary case [11, p. 243])
Suppose (X_n) is strictly stationary and (u_n) is a sequence of threshold values such that

$$n\overline{F}(u_n) \to \tau \in (0,\infty),$$

and $D(u_n)$ and $D'(u_n)$ hold. Let (N_n) be the point process of exceedances of a threshold u_n by the random variables $X_1, \dots, X_n$:

$$N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{\frac{i}{n}} I_{\{X_i > u_n\}}, \quad n = 1, 2, \dots .$$

Then $N_n \xrightarrow{d} N$, where N is a homogeneous Poisson process on $(0, 1]$ with intensity τ. The convergence holds in the space of point processes on $(0, 1]$.

The following corollary is an immediate consequence of Theorems 3.3.6 and 3.3.7.

Corollary 3.3.8. (The limiting distribution of the maxima of a stochastic volatility model) *Assume that (X_t) is a strictly stationary stochastic volatility process with regularly varying noise (Z_t) with index $\alpha > 0$. Assume that (X_n) is strongly mixing with rate function (α_h). Moreover, if the assumptions in Proposition 3.3.4 hold, then (X_n) satisfies the conditions $D(u_n)$, $D'(u_n)$ with (u_n) satisfying $P(X_1 > u_n) \sim 1/n$ and for $M_n = \max(X_1, \dots, X_n)$,*

$$P(u_n^{-1} M_n \le x) \to \Phi_\alpha(x) = e^{-x^{-\alpha}}, \quad x > 0.$$

In particular, (X_n) has extremal index 1.

Moreover, the point processes of exceedances $N_{n,x}$ converge in distribution to a homogeneous Poisson process $N_{n,x}$ on $(0, 1]$ with intensity $-\log \Phi_\alpha(x) = x^{-\alpha}, \quad x > 0$:

$$N_{n,x} = \sum_{i=1}^{n} \varepsilon_{\frac{i}{n}} I_{\{X_i > u_n x\}} \xrightarrow{d} N_x. \tag{3.3.9}$$

Remark 3.3.9. Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics of the sample $X_1, \dots, X_n$ from a stochastic volatility model satisfying the conditions of Corollary 3.3.8. An immediate consequence of (3.3.9) is that

$$\begin{aligned} P(N_{n,x}(0,1] < k) &= P(u_n^{-1} X_{(n-k+1)} \le x) \\ &\to P(N_x(0,1] < k) \\ &\quad = \Phi_\alpha(x) \sum_{i=0}^{k-1} \frac{(-\log \Phi_\alpha(x))^i}{i!} = \Phi_\alpha(x) \sum_{i=0}^{k-1} \frac{x^{-\alpha_i}}{i!}. \end{aligned}$$

3.4 The extremal behavior of a stochastic volatility model with regularly varying volatility sequence

So far we assumed that (Z_t) is iid regularly varying, independent of the volatility sequence (σ_t). Then $(X_t) = (\sigma_t Z_t)$ inherits regular variation from (Z_t) provided $E\sigma^{\alpha+\delta} < \infty$ for some $\delta > 0$. Moreover, (X_t) inherits strong mixing with essentially the same rate function as (σ_t).

We now investigate the case when $E|Z|^{\alpha+\delta} < \infty$ for some $\delta > 0$, but (X_t) is regularly varying with index $\alpha > 0$. We again assume that (σ_t) is strongly mixing with rate function $(\alpha_h(\sigma))$. Regular variation of (X_t) is easily verified. The one-dimensional case is simple: if σ is regularly varying, by Breiman's result in (3.1.19)

$$\begin{aligned} P(X > x) &\sim E(Z_+^\alpha)P(\sigma > x), \\ P(X \leq -x) &\sim E(Z_-^\alpha)P(\sigma > x), \end{aligned}$$

as $x \to \infty$. Hence X is regularly varying. A similar result holds for the finite-dimensional distributions

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & \cdots & 0 \\ 0 & \cdots & Z_n \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{pmatrix}.$$

A multivariate Breiman's result and the independence of $(Z_1, \ldots, Z_n)$ and $(\sigma_1, \ldots, \sigma_n)$ imply regular variation of $(X_1, \ldots, X_n)$. However, it can be tricky to verify that $(\sigma_1, \ldots, \sigma_n)$ is regularly varying, as the following example shows.

Example 3.4.1. Consider the AR(1) process $Y_t = \varphi Y_{t-1} + \eta_t$ for iid noise (η_t), $\varphi \in (0, 1)$. Choose $\sigma_t = e^{Y_t} = e^{\varphi Y_{t-1} + \eta_t}$. If σ_t is regularly varying with index $\alpha > 0$,

$$\begin{aligned} P(e^{\varphi Y_{t-1}} > x) &= P(\sigma^\varphi > x) = P(\sigma > x^{\frac{1}{\varphi}}) \\ &= x^{-\frac{\alpha}{\varphi}} L(x^{\frac{1}{\varphi}}), \end{aligned}$$

for a slowly varying function L. Hence σ^φ is regularly varying with index $\alpha/\varphi > \alpha$. We do expect that e^{η_t} is regularly varying. This is, however, not straightforward. One needs an inverse result to Breiman's result, i.e. we need to conclude from the product structure $e^{\varphi Y_{t-1}} e^{\eta_t}$ and the fact that $E(e^{\varphi Y_{t-1}})^{\alpha+\epsilon} = E\sigma^{\varphi(\alpha+\epsilon)} < \infty$ for some $\epsilon > 0$ that e^{η_t} is regularly varying. According to [18], one needs to verify the condition

$$E(\sigma^\varphi)^{\alpha + i\Theta} \neq 0, \quad \forall \Theta \in \mathbb{R}.$$

Using the AR(1) structure, we have

$$\begin{aligned} E(\sigma^\varphi)^{\alpha+i\Theta} &= E e^{\varphi(\alpha+i\Theta)\sum_{s=0}^\infty \varphi^s \eta_s} \\ &= \prod_{s=0}^\infty E e^{\varphi^{s+1}(\alpha+i\Theta)\eta} \neq 0, \quad \forall \Theta \in \mathbb{R}. \end{aligned}$$

This condition is satisfied if

$$E(e^{z(\alpha+i\Theta)\eta}) \neq 0, \quad \Theta \in \mathbb{R},\ 0 \leq z \leq \varphi.$$

If we assume that this condition is satisfied it is clear that the regular variation of σ is due to regular variation of e^{η}, and then by Breiman's result,

$$P(\sigma > x) = P(e^{\varphi Y_{t-1}+\eta_t} > x) \sim E(\sigma^{\varphi\alpha})P(e^{\eta} > x).$$

In what follows, assume that e^{η} is regularly with index $\alpha > 0$. This condition is satisfied if η is Exp(α) distributed:

$$P(e^{\eta} > x) = P(\eta > \log(x)) = e^{-\alpha \log(x)} = x^{-\alpha}, \quad x \geq 1.$$

We intend to show joint regular variation of $(\sigma_1, \ldots, \sigma_n)$. We have

$$(\sigma_1, \ldots, \sigma_n) = (\sigma_1, \sigma_1^{\varphi} e^{\eta_2}, \sigma_1^{\varphi^2} e^{\varphi\eta_2+\eta_3}, \ldots, \sigma_1^{\varphi^{n-1}} e^{\varphi^{n-2}\eta_2+\cdots+\eta_n}).$$

Since $E(\sigma_1^{\varphi})^{\alpha+\epsilon} < \infty, E(\sigma_1^{\varphi^2} e^{\varphi\eta_2})^{\alpha+\epsilon} < \infty, \ldots, E(\sigma_1^{\varphi^{n-1}} e^{\varphi^{n-2}\eta_2+\ldots+\varphi\eta_{n-1}})^{\alpha+\epsilon} < \infty$ for small $\epsilon > 0$ we expect that the regular variation of $(\sigma_1, \ldots, \sigma_n)$ follows from regular variation of the vector $(\sigma_1, e^{\eta_2}, e^{\eta_3}, \ldots, e^{\eta_n})$ which has independent components which are regularly varying, hence the whole vector is regularly varying with index $\alpha > 0$. We observe that

$$h(\sigma_1, e^{\eta_2}, \ldots, e^{\eta_n}) = (\sigma_1, \sigma_1^{\varphi} e^{\eta_2}, \sigma_1^{\varphi^2} e^{\varphi\eta_2+\eta_3}, \ldots, \sigma_1^{\varphi^{n-1}} e^{\varphi^{n-2}\eta_2+\cdots+\eta_n}),$$

is a continuous mapping such that $h^{-1}(B)$ is bounded for B bounded in $\overline{\mathbb{R}}^n \setminus \{\underline{0}\}$.

It is, however, not straightforward to use a continuous mapping argument to verify that

$$h(\sigma_1, e^{\eta_2}, \ldots, e^{\eta_n}),$$

inherits regular variation from regular variation of $(\sigma_1, e^{\eta_2}, \ldots, e^{\eta_n})$. This is due to the fact that h is not a homogeneous function, i.e. $h(tx) \neq t^g h(x),\ t > 0$, for some $g \in \mathbb{R}$. It remains an open problem whether $(\sigma_1, \ldots, \sigma_n)$ is jointly regularly varying.

Bibliography

[1] B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of GARCH process. *Stochastic Processes and their Applications*, 99:95–115, 2002.

[2] P. Billingsley. *Convergence of Probability Measures.* John Wiley, New York, 1968.

[3] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.

[4] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[5] P. Bougerol and N. Picard. Strict stationary of generalized autoregressive process. *The Annals of Probability*, 20(4):1714–1730, 1992.

[6] G.E.P. Box and G.M. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden–Day, San Francisco, 1970.

[7] L. Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability and its Applications*, 10, 1965.

[8] L. Breiman. *Probability.* Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[9] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer–Verlag, New York, second edition, 1991.

[10] P. Doukhan. *Mixing: Properties and Examples.* Springer–Verlag, New York, 1994.

[11] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, Berlin, Heidelberg, New York, 1997.

[12] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.

[13] R.F. Engle. *ARCH Selected Readings.* Oxford University Press, Oxford (UK), 1995.

[14] B. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

[15] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer series in statistics. Springer–Verlag, New York, 2003.

[16] W. Feller. *An Introduction to Probability Theory and its Applications.*, volume II. John Wiley and Sons, New York, London, Sydney, 1971.

[17] I.A. Ibragimov and Yu. V. Linnik. *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen, 1971.

[18] M. Jacobsen, T. Mikosch, J. Rosinski, and G. Samorodnitsky. Inverse problems for regular variation of linear filters, a cancellation property for σ-finite measures, and identification of stable laws. *Annals of Applied Probability*, to appear, 2008.

[19] A. Jakubowski. Minimal conditions in p-stable limit theorems. *Stochastic Processes and their Applications*, 44:291–327, 1993.

[20] A. H. Jessen and T. Mikosch. Regularly varying functions. *Publications de l'Institut Mathématique, Nouvelle Série*, 80:171–192, 2006.

[21] A. F. Karr. *Probability.* Springer–Verlag, New York, 1993.

[22] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.

[23] C. Klüppelberg and T. Mikosch. The integrated periodogram for stable processes. *The Annals of Statistics*, 24:1855–1879, 1996.

[24] U. Krengel. *Ergodic Theorems.* Walter de Gruyter, 1985.

[25] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, 2005.

[26] R. C. Merton. The theory of rational option pricing. *Bell Jornal of Economics and Management Science*, 4:141–183, 1973.

[27] T. Mikosch. Modeling dependence and tails of financial time series. In Bärbel Finkenstädt and Holger Rootzén, editors, *Extreme Values in Finance, Telecommunications, and the Environment*, pages 185–286. Chapman and Hall/CRC, Boca Raton, 2004.

[28] T. Mikosch. *Non-Life Insurance Mathematics.* Springer–Verlag, Berlin, Heidelberg, 2004.

[29] T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics*, 28:1427–1451, 2000.

[30] A. Mokkadem. Propriétés de mélange des modèles autorégressifs polynomiaux. *Annales Institut Henry Poincaré*, 26:219–260, 1990.

[31] P. A. P. Moran. The statistical analysis of the Canadian lynx cycle, I:Structure and prediction. *Australian Journal of Zoology*, 1:163–173, 1953.

[32] D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370, 1991.

[33] T. D. Pham and L. T. Tran. Some mixing properties of time series models. *Stochastic Processes and their Applications*, 19:297–303, 1985.

[34] S. I. Resnick. Point processes, regular variation, and weak convergence. *Advances in Applied Probability*, (18):66–138, 1986.

[35] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer, New York, 1987.

[36] W. Rudin. *Functional Analysis.* McGrow-Hill, 1973.

[37] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, London, 1994.

[38] W. F. Stout. *Almost Sure Convergence.* Academic Press, New York, 1974.

[39] D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models.* Number 181 in Lectures Notes in Statistics. Springer–Verlag, Berlin, 2005.

[40] H. Weyl. Über die Gleichverteilung von Zahlen mod Eins. *Mathematische Annelon*, 77:313–357, 1916.

[41] G. U. Yule. On a method of investigating periodicities in distributed series with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society*, A 226:267–298, 1927.