Nonlinear time series models and their extremes, with hydrological applications

Ph.D. Dissertation

Péter Elek

Supervisor: András Zempléni
Associate Professor, CSc

Doctoral School of Mathematics
Director:
Professor Miklós Laczkovich, member of HAS

Doctoral Programme of Applied Mathematics
Director:
Professor György Michaletzky, DSc

Eötvös Loránd University
Faculty of Sciences
Department of Probability Theory and Statistics
Budapest, May 2009

Contents

Chapter 1

Introduction

1.1 Motivation

By combining the methods of time series analysis and extreme value theory, this dissertation focuses on the extremal behaviour and parameter estimation of certain nonlinear time series processes. The theoretical contributions presented here were initially motivated by an applied hydrological project that aimed to build models capable of capturing both the times series dynamics and the extremal behaviour of real water discharge data sets.

Since it is natural to ask questions such as "how often will a river exceed a certain high value" or "what is the average duration of a flood above a given high level", hydrology has always played a crucial role in the development of extreme value theory (EVT). In fact, one of the earliest statistical estimation problems associated with EVT was raised after the disastrous flooding of two Dutch provinces in 1953, and concerned the determination of the height of the sea dykes such that the probability of a future flood lies below a certain pre-specified level (de Haan, 1990). Since then, EVT has acquired a recognised place both in theoretical and applied statistics, with numerous monographs appearing in the field (e.g. Embrechts et al. (1997) quickly becoming a classic) and leading figures of EVT being acknowledged also in the wider statistics community.

EVT is concerned with limit theorems about the structure of rare events in an independent identically distributed (i.i.d.) sample or in a time series. For instance, it can be proven under general conditions that the properly normalised sample maximum of i.i.d. observations asymptotically follows a parametric distribution called the generalised extreme value law. Similarly, as the threshold goes to the upper endpoint of the support of a distribution, the normalised excess above this threshold converges in distribution to the generalised Pareto law. Or, in the time series setting, the extreme observations in a stationary process under reasonable assumptions tend to occur in clusters, and the clustering tendencies can be characterised by a single number, the extremal index, the reciprocal of which gives the average size of an extremal cluster.

In the absence of other information, the parametric statistical methodology based on these limit results provides the most appropriate way of estimating the high quantiles of a distribution or the extremal clustering of a time series. They are indeed routinely used in telecommunications (in analysing periods with high rate of data transmission), finance (in conducting Value-at-Risk calculations), climate studies or in the above mentioned hydrology. However, the purely EVT-based estimation procedures rely only on a small fraction of the sample (on the highest observations), thus they require large sample sizes and can be very imprecise in small samples or in a nonstationary environment. Furthermore, they do not make use of additional, "physical" information on the dynamics of the *whole* data generating process, hence they are inefficient in cases when "background" information exists. In such instances studying the extremal behaviour of the "theoretical" data generating process may increase the accuracy of extremal event estimation and simulation, as it has been demonstrated with the analysis of heavy-tailed processes in telecommunications or of conditionally heteroscedastic models in finance. In fact, the current world financial turmoil clearly shows that no matter how sophisticated methodology is used, relying solely on the statistical information of the not-too-distant past, without having an appropriate model for the observed process, may be misleading and may not accurately provide the true probability of extreme events.[1]

Hydrology happens to be a field where a vast amount of knowledge about the dependence structure of discharge processes has accumulated over time. Hydrologists have always been concerned with developing simulation models useful to the design of reservoirs, and it was already noticed by Brochu (1978) that even modest improvements in the operation of reservoirs can lead to multi-million dollar savings per year. No wonder that new statistical techniques have quickly found their way into the hydrological practice. For instance, an intuitive definition of long range dependence was first given by Harold Edwin Hurst, a civil engineer who analysed discharge series of river Nile (Hurst, 1951). Since then, fractional ARIMA, regime switching and shot-noise processes, and other time series

[1] The unwarranted use of statistical techniques (among them extreme value ones) in mathematical finance has been criticised in the extreme value community long before the current financial crisis. For instance, Thomas Mikosch, a leading researcher in EVT initiated a discussion at the Extreme Value Analysis conference in 2005 by opposing the use of copula methods on the ground that the transformation of the data into uniform marginals (the essence of copula techniques) hides the "true" data generating process from the eyes of the modeler. Instead, more attention should be paid to the real stochastic process behind the observed phenomena. See Mikosch (2006) and the discussion in the same special issue of *Extremes*, particularly the rejoinder.

models have been frequently developed and applied in the hydrological context. Given this accumulated knowledge, it is not surprising that the combination of time series analysis and extreme value theory proves to be practically useful in hydrology as well. This has motivated our theoretical and empirical contribution.

1.2 Methods and main results

It is generally agreed that, unlike rivers with small catchments, discharge series of larger rivers are not as heavy tailed as e.g. financial series: their moments of all orders seem to be finite. Partly because of the relatively smaller attention received, extremal behaviour of such nonlinear time series models are often less studied than that of ARCH or other heavy-tailed processes, and often different methods are needed to prove extremal results on them.

In the dissertation I develop and examine two classes of such models. Both classes arose from our empirical observation that linear models – even after allowing for long range dependence and non-Gaussian innovations – are not adequate to describe the extremes of water discharge series of rivers Danube and Tisza in Hungary. As a first attempt to solve this puzzle, I present an *ARMA-ARCH-type model* of the following form:

$$\begin{aligned} X_t &= c_t + \sum_{i=1}^{p} a_i (X_{t-i} - c_{t-i}) + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \\ \varepsilon_t &= \sigma(X_{t-1}) Z_t, \end{aligned}$$

where c_t is a deterministic periodic function, Z_t is an independent identically distributed random sequence with zero mean and unit variance and, most importantly, $\sigma^2(x)$ is increasing slower than quadratic as $x \to \infty$. Besides the unusual feedback structure (i.e. that the conditional variance depends on X_{t-1} and not on ε_{t-1}), the slower than quadratic increase of $\sigma^2(x)$ makes this conditionally heteroscedastic model substantially different from the usual ARCH-specifications. Using the drift condition for Markov chains, I prove that – unlike of the quadratic ARCH-case – all moments of the stationary distribution of this model are finite provided that the corresponding linear ARMA model is stationary and invertible. More interestingly, applying various approximation techniques, I give a Weibull-like lower and upper bound for the tail of the stationary distribution. The approximation implies that the model is still heavy-tailed for certain parameter values in the sense that its moment generating function is infinite for all $z > 0$. Turning to the examination of extremal dependence, I illustrate by simulations that the model does not exhibit clustering at extreme levels.

Parameter estimation is carried out by a combination of least squares and maximum likelihood, and I prove consistency and asymptotic normality of the estimator.

Although simulations show that the model is able to reproduce the probability density, high quantiles and clustering tendencies of river discharge series at *practically relevant* thresholds, asymptotically it exhibits less (i.e. zero) clustering and heavier (i.e. close to Weibull-like) tail than hydrologists tend to assume about such series. Moreover, the model does not give back the practically important pulsatile nature of river flows, i.e. that short but steep rising periods are followed by longer, gradually falling ones. This leads to the second class of analysed models, *Markov-switching autoregressive processes*. Again, I concentrate on the case previously less examined in the literature by assuming that the process behaves as a random walk in one of the regimes, and that the noise distribution is light-tailed. More precisely, the model can be written as:

$$\begin{aligned} X_t &= \quad X_{t-1} + \varepsilon_{1,t} \quad \text{if} \quad I_t = 1 \\ X_t &= a_0 X_{t-1} + \varepsilon_{0,t} \quad \text{if} \quad I_t = 0, \end{aligned}$$

where $|a_0| < 1$, $\varepsilon_{i,t}$ are i.i.d. sequences $(i = 0, 1)$, independent of each other as well, $\varepsilon_{1,t}$ is light-tailed and I_t is a Markov chain. Using the drift condition for Markov chains and methods of renewal theory and extreme value theory, I prove under mild additional assumptions that the process has asymptotically exponential upper tail and its extremal values form nontrivial clusters. The extremal index is obtained in terms of the solution of a Wiener-Hopf equation, which can be solved explicitly in special cases. Using Laplace's method for sums, I also obtain Weibull-like bounds for the tail of the distribution of a practically important measure of extremal clustering, the limiting aggregate excess functional.

Instead of fitting a Markov-switching autoregressive (MS-AR) model to the whole river discharge series, I propose a different approach, which is more tailor-made for extremes. I prove that the limiting extremal behaviour of general Markov-switching, conditionally Markovian models can be approximated by MS-AR structures. Based on this result, I fit by maximum likelihood such a limiting MS-AR representation using only high-level exceedances of the river flow series, examine the properties of this estimator and obtain estimates for extremal clustering from the fitted representation. Simulations show that the extremal clustering behaviour obtained this way provides a reasonable approximation to the observed clustering of river flow series.

The dissertation is organised as follows. Chapter 2 gives preliminary results of time series analysis and extreme value theory, which are necessary to understand the further chapters. Chapter 3 presents the empirical features of river flow series and thus restricts the types of nonlinear time series processes that are considered in the dissertation. Chapter 4 deals with the stationarity, tail behaviour, extremes, parameter estimation and hydrological application of ARMA-ARCH type models, while Markov-switching models are examined in Chapter 5. Finally, Chapter 6 explores the relationships between the two classes of models and draws the conclusions.

1.3 Articles of the author

The dissertation is based on the following four peer-reviewed journal articles and one yet unpublished manuscript:

- Elek, P., Márkus, L., 2004. A long range dependent model with nonlinear innovations for simulating daily river flows. *Natural Hazards and Earth Systems Sciences* 4, 277-283.

- Elek, P., Márkus, L., 2008. A light-tailed conditionally heteroscedastic time series model with an application to river flows. *Journal of Time Series Analysis* 29, 14-36.

- Elek, P., Zempléni, A., 2008. Tail behaviour and extremes of two-state Markov-switching autoregressive models. *Computers and Mathematics with Applications* 55, 2839-2855.

- Elek, P., Zempléni, A., 2009. Modelling extremes of time-dependent data by Markov-switching structures. *Journal of Statistical Planning and Inference* 139, 1953-1967.

- Elek, P., Márkus, L., 2009. Tail behaviour of β-TARCH processes. *Manuscript*, submitted.

The author has also published two conference proceedings and various conference abstracts in the topic. Results on ARMA-ARCH-type models are joint work with László Márkus, while results on Markov-switching structures are joint work with András Zempléni.

Although the following three journal articles do not form the basis of the dissertation, they are loosely connected to the current topic:

- Arató, M., Bozsó, D., Elek, P., Zempléni, A., 2008. Forecasting and simulating mortality tables. *Mathematical and Computer Modelling* 49, 805-813.

- Bíró, A., Elek,P., Vincze, J., 2008. Model-based sensitivity analysis of the Hungarian economy to shocks and uncertainties. *Acta Oeconomica* 58, 367-401.

- Vasas, K., Elek, P., Márkus, L., 2007. A two-state regime switching autoregressive model with application to river flow analysis. *Journal of Statistical Planning and Inference* 137, 3113-3126.

Vasas et al. (2007) deals with MCMC-based estimation of a more general regime switching autoregressive model than considered in this dissertation. Arató et al. (2008) and Bíró et al. (2008) are applied papers in the fields of insurance and economics. Finally, the MA thesis Elek (2003) applies volatility modelling and extreme value techniques to Value At Risk calculations on the stock market in Hungary.

1.4 Acknowledgement

I would like to express my thanks to my supervisor András Zempléni and to László Márkus for piqueing my interest in statistics, and for the joint work and stimulating discussions with them in extreme value theory and time series analysis. I thank my employer, the Ministry of Finance for having allowed me to work on the PhD thesis besides my workplace duties in the last four years. The hydrological data and financial support provided by the Water Resources Research Centre (VITUKI) in the framework of the "Flood Risks" Hungarian National Research Project are also greatly acknowledged.

I would like to thank my parents for the care and support they gave to me. Finally, but most importantly, my greatest gratitude goes to my wife Krisztina, who not only helped me in technical details (being the coauthor of one of my papers) but also provided encouragement every day while I was working on the thesis. Surely, I could not have reached this point without her help.

1.5 Notations

$c_X(s)$	$\log E\exp(sX)$, the cumulant generating function of X
$F_X(x)$	$P(X < x)$, the distribution function of X
$\bar{F}(x)$	$1 - F_X(x)$, the survival function of X
$f_X(x)$	the density function of X
$L_X(s)$	$E\exp(sX)$, the moment generating function of X
$M_{k,n}$	$\max(X_k, \ldots, X_n)$
$\mathbf{R}_+$	$[0, \infty)$
$\mathbf{R}_{++}$	$(0, \infty)$
$\mathbb{R}_+^m$	$[0, \infty)^m$
$\mathbb{R}_{++}^m$	$(0, \infty)^m$
ρ_i	$\rho(X_t, X_{t-i})$, autocorrelation function of X_t
x^+	$\max(x, 0)$
x^-	$\max(-x, 0)$
x_F	$\sup\{x : F(x) < 1\}$, the upper end point of the support of the distribution
$\|\|\mathbf{x}\|\|_r$	r-norm of a vector $\mathbf{x}$
$\|\|X\|\|_{L^r}$	$(E(X^r))^{1/r}$, r-norm of a random variable X

Chapter 2

Preliminaries

2.1 Preliminaries in time series analysis

Let us first recall a few basic concepts from time series analysis, which can mainly be found in classical monographs such as Brockwell and Davis (1991). In the dissertation, we will focus on time series models in discrete time that have a stationary distribution (possibly apart from a seasonal trend in the mean and variance). Denoting the time series by X_t, (strict) stationarity means that for any $(i_1, i_2, \ldots, i_n) \in \mathbb{Z}^n$ and for any $t \in \mathbb{Z}$,

$$(X_{i_1}, X_{i_2}, \ldots, X_{i_n}) =^d (X_{t+i_1}, X_{t+i_2}, \ldots, X_{t+i_n})$$

where, as usual, $=^d$ indicates that the distributions of the left and right hand sides are identical. In the following, unless stated otherwise, all probability statements on X_t will refer to probabilities under the (unique) stationary distribution.

Classical time series analysis is mainly concerned with linear processes of the form

$$X_t = \mu + \sum_{i=0}^{\infty} \pi_i \varepsilon_{t-i} \tag{2.1}$$

where $\mu = E\,(X_t)$ and $\{\varepsilon_t\}$ is a zero-mean i.i.d. sequence of random variables with finite variance $\sigma^2_{\varepsilon_t}$ and $\{\pi_i\}$ is a real-valued sequence with $\pi_0 = 1$. In this case, if

$$\sum_{i=0}^{\infty} |\pi_i| < \infty, \tag{2.2}$$

the sum defining X_t converges a.s. and X_t has finite variance.

One of the simplest linear models is the (zero-mean) ARMA(p,q) model, which is defined by

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=0}^{q} b_i \varepsilon_{t-i}, \tag{2.3}$$

where a_i $(i = 1, \ldots, p)$ and b_i $(i = 0, \ldots, q)$ are real numbers, $b_0 = 1$ and $\{\varepsilon_t\}$ is a zero-mean i.i.d. sequence with finite variance. We also assume that

$$\Phi(z) = 1 - \sum_{i=1}^{p} a_i z^i \neq 0 \quad \text{and} \quad \Psi(z) = 1 + \sum_{i=1}^{q} b_i z^i \neq 0 \quad \text{if} \quad |z| \leq 1, \tag{2.4}$$

and $\Phi(z)$ and $\Psi(z)$ have no common zeros.

The condition on $\Phi(z)$ ensures that the model has a unique stationary solution in a causal form,[1] i.e. it can be written as (2.1) with (2.2) satisfied. The generating function of $\{\pi_i\}$ is given by

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i = \Psi(z) / \Phi(z).$$

Using the backward shift operator B, the stationary solution can be written as

$$X_t = (\Phi(B))^{-1} \Psi(B)\varepsilon_t.$$

The condition on $\Psi(z)$ in (2.4) ensures that X_t is invertible, i.e.

$$\varepsilon_t = \sum_{i=0}^{\infty} \eta_i X_{t-i} = (\Psi(B))^{-1} \Phi(B) X_t,$$

where $\sum_{i=0}^{\infty} |\eta_i| < \infty$. Throughout the dissertation, we will focus on causal and invertible ARMA models. (For a more detailed discussion of causality and invertibility, see Brockwell and Davis (1991, Chapter 3).)

Parameter estimation of ARMA processes can be carried out in many ways. Since in the presence of MA-terms full maximum likelihood estimation may not be straightforward we present here, for later reference, the simplest form of the least squares procedure. For any vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q})$, define

$$e_t(\boldsymbol{\theta}) = X_t - \sum_{i=1}^{p} \theta_i X_{t-i} - \sum_{j=1}^{q} \theta_{p+j} e_{t-j}(\boldsymbol{\theta})$$

with $e_i(\boldsymbol{\theta}) = 0$ and $X_i = 0$ for all $i \leq 0$. If (2.4) holds and ε_t has finite variance, the maximisation of the residual sum of squares

$$Q = \sum_{t=1}^{n} e_t^2(\boldsymbol{\theta})$$

[1] The unique stationary distribution exists under more general conditions, i.e. when $\Phi(z) \neq 0$ for $|z| = 1$. However, when $\Phi(z)$ has a zero inside the open unit disk the ARMA-solution is non-causal, i.e. X_t depends on future values of $\{\varepsilon_t\}$ as well.

leads to a consistent and asmptotically normal estimator.[2]

Goodness of fit of an ARMA model is usually examined by the Box-Pierce (BP) test or its various modifications such as the Ljung-Box (LB) test. Denoting the estimated i-lag autocorrelation function of the fitted innovation sequence $\hat{\varepsilon}_t$ by $\hat{\nu}_i = \hat{\rho}\left(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-i}\right)$, the test statistics are given by

$$Q_{BP} = n \sum_{i=1}^{r} \hat{\nu}_i^2 \quad \text{and} \quad Q_{LB} = n\left(n+2\right) \sum_{i=1}^{r} \frac{\hat{\nu}_i^2}{n-i}, \tag{2.5}$$

respectively. If the true model is a (linear) ARMA(p, q) process, both Q_{BP} and Q_{LB} follow asymptotically a χ^2_{r-p-q}-distribution (however, the Ljung-Box statistic has more favourable small sample properties). Thus in large samples a χ^2-test can be constructed to test the appropriateness of the model.

ARMA models are routinely used in natural sciences, economics and other fields to describe the linear dependence structure of observed phenomena. Their popularity partly stems from the easy to implement ARMA fitting procedure known as the Box-Jenkins methodology (Box and Jenkins, 1970). The procedure consists of five steps: in the first step differencing is applied until the modelled process becomes stationary, while in the second the orders p and q of the ARMA-process are identified by the inspection of the autocorrelation and partial autocorrelation functions. The third step deals with model estimation (e.g. with least squares), the fourth with investigating goodness of fit (e.g. by the Box-Pierce or Ljung-Box tests), and the fifth step covers forecasting. Nowadays, statistical softwares make it easy to fit ARMA models even in the applied statistical practice.

Basically, there are two possibilities to generalise the ARMA model that are of interest for us: to allow long memory, and to allow nonlinearities.

First, denoting the autocorrelation function by $\rho_i = \rho\left(X_t, X_{t-i}\right)$, it can be shown easily that ρ_i decays exponentially in the case of ARMA processes as $i \to \infty$. This behaviour is too restrictive for some applications e.g. in hydrology or telecommunications, thus there is considerable research in the field of the so-called long range dependent processes. The following is one possible definition of long memory. (The concepts and basic results below can be found e.g. in the monograph of Beran (1994).) As usual, we use the notation $a(x) \sim b(x)$ as $x \to \infty$ if $\lim_{x\to\infty} a(x)/b(x) = 1$.

Definition 2.1. *A stationary process X_t is called long range dependent (LRD) if there exist $K > 0$ and $1/2 < H < 1$ such that $\rho_i \sim Ki^{2H-2}$ as $i \to \infty$.*

In such a case H is called the Hurst-parameter. As the following Propositions state LRD processes behave very differently than processes with exponentially decaying auto-

[2] Various small-sample improvements to this procedure (such as back-forecasting) exist.

correlations. Their spectral density has a pole at zero, and their aggregate variance grows faster than Kn as $n \to \infty$.

Proposition 2.2. *Let $\eta(x)$ denote the spectral density of the stationary process X_t. X_t is LRD if and only if there exist $K > 0$ and $1/2 < H < 1$ such that $\eta(x) \sim K|x|^{1-2H}$.*

Proposition 2.3. *If X_t is LRD there exists $K > 0$ such that, as $n \to \infty$,*

$$D^2\left(\sum_{t=1}^{n} X_t\right) \sim Kn^{2H}.$$

The easiest way to model LRD series is by fractional ARIMA processes. With the notations of (2.4), a zero-mean fractional ARIMA (FARIMA) process X_t satisfies

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t \tag{2.6}$$

where d is the order of fractional differencing (lying between 0 and 0.5 in cases of our interest) and ε_t is an independent zero-mean innovation (noise) sequence with variance σ_ε^2. If $0 \leq d < 0.5$ and all roots of the polynomials $\Phi(z)$ and $\Psi(z)$ lie outside the unit circle the model is stationary. Then, the $d = 0$ case gives a simple ARMA process, while if $0 < d < 1/2$, the process is LRD with $H = d + 1/2$.

Parameters of FARIMA models (including d and the short-run coefficients) can be estimated by exact Gaussian maximum likelihood or by the Whittle-procedure. The latter basically depends on the approximation of the Gaussian likelihood in the spectral domain, and is consistent and asymptotically normally distributed – under mild regularity conditions which FARIMA models satisfy – for linear processes driven by i.i.d. ε_t innovations with finite fourth moments (Giraitis and Surgailis, 1990).

Hence a straightforward parametric method to estimate the Hurst-parameter of a series is to fit a FARIMA model and obtain $H = d+1/2$ from the estimated d coefficient. Alternatively, there exist various nonparametric estimators which make use of certain properties of LRD series. For instance, the autocorrelation-based method estimates H by regressing the logarithms of the estimated sample autocorrelations $\hat{\rho}_i$ on $\log i$ for large values of i, while the periodogram-based method and the aggregate variance method utilise Propositions 2.2 and 2.3, respectively. Other popular procedures include the rescaled-range statistic and the Geweke-Porter-Hudak estimator.

Leaving the field of LRD models, a second possible deviation from ARMA processes is the modelling of nonlinearities. It follows from Wold's decomposition, one of the basic results of time series analysis, that a stationary time series with finite variance and vanishing memory[3] can be written in the form of (2.1) where ε_t are zero-mean uncorrelated random variables with finite variance. (See e.g. Bierens (2004, Chapter 7).) This "weak" MA(∞)-representation may be useful for forecasting purposes but it is not a proper data generating process, i.e. it does not give a full probabilistic description of time-dependence. The class of nonlinear models, therefore, is very wide and difficult to characterise because nonlinearities may arise for several reasons. A thorough treatment of nonlinear time series analysis is given in Tong (1990).

If, furthermore, the nonlinear model can be represented as (2.3) with ε_t an uncorrelated sequence, it is said to have a "weak" ARMA-representation (as opposed to the strong ARMA-representation when ε_t is i.i.d.). Between the weak and the strong cases there is the semi-strong ARMA-representation, when ε_t is assumed to be a martingale difference sequence. Both classes of nonlinear processes examined in this dissertation (the ARMA-ARCH-type model and the Markov-switching autregressive model) possess a weak ARMA-representation.

2.2 Preliminaries in extreme value theory

A fast growing field of research, extreme value theory (EVT) is concerned with the analysis and estimation of high observations of data generating processes. EVT for i.i.d. data gives limit results on the distribution of the (properly normalised) maximum of a high number of observations, or on the distribution of observations above a high threshold, assuming as little as we can about the underlying distribution of observations. If the i.i.d. assumption is relaxed, further questions arise about the pattern of clustering among high observations (i.e. how these observations occur together in time), which is the topic of extreme value theory for dependent sequences.

The probabilistic results then provide the basis of parametric estimation techniques for high quantiles of a distribution or for the clustering tendencies of a time series. Nowadays, these procedures have a wide range of applications in e.g. actuarial science, finance, economics, telecommunications, hydrology or environmental modelling.

In this section we outline the main results of extreme value theory of i.i.d. observations first, and then focus on its extension for stationary sequences. Unless otherwise indicated,

[3] A time series X_t has vanishing memory if all the events in the σ-algebra $\mathfrak{F}_X^{-\infty} = \cap_{t=0}^{\infty}\sigma\left(X_{-t}, X_{-t-1}, \ldots,\right)$ have 0 or 1 probability. This is a natural assumption for series of practical interest. If the vanishing memory condition is not assumed, the stationary time series can only be written as the sum of an MA(∞)-term and a deterministic term.

these basic results can be found in the monographs of Embrechts et al. (1997) or Coles (2001).

2.2.1 Extreme value theory for i.i.d. random variables

Let $\{X_i\}$ be an i.i.d. sequence with common distribution function F and let us denote the maximum of the first n observations by $M_{1,n} = \max\left(X_1, X_2, \ldots, X_n\right)$. If $x_F = \sup\{x : F(x) < 1\}$ is the upper end point of the support of the distribution, then it is clear that $M_{1,n} \to x_F$ a.s. as $n \to \infty$. Thus, we are rather interested in the non-degenerate limit of $M_{1,n}$, properly normalised. The Fisher-Tippet theorem states that the limit distribution, if exists, is in the class of the so-called generalised extreme value (GEV) distributions.

Definition 2.4. *The generalised extreme value distribution with shape parameter ξ has the following distribution function. If $\xi \neq 0$,*

$$F_{\xi}^{GEV}(x) = \exp\left[-\left(1+\xi x\right)^{-1/\xi}\right]$$

for $1 + \xi x > 0$ (and otherwise 0 if $\xi > 0$ and 1 if $\xi < 0$.) If $\xi = 0$,

$$F_{\xi}^{GEV}(x) = \exp\left[-e^{-x}\right].$$

The $\xi = 0$ case can also be obtained from the $\xi \neq 0$ case by letting $\xi \to 0$. The name of the distribution is Frechet for $\xi > 0$, Gumbel for $\xi = 0$ and Weibull for $\xi < 0$. We can also define the corresponding location-scale family $F_{\xi,\mu,\sigma}^{GEV}$ by replacing x above by $\left(x-\mu\right)/\sigma$ for $\mu \in \mathbb{R}$ and $\sigma > 0$ and changing the support accordingly. This latter family is also referred to as GEV.

Theorem 2.5. *(Fisher and Tippet, 1928) If there exist $\{a_n\}$ and $\{b_n\} > 0$ sequences such that*

$$\frac{M_{1,n} - a_n}{b_n} \to_d Z$$

where Z is a non-degenerate random variable, then Z is distributed as GEV. In this case we say that the distribution of X_i belongs to the max-domain of attraction of a generalised extreme value distribution.

More importantly from our point of view, the fact that the distribution of a random variable X lies within the max-domain of attraction of a GEV, implies that the distribution of exceedances of that random variable above a high threshold has a particular limiting representation. To be precise, using the notation $F = F_X$, the following theorem, due to Balkema and de Haan, holds.

Theorem 2.6. *(Balkema and de Haan, 1974) The distribution of X lies within the domain of attraction of a GEV with shape parameter ξ if and only if there exists a positive measurable function $a(u)$ such that for all $1+\xi x>0$*

$$\lim_{u\to x_F}\frac{\bar{F}\left(u+xa\left(u\right)\right)}{\bar{F}\left(u\right)}=\bar{F}_{\xi}^{GPD}\left(x\right) \tag{2.7}$$

where $\bar{F}_{\xi}^{GPD}$ stands for the survival function of the generalised Pareto distribution with shape parameter ξ (defined in the following).

Definition 2.7. *The generalised Pareto distribution (GPD) with shape parameter ξ has the following form. Its support is $[0,\infty]$ if $\xi\geq 0$ and $[0,-1/\xi]$ if $\xi<0$. If x is within the support, the distribution function can be written as*

$$\begin{aligned} &F_{\xi}^{GPD}\left(x\right)=1-\left(1+\xi x\right)^{-1/\xi} && \textit{if}\quad \xi\neq 0\\ &F_{\xi}^{GPD}\left(x\right)=1-\exp\left(-x\right) && \textit{if}\quad \xi=0. \end{aligned}$$

One can also introduce the location-scale family $F_{\xi,\mu,\sigma}^{GPD}$ by replacing x by $\left(x-\mu\right)/\sigma$ for $\mu\in\mathbb{R}$ and $\sigma>0$ and by adjusting the support accordingly. $F_{\xi,\mu,\sigma}^{GPD}$ will also be referred to as GPD.

Note that the $\xi=0$ case is just the exponential distribution, and it can be obtained from the $\xi\neq 0$ case by letting $\xi\to 0$.

Thus, intuitively, Theorem 2.5 states that normalised maxima of many i.i.d. random variables follow approximately a GEV distribution, while Theorem 2.6 states that the scaled excesses over high thresholds (i.e. $\left(X-u\right)/a\left(u\right)$ provided that $X\geq u$) are approximately GPD. These facts can be used to construct nowadays commonly used estimation methods for the distribution of maxima and high quantiles.

One of the most popular estimation procedures is the peaks over threshold (POT) method. In the i.i.d. setting this consists of estimating high quantiles of an observed distribution by fitting a GPD to exceedances above a sufficiently high threshold by maximum likelihood and then calculating the quantiles of the estimated GPD. (The maximum likelihood estimator is consistent and asymptotically normal if $\xi>-1/2$.) Although in the absence of other information this is an appropriate quantile estimating procedure, threshold choice may be difficult during its application because of the bias-variance tradeoff. Since Theorem 2.6 is an asymptotic result choosing a low threshold may lead to an estimation bias, while a high threshold necessarily yields an increased variance of the parameter estimates. However, the GPD has an interesting property: if $X\sim GPD_{\xi,\mu,\sigma}$ then $X|\left(X>u\right)\sim GPD_{\xi,\mu(u),\sigma(u)}$, i.e. taking threshold exceedances does not alter the shape parameter. Hence the bias-variance problem can be partly resolved by estimating the GPD model above various thresholds and choosing one above which the shape parameter looks roughly constant.

The EVT-based statistical procedures implicitly assume that most distributions of practical interest lie within the max-domain of attraction of a GEV (and hence in the domain of attraction of a GPD). Therefore, a natural question arises: how wide is the class of distributions for which the limit results in Theorems 2.5 and 2.6 hold? It is not difficult to find counterexamples (e.g. among discrete distributions) but most "well-behaved" continuous distributions lie within the domains of attraction. In the Frechet- and Weibull-cases (i.e. when $\xi \neq 0$) there exist relatively easily verifiable conditions.

A distribution function F belongs to the domain of attraction of a GPD with $\xi > 0$ (i.e. to the max-domain of attraction of the Frechet distribution) if and only if $\bar{F}$ is regularly varying, that is, $\bar{F}\left(u\right) = R(u)u^{-1/\xi}$ where $R(u)$ is a slowly varying function. A function is said to be slowly varying if

$$\lim_{u\to\infty} \frac{R\left(au\right)}{R\left(u\right)} = 1 \quad \text{for all} \quad a > 0.$$

(An example of a nonconstant slowly varying function is the logarithmic function.) Thus we can say that the domain of attraction of the Frechet-distribution roughly contains distributions with polynomially decaying survival function. It follows easily from the characterisation that if the distribution of X belongs to the Frechet(ξ)-domain, then $E\left(X^{+}\right)^{m}$ is infinite for $m > 1/\xi$ and finite for $m < 1/\xi$. (Here we use the notation $x^{+} = \max(0, x)$.) Hence such a distribution has infinite moments for sufficiently large m. Examples for distributions belonging to the domain of attraction of the Frechet-distribution include the Pareto, the Cauchy and (for $\alpha < 2$) the stable distributions.

In contrast to these heavy-tailed distributions, the Weibull-domain ($\xi < 0$) contains distributions which have a finite right endpoint $x_F < \infty$. More precisely, the distribution function F belongs to the domain of attraction of the Weibull-distribution with shape parameter $\xi < 0$ if and only if $x_F < \infty$ and $\bar{F}\left(x_F - x^{-1}\right) = x^{1/\xi} R\left(x\right)$ for some slowly varying function R. Examples for this domain of attraction include the uniform and the beta distributions.

The third, Gumbel-case ($\xi = 0$) is much more complicated. Although there exist necessary and sufficient conditions here as well (in terms of von Mises functions, see Embrechts et al. (1997, Thms 3.3.26 and 3.3.27)), they are hardly used in practice. Indeed, the Gumbel-domain consists of distributions whose tails can be very different, although they have the common feature that all of their moments are finite (more precisely,

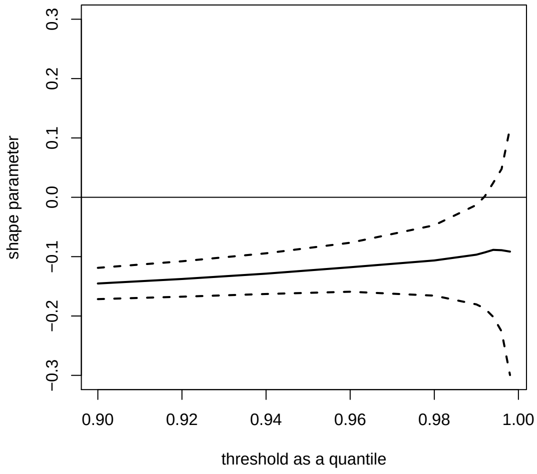


Figure 2.1: The mean and 95% confidence interval of the shape parameter of GPDs as a function of the threshold (chosen and displayed as a high quantile) for normally distributed samples of size 50000

$E\left(X^+\right)^m < \infty$ for all $m > 0$). To give some classification of the distributions belonging to this domain of attraction, let us introduce the notion of heavy-tailedness.

Definition 2.8. *Let $L_X(s) = E \exp(sX)$ denote the moment-generating function of a random variable X. X is heavy-tailed if $L_X(s) = \infty$ for all $s > 0$, and light-tailed if there exists an $s > 0$ such that $L_X(s) < \infty$.*

It is clear from above that all distributions in the $\xi > 0$ domain are heavy-tailed, while all distributions in the $\xi < 0$ domain are light-tailed. In contrast, it can be shown that within the Gumbel-domain $(\xi = 0)$ there are heavy-tailed distributions whose all moments are finite (e.g. the lognormal distribution), but also light-tailed distributions (e.g. the normal, the exponential or the gamma distribution, and even some distributions whose support is bounded to the right).

Hence the fact that a distribution belongs to the Gumbel-domain only gives a first approximation to its tail. Although it ensures the asymptotic exponentiality of scaled exceedances (Theorem 2.6), their behaviour above *finite* thresholds may differ very much from this asymptotics. As an illustration, Figure 2.1 shows the shape parameters of the GDPs fitted to exceedances above various thresholds (chosen as quantiles) for different normally distributed samples of size 50000. (Estimation of GPD parameters were carried out by maximum likelihood. According to the Figure, the mean of the estimated shape parameters is lower than the asymptotic value (zero) even for very high thresholds, which is not surprising because the normal distribution is lighter tailed than the exponential. The 95% confidence intervals on the Figure also illustrate the bias-variance trade-off mentioned previously.

An interesting family within the domain of attraction of the Gumbel-distribution, the class of distributions with Weibull-like tail, will be recalled throughout the dissertation. It is easy to show that such a distribution is heavy-tailed if $\alpha < 1$ and light-tailed if $\alpha \geq 1$. [4]

Definition 2.9. *The distribution of a random variable X has Weibull-like tail with exponent $\alpha > 0$ if there exist $K_1 > 0$, K_2 and $\lambda > 0$ such that, as $u \to \infty$,*

$$\bar{F}_X(u) \sim K_1 u^{K_2} \exp(-\lambda u^{\alpha}).$$

2.2.2 Extreme value theory for stationary sequences

When the i.i.d. assumption is relaxed and stationary time series are examined instead, a lot of new questions arise. The first type of questions is about the one-dimensional (marginal) distribution of the time series model: does the stationary distribution belong to the domain of attraction of a GEV, and if yes, to which one? If it belongs to the Gumbel domain, what is the more precise asymptotic behaviour of its tail? Answering these questions usually requires a thorough investigation of the data generating process.

The second type of questions is about the extremal dependence structure of the time series: how does the distribution of maxima change compared to the i.i.d. case, or what are the features of the clusters of high-level exceedances, which occur together? To make these questions more precise, let us first introduce the most basic concept of extremal dependence, the extremal index.

Extremal index and point process convergence

Let X_t be a stationary time series with continuous marginal distribution function F (and survival function $\bar{F}$) and let u_n denote a sequence of thresholds such that

$$n(1 - F(u_n)) \to \tau > 0, \tag{2.8}$$

which is equivalent to $F^n(u_n) \to \exp(-\tau)$. Here, as $F(u_n) \to 1$, u_n plays the role of a high threshold.

If X_t is i.i.d., it follows that $P(M_{1,n} \leq u_n) = F^n(u_n) \to \exp(-\tau)$ as $n \to \infty$. Instead of this relationship, for a wide range of stationary sequences there exists a θ real number

[4] Note that the naming may cause a slight confusion. In most fields of probability theory (e.g. in reliability), the Weibull-distribution is defined as having survival function $\bar{F}(u) = \exp(-\lambda u^{\alpha})$ for $u > 0$, hence the name "Weibull-like" in the definition. If X is a Weibull-distributed random variable in this sense and $\lambda\alpha^{\alpha} = 1$, then $-(X - \alpha)$ is a Weibull-distribution in the extreme value sense (see Definition 2.4) and $\alpha = -1/\xi$. Both wordings will be used in the dissertation, but it will not cause any ambiguities.

such that

$$P\left(M_{1,n} \leq u_n\right) \rightarrow \exp\left(-\theta\tau\right) \tag{2.9}$$

for each $\tau > 0$ and u_n satisfying (2.8). In this case θ is called the extremal index of the stationary series. Equations (2.8) and (2.9) together imply that

$$\theta = \lim_{n\to\infty} \frac{\log P\left(M_{1,n} \leq u_n\right)}{n \log F\left(u_n\right)}. \tag{2.10}$$

It follows easily from the definition that $\theta \in [0,1]$, thus a first intuitive meaning of θ (see (2.10)) is that the distribution of $M_{1,n}$ may be stochastically smaller because of the clustering of high values than it would be if the data were i.i.d. with the same marginal distribution function F. For a deeper meaning of the extremal index, one needs to introduce the point process of exceedances.

Define the point process of exceedances N_n on the state space $E = (0,1]$ as

$$N_n\left(.\right) = \sum_{t=1}^{n} \nu_{t/n}\left(.\right) \chi_{\{X_t > u_n\}}$$

where $\nu_x(.)$ is the Dirac-measure, i.e. for any $A \subset E$ Borel-set $\nu_x(A) = 1$ if $x \in A$ and $\nu_x(A) = 0$ otherwise. Thus N_n puts a unit mass to t/n if $X_t > u_n$. The importance of N_n comes from the fact that the most relevant indicators in extreme value theory can be deduced from it. For instance, taking the whole interval $(0,1]$, we obtain

$$N_n(0,1] = \text{card}\{t : 0 < t/n \leq 1 \text{ and } X_t > u_n\} = \text{card}\{t \leq n : X_t > u_n\}$$

hence the $N_n(0,1] = 0$ event is equivalent to $M_{1,n} \leq u_n$, which already appeared in the definition of the extremal index.

If the $\{X_t\}$ sequence is i.i.d., then N_n converges weakly in $M_p(E)$ (the state of all point measures on E equipped with an appropriate σ-algebra) to a homogenous Poisson-process N on $E = (0,1]$ with intensity τ (see Embrechts et al. (1997, Thm. 5.3.2)). Intuitively, this means that high-level exceedances occur independently (i.e. without clustering) in the i.i.d. case. But how does this limit result change if $\{X_t\}$ is a dependent process?

Obviously, we cannot expect a limit result to hold in general for N_n. For instance, if $X_t = X$ for a fixed random variable X for all t, the point process of exceedances will not converge to a reasonable limit. To obtain a meaningful theorem, we have to restrict our attention to processes with certain mixing properties. In the following we give four definitions of mixing, which will be useful further in the dissertation. The first concept, strong mixing, is widely used in all areas of probability, while the other three are rather motivated by extreme value theory.

Definition 2.10. *(strong mixing) Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{X_i : a \leq i \leq b\}$ and let*

$$\alpha_l = \sup\{|P(A \cap B) - P(A)P(B)| : -\infty < t < \infty, A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+l}^\infty\}.$$

X_t is strong mixing if $\alpha_l \to 0$ as $l \to \infty$.

Definition 2.11. *(Condition $\Delta(u_n)$) Let $\mathcal{H}_a^b(u_n)$ be the σ-algebra generated by the events $\{X_t \leq u_n : a \leq t \leq b\}$. For $1 \leq l \leq n-1$ write*

$$\beta_{n,l}^\Delta = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{H}_1^k, B \in \mathcal{H}_{k+l}^n, 1 \leq k \leq n\}.$$

Condition $\Delta(u_n)$ is said to hold if $\beta_{n,l_n}^\Delta \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

Definition 2.12. *(Condition $D(u_n)$) Let*

$$\begin{aligned}\beta_{n,l}^D = \sup\{|P(\max(X_t : t \in A_1 \cup A_2) \leq u_n) - \prod_{i=1}^2 P(\max(X_t : t \in A_i) \leq u_n)| :&\\ A_1 = \{i_1, \dots, i_p\}, A_2 = \{j_1, \dots, j_q\},&\\ 1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n, j_1 - i_p \geq l, p \in \mathbb{Z}, q \in \mathbb{Z}\}.&\end{aligned}$$

Condition $D(u_n)$ holds if $\beta_{n,l_n}^D \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

Definition 2.13. *Condition $D'(u_n)$ is said to hold if*

$$\lim_{k \to \infty} \limsup_{n \to \infty} n \sum_{i=2}^{[n/k]} P(X_1 > u_n, X_i > u_n) = 0.$$

It is clear that strong mixing implies condition $\Delta(u_n)$, and condition $\Delta(u_n)$ implies $D(u_n)$. The reasons for introducing them are the following theorems.

Theorem 2.14. *Let u_n be a sequence satisfying* (2.8)*, and assume conditions $D(u_n)$ and $D'(u_n)$ for X_t. Then the point process of exceedances, N_n converges weakly in $M_p(E)$ to a homogenous Poisson-process with intensity τ.*

Theorem 2.15. *(Hsing et al., 1988) Suppose that X_t has extremal index $\theta > 0$ and condition $\Delta(u_n)$ holds with a u_n satisfying* (2.8)*. Let k_n be a sequence of integers such that $k_n \to \infty$ and $k_n \beta_{n,k_n}^\Delta \to 0$ as $n \to \infty$. Let $r_n = [n/k_n]$ and define for $i > 0$ integers*

$$\pi_n(i) = P\left(\sum_{j=1}^{r_n} \chi_{\{X_j > u_n\}} = i | \sum_{j=1}^{r_n} \chi_{\{X_j > u_n\}} > 0\right).$$

Then, if $\pi_n(i)$ has a limit $\pi(i)$ for each $i = 1, 2, \ldots$, then π is a probability distribution and the point process of exceedances N_n converges weakly to a compound Poisson-process $\tilde{N}$ with intensity $\theta\tau$ and cluster size distribution π. That is,

$$\tilde{N} = \sum_{i=1}^{\infty} \xi_i \nu_{\Gamma_i}$$

where $\{\xi_i\}$ is a sequence of i.i.d. positive integer valued random variables with common distribution π, $\{\Gamma_i\}$ are the points of a homogenous Poisson-process N with intensity $\theta\tau$, and N is independent of all ξ_i.

The heuristic interpretation of Theorem 2.15 is that high-level exceedances of a stationary time series (satisfying some weak mixing conditions) occur in clusters. Under some additional summability conditions on $\pi_n(i)$:

$$\tau = \lim_{n\to\infty} n\bar{F}\left(u_n\right) = \lim_{n\to\infty} EN_n(0,1] = EN(0,1] = \theta\tau E\xi_1,$$

hence $\theta = \left(E\xi_1\right)^{-1}$, the reciprocal of the average cluster size. This interpretation makes the extremal index the most common measure of extremal clustering. A lower θ indicates a more pronounced clustering at extreme levels.

Theorem 2.14 has the important consequence that even some dependent sequences (which satisfy conditions $D(u_n)$ and $D'(u_n)$) behave in the same way at extreme levels as i.i.d. sequences do, and hence their extremal index is equal to one. For instance, for a stationary Gaussian sequence with autocorrelation function ρ_n, a sufficient condition for these conditions to hold is $\rho_n \log n \to 0$ as $n \to \infty$, which is indeed very weak. It implies that Gaussian ARMA models (and long range dependent FARIMA models as well) all have $\theta = 1$.

Based on this example, one could conjecture that very mild assumptions on the autocorrelation function ensure $\theta = 1$ for a linear model. However, this is not the case: θ is determined together by the innovation distribution and the autocorrelation function. For instance, if the assumption of normality is dropped, even an AR(1) model may produce nontrivial extremal clustering (i.e. $\theta < 1$). Let $X_t = aX_{t-1} + \varepsilon_t$ be a linear process with $0 < a < 1$. If ε_t belongs to the max-domain of attraction of the Frechet distribution with shape parameter $\xi > 0$ and a tail balance condition holds (i.e. $\bar{F}_{\varepsilon_t}(u) \sim p\bar{F}_{|\varepsilon_t|}(u)$ for a $p \in (0,1]$), then the extremal index of X_t is given by

$$\theta = p\left(1 - a^{1/\xi}\right) < 1.$$

On the other hand, if ε_t is light-tailed, and some general conditions on its probability density function are satisfied, the extremal index of X_t is equal to one, see e.g. Klüppelberg and Lindner (2005). (For a more detailed analysis of extremes of linear processes see Embrechts et al. (1997).) In the case of nonlinear processes, calculating the extremal index can be even more complicated.

How to estimate the extremal index? A classical procedure is the so-called blocks method, which makes use of representation (2.10) of θ. Let us divide our series of length n into k pieces of consecutive blocks of length $r = [n/k]$ and let K denote the number of blocks that contain at least one exceedance of u. Furthermore, let N be the total number of exceedances of u. Then $P\left(M_{1,r} \leq u\right) \approx 1 - K/k$ and $F\left(u\right) \approx 1 - N/n$, hence the blocks estimator of θ is given by

$$\hat{\theta}_n(u,r) = \frac{k\log(1-K/k)}{n\log(1-N/n)}.$$

As an approximation we get $\hat{\theta}_n \approx K/N$, which is just an approximation of the reciprocal of the average cluster size at extreme levels.

Another widely used method, the runs procedure, uses a different declustering scheme. Two exceedances of u are considered to belong to different clusters if there are at least r consecutive observations smaller than u between them. Then, denoting by K^* the number of exceedances where the following r observations lie below u (i.e. the number of rightmost members of the clusters), a natural estimator for θ is again the reciprocal of the average cluster size:

$$\hat{\theta}_n^*\left(u,r\right) = K^*/N. \tag{2.11}$$

If $u = u_n$ and r tends to infinity at an appropriate rate, both $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are consistent.

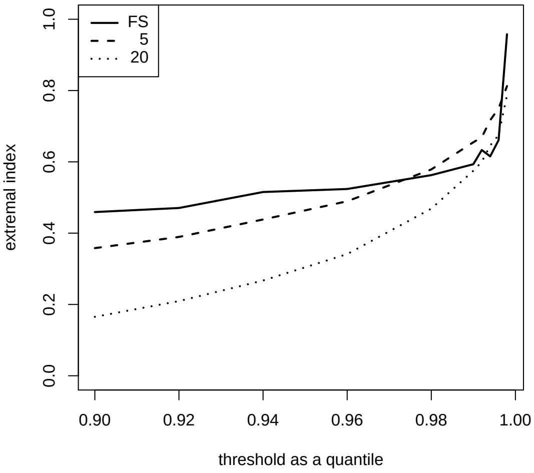
Both the blocks and the runs method have the drawback that they require to specify not only threshold u but also a cluster length r and $\hat{\theta}$ may depend very much on these particular choices. In an influential article, Ferro and Segers (2003) proposed a partial solution to this problem by analysing the distribution of the time between successive exceedances of u, denoted by $T(u)$. They prove under general mixing conditions that as $u \to x_F$,

$$\bar{F}\left(u\right)T\left(u\right) \to_d T_\theta,$$

where T_θ is the mixture of a degenerate distribution concentrated at zero and an exponential distribution with parameter θ, each having weights $1-\theta$ and θ, respectively. Thus the extremal index has a double role: in the limit it is both the proportion of non-zero interexceedance times (which is in line with the fact that the average size of an extremal cluster is $1/\theta$) and the reciprocal of the mean of the non-zero interexceedance times, properly normalised.

Based on this observation, Ferro and Segers (2003) give a moment-based estimator for θ, and prove its consistency at least for m-dependent processes. An automatic declustering

Figure 2.2: $\hat{\theta}$ as a function of the threshold (chosen and displayed as a high quantile) for a Gaussian AR(1) process (length 40000) with autoregressive coefficient 0.7. The runs method ($r = 5$ and $r = 20$) and the method of Ferro and Segers (FS) were used.



scheme is also proposed: given an initial choice of θ, the limit result suggests to consider the largest $\lfloor\hat{\theta}N\rfloor$ interexceedance times as intercluster times, i.e. as times that separate different clusters of high level exceedances from each other. (This is equivalent to choosing r as the $\lfloor\hat{\theta}N\rfloor$-th largest interexceedance time in the runs declustering scheme.) Then a final estimate for θ can be given e.g. the same way as in (2.11).

This way, we obtain an asymptotically motivated estimator of the extremal index, which is a function of only u, and is considered to be superior to the classical blocks or runs method. Although the problem of choosing r is eliminated, the estimation remains notoriously difficult because the behaviour of a process at finite u thresholds may differ substantially from the asymptotic behaviour measured by θ.

As an illustration, Figure 2.2 displays $\hat{\theta}$ as a function of the threshold (chosen as a quantile of the marginal distribution) estimated by the runs method (with $r = 5$ and $r = 20$) and with the method of Ferro and Segers for a Gaussian AR(1)-process with autoregressive coefficient 0.7. The estimates are substantially lower than the theoretical extremal index of the process (which is equal to one) even for high quantiles because dependencies die out very slowly as the threshold tends to ∞. This phenomenon is similar to what was observed in Figure 2.1, where the slow convergence of the shape parameter estimate of the GPD was presented.

Apart from illustrating the difficulties of estimating extremal characteristics, this simple example also teaches us that in many practical situations not only the asymptotic (extremal) behaviour of the process, but also the behaviour at finite (but high) thresholds is of interest. The asymptotic independence (as in the above AR(1) process) may be less important if the clustering is very pronounced at thresholds of practical interest.

Extremal cluster functionals

EVT for dependent observations traditionally focuses on the extremal index and (to a smaller extent) on the distribution of the size of an extremal cluster, π. However, in practical situations, other extremal quantities may be equally important. For instance, in the context of flood risk assessment, naturally arising quantities include not only the distribution and the mean of the duration of a flood (measured by π and $1/\theta$, respectively), but also the distribution of the flood peak (i.e. of the cluster maximum) or of the aggregate flood volume (i.e. of the aggregate excess above a high threshold).

Among these quantities, the distribution of the cluster maximum is quite easy to determine. Under general conditions (see e.g. Smith et al. (1997)) the cluster maximum has the same limit distribution as an arbitrary exceedance, hence it can be modelled as a GPD. For modelling other quantities (e.g. aggregate excesses), let us introduce the concept of extremal functional.

Definition 2.16. *For a process X_t and a threshold u, an extremal functional is defined by*

$$C_n(u) = \sum_{t=1}^{n-m+1} g\left(X_t - u, \ldots, X_{t+m-1} - u\right)$$

where g is a $\mathbb{R}^m \to \mathbb{R}_+$ function satisfying $g(\mathbf{x}) = 0$ for all $\mathbf{x} \notin \mathbb{R}_+^m$ (i.e. for vectors that have at least one negative component).

A few examples are the following. (Somewhat more general extremal functionals are considered in Segers (2003).)

Example 2.17. *The total number of exceedances above threshold u is obtained by choosing $m = 1$ and $g(x) = \chi_{\{x>0\}}$.*

Example 2.18. *The aggregate excess is defined by $m = 1$ and $g(x) = x^+$.*

Example 2.19. *Let $m > 1$, and for some $z > 0$ let*

$$g\left(x_t, \ldots, x_{t+m-1}\right) = 1 \quad \textit{if} \quad \min\left(x_t, \ldots, x_{t+m-1}\right) > z,$$

and $g\left(\mathbf{x}\right) = 0$ otherwise. Then $C_n(u)$ is the number of times that there are m consecutive exceedances of $u + z$.

It is not surprising in light of the point process convergence result (Theorem 2.15) that $C_n(u_n)$ has a compound Poisson limiting distribution when u goes to the upper end point of the support of the marginal distribution at a particular rate. Indeed, the following theorem, due to Smith et al. (1997), states that under some technical conditions, which most series of practical interest satisfy, the distribution of $C_n(u_n)$ (with u_n defined by (2.8)) converges as $n \to \infty$ to the distribution of a Poisson sum of i.i.d. variables. The essence is that each cluster of high-level exceedances of a stationary time series contributes independently to the determination of $C_n(u_n)$.

Theorem 2.20. *(Smith et al., 1997)*[5] *Let us assume that the X_t process is strong mixing with mixing function α_l. We can define a p_n sequence satisfying*

$$p_n \to \infty, \qquad \frac{p_n}{n} \to 0, \qquad \frac{n\alpha_{p_n}}{p_n} \to 0. \tag{2.12}$$

We also assume that, with p_n chosen this way and a u_n sequence defined in (2.8)*, the following conditions hold:*

$$\lim_{p\to\infty} \lim_{n\to\infty} \sum_{k=p}^{p_n} P(X_k > u_n | X_0 > u_n) = 0 \tag{2.13}$$

and

$$E\left(C_{p_n}(u_n) | M_{1,p_n} > u_n\right) \leq K < \infty. \tag{2.14}$$

Finally, we assume that there exists a C^* random variable such that

$$P(C^* \leq y) = \lim_{p\to\infty} \lim_{u\to\infty} P(C_p(u) \leq y | M_{1,p} > u). \tag{2.15}$$

Then the distribution of $C_n(u_n)$ converges as $n \to \infty$ to the distribution of $C_1^* + C_2^* + \cdots + C_L^*$ where L is a Poisson random variable and $C_1^*, C_2^*, \ldots$ are independent, of each other and of L, random variables with the same distribution as C^*. The mean of L is $\theta\tau$ where τ is defined by (2.8) *and θ is the extremal index. The latter can be calculated, for instance, as*

$$\theta = \lim_{p\to\infty} \theta_p \tag{2.16}$$

where

$$\theta_p = \lim_{u\to\infty} \theta\,(u, p) = \lim_{u\to\infty} P(M_{1,p} \leq u | X_0 > u). \tag{2.17}$$

Definition 2.21. *C^* in the theorem is called the extremal cluster functional. Its distribution is the limiting cluster size distribution in the case of Example 2.17 (then it is equal to π) and the limiting aggregate excess distribution in the case of Example 2.18.*

[5] Note that Smith et al. (1997) focused on Markov chains and thus stated the theorem assuming that X_t is a stationary aperiodic Harris chain. But only the strong mixing property of X_t was used in the proof.

How to estimate the extremal cluster functionals? In contrast to the GPD which describes threshold exceedances and cluster maxima, there do not exist unique parametric families to model the distributions of extremal cluster functionals. Moreover, as we illustrated, even the estimation of the extremal index (a single number) may pose great difficulties, hence the estimation of the cluster functionals is almost impossible in practice without additional assumptions on the data generating process (e.g. Markovity or a certain time series structure). This will be a recurrent topic in the dissertation.

2.2.3 Extreme value theory for Markov chains

To give an example how assuming a certain structure on the time series may help estimate the extremal index and the extremal cluster functionals, let us consider a discrete time Markov chain X_t with continuous state space. Assume that its stationary distribution has asymptotically a unit exponential tail: $\bar{F}_X(u) \sim K \exp(-u)$. There is no loss of generality in this since we are interested in the extremal clustering of the process and the general case can be derived by a marginal transformation. Concerning the bivariate dependence structure, a natural assumption, which is satisfied by all practically relevant bivariate distributions, is that the joint distribution of (X_{t-1}, X_t) belongs to the domain of attraction of a bivariate extreme value law (see Coles (2001) and Coles and Tawn (1991)). A sufficient condition for this assumption is given below, adapted from Resnick (1987, Prop. 5. 15.) to the case of exponential marginals.

Let (Y_1, Y_2) be a bivariate random variable with distribution function $G(y_1, y_2)$ and marginals $G_1(y_1)$ and $G_2(y_2)$, respectively. Using the notation $Z_j(y_j) = -\log(G_j(y_j))$, let $Y_j^* = Z_j(Y_j)$ $(j = 1, 2)$, then (Y_1^*, Y_2^*) has unit exponential marginals and bivariate distribution function

$$G_*(s_1, s_2) = G\left(Z_1^{-1}(s_1), Z_2^{-1}(s_1)\right).$$

(Y_1, Y_2) (or, alternatively, G) belongs to the domain of attraction of a bivariate extreme value distribution if for all s_1 and s_2

$$\lim_{u \to \infty} \frac{1 - G_*(u + s_1, u + s_2)}{1 - G_*(u, u)} = \frac{V(e^{s_1}, e^{s_2})}{V(1, 1)} \tag{2.18}$$

with

$$V(v_1, v_2) = \int_0^1 \max\left(\frac{w}{v_1}, \frac{1 - w}{v_2}\right) dH(w), \tag{2.19}$$

where H is a nonnegative measure on $[0, 1]$ (see e.g. Coles and Tawn (1991)), satisfying

$$\int_0^1 w dH(w) = \int_0^1 (1 - w) dH(w) = 1. \tag{2.20}$$

H is called the spectral measure of the bivariate extreme value distribution. It completely determines the extremal dependence of the two univariate variables. For instance, if Y_1 and Y_2 are (asymptotically) independent H puts all its mass equally to 0 and 1, and if there is a monotone increasing deterministic relationship between Y_1 and Y_2, H puts all the mass to 1/2.

Taking (2.18) as an identity for large $u_i = u + s_i$ $(i = 1, 2)$ we can obtain by using the fact that V is homogeneous of order -1 (c.f. Smith et al. (1997)):

$$1 - G_* (u_1, u_2) = \frac{e^u (1 - G_* (u, u))}{V(1,1)} V(e^{u_1}, e^{u_2}) = K(u) V(e^{u_1}, e^{u_2}) . \tag{2.21}$$

$K(u)$ is determined by the marginal distributions. If we set $u_1 = \infty$ and utilise the identity $V(\infty, x) = x^{-1}$ (see (2.19) and (2.20)) and the exponential marginals of G_*, we get from (2.21) that $e^{-u_2} = K(u)e^{-u_2}$ and hence $K(u) = 1$.

Then, if (2.21) exactly holds, using the fact that V_1, the partial derivative of V, is homogeneous of order -2, we can obtain (see Bortot and Coles (2000))

$$\begin{aligned} F^*(z) :&= \lim_{u\to\infty} P(Y_2^* < u + z | Y_1^* = u) = \lim_{u\to\infty} \left[\left(e^{-u}\right)^{-1} \frac{\partial G_*(x,y)}{\partial x} |_{(x,y)=(u,u+z)} \right] \\ &= \lim_{u\to\infty} \left[- \left(e^{-u}\right)^{-1} e^u V_1 \left(e^u, e^{u+z}\right) \right] \\ &= \lim_{u\to\infty} \left[-e^{2u} e^{-2u} V_1(1, e^z) \right] = -V_1(1, e^z) . \end{aligned} \tag{2.22}$$

Thus, provided that the bivariate dependence structure exactly follows an extreme value law above a sufficiently high threshold and Y_1^* is large enough, the distribution of $Y_2^* - Y_1^*$ does not depend on Y_1^*. If the bivariate extreme value law is only an approximation, then the above property holds only under some regularity conditions (i.e. if taking the derivative and taking the limit can be interchanged).

Thus on the basis of these considerations we can expect that for X_t Markov chains with asymptotically exponential marginals that are important from a statistical modelling perspective there exists a limiting step distribution function

$$F^*(z) = \lim_{u\to\infty} P(X_t < u + z | X_{t-1} = u) . \tag{2.23}$$

Note that $F^*(z)$ – even if it exists – is not always a proper distribution function. For instance, if X_{t-1} and X_t are (asymptotically) independent, then $F^*(z) = 1$ for all z.

When deriving the extremal properties of Markov chains the literature in extreme value theory generally assumes (2.23) or a version of it rather than making the conditions behind it (which we outlined above) explicit. A typical article of this kind proves under further conditions that the chain behaves at extreme levels as a random walk for many steps ahead, sufficiently for the extremal index or extremal cluster functionals to converge to their random walk versions. For instance, Smith (1992) proves the following proposition.

Proposition 2.22. *(Smith, 1992) Assume that* (2.13) *holds and the transition density* $q(x, y)$ *of the Markov chain satisfies* $\lim_{u\to\infty} q(u, u+z) = h(z)$ *for some limiting function* $h(z) \geq 0$, $\int_{-\infty}^{\infty} h(z)dz \leq 1$. *Moreover, suppose that there exists a* u^* *such that, for all* M, $q(u, u+z)$ *is uniformly bounded over* $u \geq u^*$, $y^* \geq M$ *and*

$$\lim_{M\to\infty} \lim_{u\to\infty} \sup_{x\leq u-M} P\left(X_t > u | X_{t-1} = x\right) = 0. \tag{2.24}$$

Under these assumptions the limit in (2.23) *exists, and the extremal index of the process is given by*

$$\theta = \int_{-\infty}^{0} \exp\left(x\right) Q\left(x\right) dx,$$

where $Q(x)$ *is the solution of the Wiener-Hopf equation*

$$Q\left(x\right) = \int_{0}^{\infty} Q\left(y\right) F^{*}(x - dy).$$

Note that (2.24) ensures that the Markov chain jumps from a very low level to a high level in one step only with vanishing probability. The appearence of the Wiener-Hopf equation in the calculation of θ is not a surprise. If X_t has a unit exponential marginal and happens to follow exactly a random walk above a certain level, then (2.16)-(2.17) yield under some regularity conditions that, using the notation $M_{k,\infty} = \max\{X_i : i \geq k\}$,

$$\begin{aligned}\theta = P\left(M_{1,\infty} \leq u | X_0 > u\right) &= \int_{-\infty}^{0} P\left(M_{1,\infty} \leq u | X_0 = u - x\right) \exp\left(x\right) dx \\ &= \int_{-\infty}^{0} Q\left(x\right) \exp\left(x\right) dx,\end{aligned}$$

where $Q(x) := P\left(M_{1,\infty} \leq u | X_0 = u - x\right)$ satisfies

$$Q\left(x\right) = \int_{0}^{\infty} P\left(M_{2,\infty} \leq u | X_1 = u - y\right) F^{*}\left(x - dy\right) = \int_{0}^{\infty} Q\left(y\right) F^{*}\left(x - dy\right).$$

Proposition 2.22 for the extremal index was later generalised to the determination of extremal cluster functionals by Perfekt (1994). Based on these findings and generalising them, Smith et al. (1997) developed an estimation and simulation scheme for the extremal cluster functionals of Markov chains with exponential tail. In the spirit of the threshold methods of extreme value theory (e.g. of the POT procedure described above), they assume that the values of the analysed process are censored. That is, instead of X_t, only $W_t =$

(V_t, δ_t) is observed, where $V_t = \max(X_t, u)$ and $\delta_t = \chi_{\{X_t > u\}}$ with a sufficiently high u threshold.

As a result of Markovity, the joint density for $\{X_t\}$ factorises into

$$f(X_1, X_2, \ldots, X_n) = f(X_1) \prod_{t=2}^{n} \frac{f(X_{t-1}, X_t)}{f(X_{t-1})}.$$

Taking into account censoring, an approximate likelihood can be obtained for $\{W_t\}$ in a straightforward way. In view of Theorem 2.6 the $f(X_{t-1})$ term is replaced by the density of the GPD if $X_{t-1} > u$ and by $F_X(u)$ otherwise. The $f(X_{t-1}, X_t)$ term can be modelled with a parameteric bivariate extreme value distribution (i.e. assuming a parametric structure on V or H in (2.19)) by using approximation (2.21)[6] if $\min(X_{t-1}, X_t) > u$ and by appropriate modifications otherwise. Then, the approximate likelihood can be maximised to yield the parameters of the GPD and of the bivariate extreme value law governing the transition of the Markov chain at extreme levels.

In the second step of the method proposed by Smith et al. (1997), after transformation of the marginals into exponential tail, the limiting step distribution function of the Markov chain is calculated from (2.22) and then the extremal cluster functionals are simulated on the basis of this limiting random walk representation. With this procedure, the accuracy of extremal estimation and simulation is improved substantially compared to the case when no prior knowledge is given about the structure of dependence and hence when only nonparametric estimation of extremal functionals is possible. The method has become a popular tool in EVT of time series, and gave rise to various modifications and generalisations such as in Bortot and Tawn (1998) or Sisson and Coles (2003). For a practical application, see Fawcett and Walshaw (2006).

[6] Here the parameters of the marginal generalised Pareto distribution also appear indirectly through Z_1 and Z_2.

Chapter 3

Motivation: empirical features of water discharge data

In this chapter we present the most important time series and extreme value features of water discharge series of rivers Danube and Tisza. These "stylised facts" help us restrict the classes of time series processes that can arise as possible candidates for modelling river flow series. The final aim (achieved in Chapters 4 and 5) will be to combine the knowledge on the time-dependence structure of the series with extreme value theory to understand the extremes of river flows better than it would be possible with the techniques of extreme value theory or time series analysis alone. The motivation behind doing such a research comes from the fact that both rivers have a long history of damaging inundations and record-high floods indeed occured repeatedly in the last few years as well, causing losses of property worth hundreds of billion HUFs.

3.1 Time series properties

The data examined in this dissertation consist of daily water discharge measurements at three monitoring stations along river Danube (Komárom, Nagymaros, Budapest) and also three stations at river Tisza (Tivadar, Vásárosnamény, Záhony).[1] All series finish in year 2000 but they start in different years. The starting point is 1901 for Nagymaros and Vásárosnamény, 1925 for Budapest, 1951 for Tivadar, 1953 for Záhony and 1961 for

[1] Although water level data are also available we focus on water discharge measurements. Level data tend to be more unstable because of the changing shape of the river basin, hence it is a common practice to use discharge data in the hydrological literature in spite of their slightly smaller practical value. Discharge data are calculated from level data by appropriate hydrological transformations, and inverse transformations also exist.

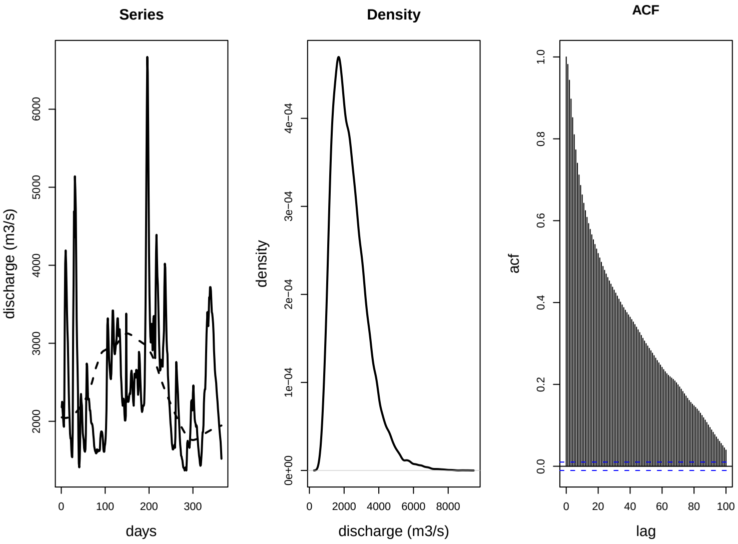


Figure 3.1: A one year portion (along with the estimated seasonal component), the probability density and the autocorrelation function of the water discharge series at Nagymaros (river Danube, Hungary)

Komárom. Thus each data set consists of at least 14600 observations, and some as many as 36500 days.

To obtain a first impression of the data, Figure 3.1/a displays a one-year portion of the daily water discharge series at Nagymaros. Similarly to the other five series, a distinctive feature of this data set is its pulsatile nature, i.e. that apparently random short and steep rising periods are followed by longer, gradually falling ones. As Figure 3.1/b shows, the series has a highly skewed marginal distribution.

All series exhibit substantial seasonality both in mean and in other features of their distribution, and there is also a small upward trend in all series. Although there have been important water resource developments at the Hungarian segments of Danube and Tisza during the examined period (e.g. the construction of the Tiszalök and Kisköre hydroelectric power plants), no practically important structural break can be detected in the series. The reason is that the examined monitoring stations lie in segments relatively unaffected by such dammings.

As a starting point, the seasonal and trend component (c_t) in the mean of each series was estimated using a local polynomial fitting (LOESS) procedure as proposed by Cleveland et al. (1990). (Figure 3.1/a also displays the estimated seasonal component at Nagymaros.) The remaining $X_t - c_t$ series is stationary in mean but still has a seasonal component e.g. in its variance, which is an issue to be solved during modelling.

3.1.1 ARMA modelling

The deseasonalised $X_t - c_t$ series have zero mean and are strongly autocorrelated at all sites: the one-lag autocorrelations are e.g. around 0.95. (See the autocorrelation function at site Nagymaros in Figure 3.1/c.) As a first approach to tackle the dependence structure, ARMA(p,q) models were fitted to $X_t - c_t$ with various p and q orders.

As already mentioned in section 2.1, the standard way to examine the adequacy of an ARMA model is to check the autocorrelation function of the fitted innovations. At first sight ARMA(3,1) or ARMA(2,1) models (depending on the site) seem appropriate at the monitoring stations in the sense that there remains little autocorrelation in the $\{\hat{\varepsilon}_t\}$ sequences. However, a formal evaluation of the goodness of fit is not an easy task because $\{\hat{\varepsilon}_t\}$ – although roughly uncorrelated – is far from an independent series: there is a strong autocorrelation in its square and also in the absolute values of the fitted innovations at each station. (Figure 3.2 displays the autocorrelation function of the fitted innovations, of the squared and absolute valued innovations at Nagymaros.) Since the p-values derived from the standard Box-Pierce or Ljung-Box tests (equation (2.5)) are valid only if the ε_t innovations are independent (i.e. if $X_t - c_t$ follows a strong ARMA process), they are not applicable in this case. In fact, if the ARMA representation is only a semi-strong or weak one, the asymptotic distribution of the Box-Pierce and Ljung-Box statistics and thus the critical values depend substantially on the true data generating process (Francq et al., 2005).

Therefore a posterior evaluation strategy was used to justify the adequateness of the choice $p = 3$ (or $p = 2$), $q = 1$. After estimating an ARMA-ARCH-type nonlinear model (which fits well to the observations, see section 4.6), we simulated synthetic ARMA-ARCH-type series with the estimated parameters, fitted the ARMA models to them and obtained empirical critical values of the Box-Pierce statistics of the resulting innovations by simulation. The fit was accepted at all sites at the 99% level for not too large lags in the autocorrelations.[2] Note also that model selection based on the standard Box-Pierce p-values (i.e. assuming that the process is a strong ARMA) would have rejected the chosen ARMA models and thus would have resulted in overfitting. For instance, at site Nagymaros the simulated 99% critical value of the test statistic of the first $r = 8$ autocorrelations of the innovations was 22.1, while the observed value was 21.6. In comparison, the 99% quantile

[2] The use of the 99% level may be justified by the large sample sizes (15000-36000 observations). With the somewhat vag ue concept of "practical" (as opposed to "statistical") significance, one can surely say that the deviation from the chosen weak ARMA representations – even if it exists – is not practically significant for not too large lags.

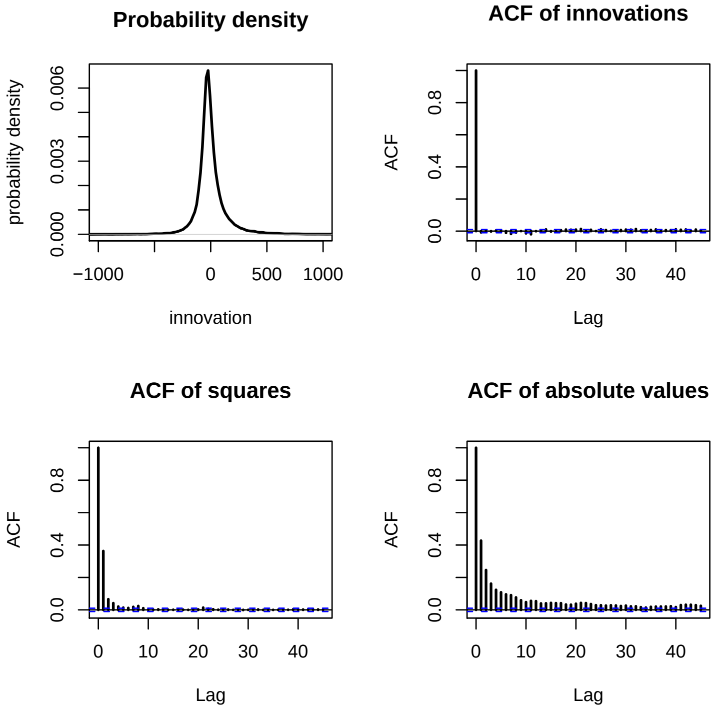


Figure 3.2: Probability density of the innovations, autocorrelation function of the innovations, of their squares and of their absolute values at Nagymaros

of a χ^2 distribution with $r - p - q = 4$ degrees of freedom is 13.3.[3] It is worth noting that more restricted models (e.g. ARMA(1,1)) were not appropriate at any site.

According to the Box-Pierce test there remain some autocorrelations at *high* lags at all sites but this problem cannot be overcome by increasing the ARMA model order. This fact is related to the otherwise detectable long range dependence of the series (see section 3.1.2).

Therefore, modelling the water discharge data by a strong ARMA process has the shortcoming that it neglects both the nonlinearity (as evidenced by the autocorrelation in the squared innovations) and long range dependence present in the data. However, from the natural hazards (flood) perspective, the main question is whether this simplification affects the model's performance in terms of approximating the probability distribution and high quantiles of the observed discharge series. To examine that, we need to simulate water discharge series from the ARMA-model.

[3] Naturally, other nonlinear structures such as the regime switching autoregressive models of Chapter 5 could also be used to simulate the true data generating process. The p-values derived this way would slightly differ from those in the ARMA-ARCH-case, but would still be higher than in the strong ARMA-case.

If the independence of innovations is assumed (i.e. the model is strong ARMA), the simulation can be carried out in a straightforward way even when the fitted innovations are strongly non-Gaussian. Non-Gaussianity is the case here as well: the probability density of $\hat{\varepsilon}_t$ is highly peaked and somewhat skewed, see the case of Nagymaros in Figure 3.2. In order to concentrate on the time-dependence structure, we can use a seasonal bootstrap procedure to generate the innovation sequence: that is, a synthetic innovation in month A is randomly selected from all observed innovations in the same month but in a possibly different year. This method is commonly used in hydrology (see e.g. Montanari et al. (1997)) and has the advantage of not making any artificial distributional assumptions, moreover, it takes into account the observed seasonality of the series even after having removed the seasonal component in the mean (c_t). However, it may be sensitive to a few extreme observations, therefore the financial econometric literature (e.g. McNeil and Frey (2000)) prefers to use its slight modification where only the central part of the distribution is generated by bootstrap, while the upper and lower tail (e.g. the upper and lower 5%) are simulated by fitting a GPD to them. We pursued both the full bootstrap and the mixed method but did not find a substantial difference between them, therefore we present only the results with the simpler (full bootstrap) procedure here.

After simulating the independent innovation sequence, the synthetic water discharge series is generated by applying the linear (ARMA) filter to the innovations and finally adding back the seasonal and trend component c_t. This way, the autocorrelation structure of the series is properly reproduced. However, the probability densities of the simulated and the observed series do not fit well and high quantiles are seriously underestimated by the simulations especially for sites at river Tisza. (This is illustrated in Figure 3.3 for two selected monitoring stations: Nagymaros at river Danube and Tivadar at river Tisza.) Thus the use of simple ARMA processes – although often applied in practice – may be very misleading in flood risk assessment.

3.1.2 A note on fractional ARIMA modelling

Before turning to the nonlinear models one should also consider whether long range dependent (LRD) processes may help improve the density and quantile forecasts. The river flow series exhibit LRD patterns: the slow decay of the autocorrelation function is already displayed in Figure 3.1/c, and other nonparametric procedures (e.g. aggregate variance method, R/S statistics) also point to the presence of long memory (Elek, 2002). This is not a particularly surprising fact: the detection of long range dependence in certain hydrologic time series dates back to the early works of Hurst (1951) and since then a plenty of articles

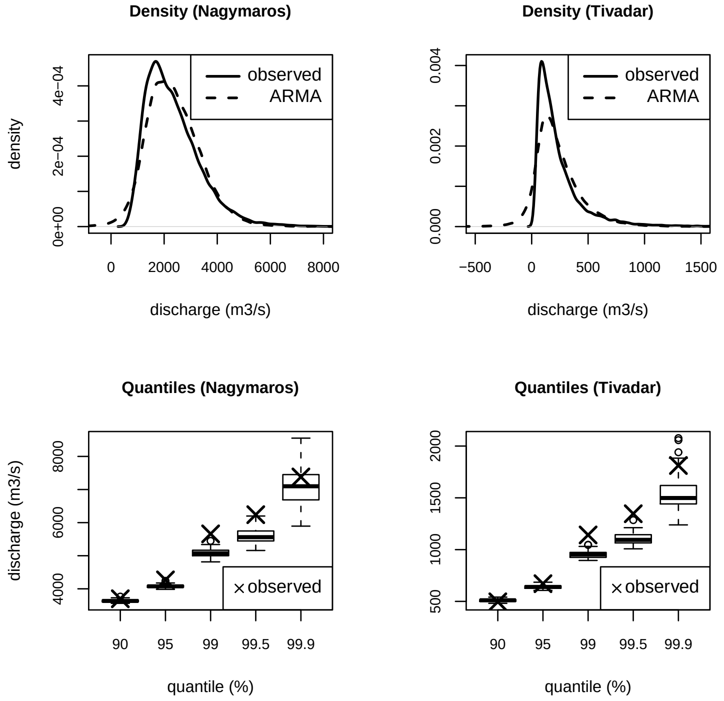


Figure 3.3: Probability density and high quantiles of empirical (observed) and ARMA-simulated series at two selected monitoring stations (Nagymaros and Tivadar). The box-plots of simulated quantiles are constructed from 100 simulated series of the same length as the original ones. The boxes show the middle 50% of the distribution.

have dealt with this phenomenon in river flow series. (For recent examples see Montanari et al. (1997) or Ooms and Franses (2001).) However, the LRD properties of rivers Danube and Tisza had not been investigated before.

In Elek and Márkus (2004) we fitted fractional ARIMA processes to the water discharge series of six monitoring stations at river Tisza by the Whittle estimation procedure and obtained that the estimated Hurst parameter lies in the range of 0.75-0.82 and turns out to be significantly greater than 1/2 at all stations. The introduction of the fractional differentiation filter $(1-B)^d$ on top of the ARMA one completely eliminates all high-lag autocorrelations in the innovations of the ARMA fit. At the same time, however, simulations from the fitted fractional ARIMA model (performed in a similar way than in the ARMA case) leave the obtained probability densities and high quantiles practically the same as in the simpler ARMA case. This is illustrated in Figure 3.4 where the simulation results of the ARMA and FARIMA models are compared for Tivadar, and only slight differences occur.

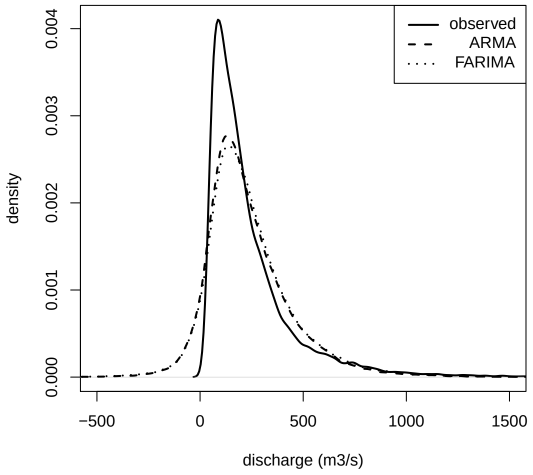


Figure 3.4: Comparison of the fit of the ARMA and FARIMA models in terms of approximating the probability density and high quantiles at Tivadar. The two lines are almost indistinguishable.

Therefore it can be concluded that even a long range dependent linear model with independent non-Gaussian innovations does not approximate well the probability density and high quantiles of the discharge series. This result differs from the findings of Montanari et al. (1997), where a fractional ARIMA model with independent, seasonally changing non-Gaussian innovations is used to simulate daily inflows to Lake Maggiore in Italy and the fit of the probability density is adequate. Thus using nonlinear models in the cases of Danube and Tisza is essential because the fitting criteria of the linear processes do not guarantee and in fact do not produce the fit of the tails.

3.2 Extreme value properties

The class of nonlinear processes which arise as possible candidates for river flow modelling is very wide. In this section we restrict this class by examining a few important extreme value properties of the observed series. These properties should hold for reasonable models of water discharges of Danube and Tisza.

Table 3.1 displays the shape parameter estimates of the GPD-s (with the corresponding standard errors) fitted to exceedances over two different thresholds, roughly the 90% and 95% quantiles, at each site.[4] It turns out that the point estimates are close to zero and are generally negative, being only the estimate at Záhony significantly lower than zero.[5] This

[4] Since the observations are not i.i.d., a separation method should be used here as well to identify approximately independent exceedances. The runs method with run length of 15 days was used for this purpose.

[5] Standard errors and thus p-values were determined from the Fisher information matrix, using asymptotic likelihood theory. More precise (asymmetric) confidence intervals could have been obtained with the profile likelihood method.

gives a strong indication against time series with positive shape parameter $(\xi > 0)$. (For more details on extreme value analysis of floods at River Tisza, see Bozsó et al., 2005.)

It is worth noting that these findings are supported by other international hydrological studies, too. A comprehensive analysis conducted in the mid-eightees by the World Meteorological Organization indicated that the Gumbel distribution is one of the most frequently applicable tool for annual maxima of discharge series (Cunnane, 1989). (It follows from Theorem 2.6 that if a Gumbel distribution fits well to the annual maxima, then the exceedances over high thresholds can be approximated by the GPD with $\xi = 0$.) Out of the 54 hydrological agencies (mostly in Europe and North America) 28 used this distribution as the recommended or "standard" one for medium and large rivers. (The Frechet distribution, i.e. $\xi > 0$, is usually advised for smaller catchments.)

Table 3.1: Shape parameter estimates (with standard errors in parentheses) of the GPDs fitted to exceedances over the specified thresholds at different monitoring stations

Monitoring station	Threshold	ξ	Threshold	ξ
	$(\mathrm{m}^3/\mathrm{s})$		$(\mathrm{m}^3/\mathrm{s})$	
Komárom	3400	-0.022 (0.077)	4000	0.055 (0.111)
Nagymaros	3700	-0.059 (0.121)	4300	-0.095 (0.060)
Budapest	3600	-0.034 (0.123)	4200	-0.116 (0.106)
Tivadar	500	-0.048 (0.062)	700	0.032 (0.090)
Vásárosnamény	800	-0.085 (0.075)	1100	-0.131 (0.112)
Záhony	850	-0.230 (0.048)	1150	-0.214 (0.055)

Turning to the extremal clustering properties, all series display substantial clustering even above high thresholds. This is illustrated in Figure 3.5 where the extremal index estimated with the method of Ferro and Segers (2003) is shown as a function of the quantile of the marginal distribution for two sites, and the estimates turn out to be lower than 1/2 even at the 99.5% quantile. (In order to eliminate the potential effect of the seasonality, the extremal index was also estimated for the deseasonalised series at Tivadar but the results did not differ substantially.) Thus, a suitable model of river flows should reflect this feature of clustering at high level. However, it does not necessarily follow that the theoretical extremal index of such a model should be less than one: as Figure 2.2 illustrated even an asymptotically non-clustering series can exhibit strong clustering at finite (but high) thresholds.

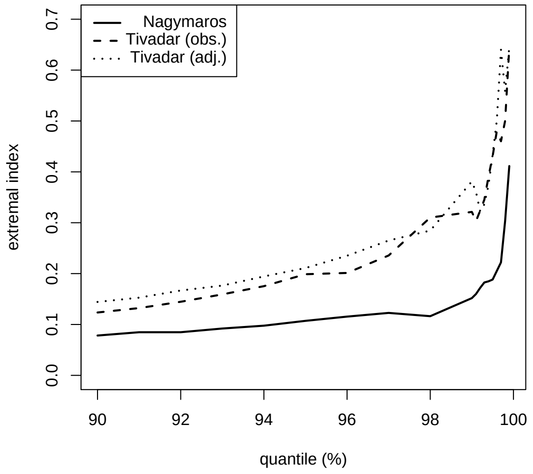


Figure 3.5: Extremal index as a function of the threshold chosen as a quantile for Nagymaros and for the observed and seasonally adjusted series at Tivadar. The index is estimated with the method of Ferro and Segers.

3.3 Summary

In this chapter we outlined those statistical properties of the water discharge series that are most important from a modelling point of view. The nonlinearity of the data sets was illustrated by the inadequate fit of ARMA (and also fractional ARIMA) models, even when generated by non-Gaussian innovations. As far as extremes are concerned, the series possess tails in the max-domain of attraction of the Gumbel distribution, rather than polynomially decaying tails, and they exhibit remarkable clustering of high values at least at not too high levels.

Therefore, in order to model successfully the time series and extreme value properties of the water discharge series, one should use nonlinear processes that are not "too" heavy tailed, i.e. that the shape parameter of their corresponding GPD is zero. In the following two chapters we present two classes of such models. The first, a conditionally heteroscedastic process, models the remaining (nonlinear) structure of the innovations of an ARMA model fitted to the series. Therefore, this approach can be regarded as a rather statistical solution to the nonlinearity problem. To the contrary, the second model, a regime switching autoregressive one, directly captures the "pulsatile" nature of the river flow data and therefore is a rather "structural" (or "physical") approach to modelling. Besides proving the usefulness of both models in hydrological analysis, we raise and answer interesting mathematical questions about their stationarity, tail behaviour, extremal clustering properties and estimation of their parameters.

Chapter 4

Conditionally heteroscedastic models

4.1 The model

In this chapter we introduce a specific class of conditionally heteroscedastic models and investigate its stationarity, tail behaviour, extremes, estimation of its parameters and its applicability in river flow analysis.

An ARCH(r)-type model (ARCH stands for autoregressive conditional heteroscedasticity) is defined by the equation

$$\varepsilon_t = \sigma\left(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-r}\right) \cdot Z_t \tag{4.1}$$

where Z_t is an i.i.d. sequence with zero mean and unit variance, and $\sigma(\mathbf{x})$ is an appropriate $\mathbb{R}^r \rightarrow \mathbb{R}^+$ nonconstant function. Then, if the model has a stationary solution, $E\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0$ where $\mathcal{F}_t$ is the $\sigma-$algebra generated by $\{X_i : i \leq t\}$. That is, ε_t is a martingale difference sequence (hence it is uncorrelated provided that it has finite variance). Nevertheless, it is not independent because $E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) = \sigma^2\left(\varepsilon_{t-1}, \ldots, \varepsilon_{t-r}\right)$. Therefore, ARCH-type processes provide a natural (and statistically inspired) way to introduce nonlinearity by allowing to model the conditional variance structure of an already uncorrelated process.

Since volatility clustering is one of the "stylised facts" of financial time series (e.g. of stock returns), ARCH-type processes have become the standard tools in financial analysis in the last two decades. Restricting the attention to ARCH(1)-type models, the original specification, due to Engle (1982), gives the conditional variance as a quadratic function of the previous value:

$$\sigma^2(x) = \alpha_0 + \alpha_1 x^2. \tag{4.2}$$

(This model will be referred to as the standard / quadratic specification.) Numerous generalisations have been published since, which are successfully applied in financial econometrics. For instance, the TARCH(1)-model (Glosten et al., 1993) is defined as

$$\sigma^2(x) = \alpha_0 + \alpha_{1+}\left(x^+\right)^2 + \alpha_{1-}\left(x^-\right)^2 \tag{4.3}$$

where $\alpha_0 > 0$, $\alpha_{1+} \geq 0$ and $\alpha_{1-} \geq 0$. This model is motivated by the fact that negative shocks to the stock market tend to have larger impact on the variance than positive shocks, hence $\alpha_{1+} > \alpha_{1-}$ generally holds for stock return series. Further generalisations (the so-called GARCH-type models) arise when the conditional variance is allowed to depend on all lags of ε_t.[1]

A typical ARMA-ARCH-type model is then defined by combining a version of equation (4.1) with the ARMA-equation (2.3), see e.g. Hamilton (1994). With this choice the latter equation drives the conditional expectation of the process, while the former determines the conditional variance.

Probabilistic and statistical properties of the above, typically used ARCH- and ARMA-ARCH-type models have been researched thouroughly, and a number of articles and monographs has been published about them (some of which will be cited later in this dissertation). However, they are not appropriate without modifications for river flow modelling for at least two reasons.

First, as equations (4.2) or (4.3) show, the typical models have the common feature that their conditional variance is asymptotically a quadratic function of the past observations. This property yields that the stationary distribution (if it exists) has a polynomially decaying tail (see e.g. Borkovec and Klüppelberg (2001)), which contradicts the empirical findings on river flow series. Hence, to get the tail behaviour right, the conditional variance of a suitable model should increase slower than a quadratic function, i.e. assuming only one lag, $\lim_{|x|\to\infty} \sigma^2(x)/x^2 = 0$.

Second, the usual definition of the ARMA-ARCH process (that is, the simple combination of (2.3) and (4.1)) is not easily interpreted in hydrological terms since in that setting the conditional variance of ε_t depends on ε_{t-1} (the previous innovation). Instead, conditioning should be made directly on the past value of X_t itself to capture a proper feedback from the modelled process. High river discharge, as a rule, goes together with a more saturated watershed, allowing any further precipitation a more straightforward reach to the river and thus leading to a greater possible increase in the water supply. On the other

[1] More precisely, ε_t is a GARCH(r, s)-type model if $\varepsilon_t = \sigma_t Z_t$ where

$$\sigma_t^2 = f\left(\varepsilon_{t-1}, \ldots, \varepsilon_{t-r}, \sigma_{t-1}^2, \ldots, \sigma_{t-s}^2\right)$$

with an appropriate function f. In the original GARCH(1,1) model (Bollerslev, 1986) $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$. A GARCH model can be written in the form of an ARCH(∞) model.

hand, saturated watershed gives away water quicker, producing greater possible decrease of the series. As a result, higher water discharge indicates higher uncertainty for the next day's discharge, corresponding to the mentioned feedback effect. Though models where the conditional variance of ε_t is defined directly as a function of the past values of X_t are known (Borkovec and Klüppelberg, 2001; Masry and Tjostheim, 1995), they are much less elaborated in the literature than the usual specification.

Based on the above considerations, we introduce and examine the following ARMA-ARCH-type model in the dissertation:

$$X_t = c_t + \sum_{i=1}^{p} a_i (X_{t-i} - c_{t-i}) + \sum_{i=1}^{q} b_i \varepsilon_{t-i} \tag{4.4}$$

$$\varepsilon_t = \sigma(X_{t-1}) Z_t = \sigma_t Z_t, \tag{4.5}$$

where

$$\sigma^2(x) = \alpha_0 + \alpha_{1+} \left((x-m)^+ \right)^{2\beta} + \alpha_{1-} \left((x-m)^- \right)^{2\beta}. \tag{4.6}$$

(Hence the notation $\sigma_t = \sigma(X_{t-1})$ will also be used.) Here, c_t is a deterministic periodic function (with period of e.g. one year), representing the seasonal component of the mean, and the following assumptions are imposed on the parameters and on the noise sequence.

Assumption 4.1. *Z_t is an independent identically distributed random sequence (or more generally, an independent sequence with seasonally changing distribution) with zero mean and unit variance. The distribution of Z_t is absolutely continuous with respect to the Lebesgue-measure, and its support is the whole real line. (In contrast to the term innovation, Z_t will be referred to as the noise in the model.)*

Assumption 4.2. *For the characteristic polynomials*

$$\Phi(z) = 1 - \sum_{i=1}^{p} a_i z^i \neq 0 \quad \textit{and} \quad \Psi(z) = 1 + \sum_{i=1}^{q} b_i z^i \neq 0 \quad \textit{if} \quad |z| \leq 1,$$

and $\Phi(z)$ and $\Psi(z)$ have no common zeros.

Assumption 4.3. $0 < \beta < 1$.

Assumption 4.4. $\alpha_0 > 0$, $\alpha_{1+} \geq 0$ *and* $\alpha_{1-} \geq 0$.

Putting (4.4)–(4.6) into words the process X_t is generated from the ε_t innovation sequence like an ARMA process with seasonally changing mean. Assumption 4.2 ensures that the corresponding (strong) ARMA model with independent innovations and constant $c_t = c$ is stationary and invertible. The ε_t innovations, however, are uncorrelated but not independent here as their temporally changing variance creates interdependence. (Hence representation (4.4) is a weak ARMA one.) Their variance depends on the lagged value of the observed process X_t. The form of dependence – characterised by $\sigma^2(x)$ – is asymmetric and goes to ∞ proportionally to $x^{2\beta}$ for both positive and negative x-s. Hence, by Assumption 4.3, $\lim_{|x|\to\infty} \sigma^2(x)/x^2 = 0$. We will prove that this property implies the finiteness of all moments of X_t provided that the same is true for Z_t, making the model suitable for river flow analysis in this respect.

The asymmetric nature of $\sigma(x)$, together with the specification of the conditional variance as a function of the past value of X_t, yields that X_t has a highly skewed marginal distribution even in the case of a symmetrically distributed Z_t noise, in accordance with the aims of river flow modelling. If a linear ARMA model were used instead, innovations with asymmetric distribution would be necessary in order to produce skewness.

Although the model is new in its generality, some special cases have been introduced before. The pure ARCH-type version of the model (i.e. when no ARMA-equation is present, $p = q = 0$) was first proposed by Guegan and Diebolt (1994) and it was called a β-ARCH process. Following this terminology, the model presented in this dissertation could be named an ARMA-β-TARCH model to reflect both the presence of the ARMA filter and the asymmetric form of $\sigma(x)$. If the ARMA term is absent and $m = 0$ and $\beta = 1$, the model reduces to the standard TARCH model of (4.3) but this case (because of $\beta < 1$) lies outside the scope of our analysis.

In the following sections we examine the probabilistic and statistical properties of the model and compare them to the corresponding properties of the usual (quadratic) specifications.

4.2 Existence and moments of the stationary distribution

First we address the problem of existence of a stationary solution to model (4.4)-(4.6). Geometric ergodicity and hence existence of a unique stationary distribution was proven in the case of $q = 0$ (i.e. in the AR-β-TARCH case) in Masry and Tjostheim (1995) provided that all roots of the characteristic equation of the corresponding AR model lie within the unit circle (i.e. the corresponding linear AR model has a stationary solution). Guegan and Diebolt (1994) studied in detail the case $p = 0$ and $q = 0$ (i.e. the pure β-TARCH case without AR- or MA-terms). They proved that under the finiteness of all moments of Z_t, all moments of the stationary distribution of X_t are finite, too. We now prove the stationarity and moment properties of the general model (4.4)-(4.6). To avoid technical problems with seasonality, it is assumed that $c_t = c$ is a constant.

Theorem 4.1. *Assume that $c_t = c$ and Assumptions 4.1-4.4 hold. Then, the X_t process defined by* (4.4)-(4.6) *is geometrically ergodic and has a unique stationary distribution. If, moreover, $E(|Z_t|^r) < \infty$ for an $r \geq 2$ real, then $E\left(|X_t|^r\right) < \infty$ under the stationary distribution.*

Before the proof let us introduce a few notations and a technical lemma. We may assume that $p \geq 1$ and $q \geq 1$. Clearly,

$$\mathbf{Y_t} = (X_t - c, X_{t-1} - c, \ldots, X_{t-p+1} - c, \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q+1})^T$$

is a $(p+q)$-dimensional Markov chain, and there exists a matrix $\mathbf{A}$ for which

$$\mathbf{Y_t} = \mathbf{A}\mathbf{Y_{t-1}} + \mathbf{E}_t, \tag{4.7}$$

where $\mathbf{E}_t = (\varepsilon_t, 0, \ldots, 0, \varepsilon_t)^T$ is a $(p+q)$-dimensional random vector, q eigenvalues of $\mathbf{A}$ are 0, and the other p eigenvalues lie within the unit circle. There exists a real matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P^{-1}}$ is in real Jordan form. If λ_j is a real eigenvalue, a corresponding component of $\mathbf{B}$, depending on the multiplicity, is:

1. either $(\lambda_j) \in \mathbf{R}^{1\times 1}$
2. or a matrix where the diagonal elements are λ_j, the elements above the diagonal are 1 and the other elements are zero.

Similarly, if $a_j \pm b_j i$ is a complex eigenvalue pair, a corresponding component matrix is

1. either $\left(\begin{smallmatrix} a_j & b_j \\ -b_j & a_j \end{smallmatrix}\right) \in \mathbf{R}^{2\times 2}$
2. or a matrix where the diagonal structure consists of $\left(\begin{smallmatrix} a_j & b_j \\ -b_j & a_j \end{smallmatrix}\right)$ matrices, and there are matrices $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ directly above them (all other elements being zero).

Lemma 4.2. *For all $r \geq 2$ there exist a real number $\mu < 1$ and a diagonal matrix $\mathbf{Q} = diag(q_i)$ with positive diagonal elements such that for all $\mathbf{y} \in \mathbf{R}^{p+q}$*

$$||\mathbf{QBy}||_r \leq \mu ||\mathbf{Qy}||_r \tag{4.8}$$

where $||\mathbf{x}||_r$ denotes the r-norm of a vector $\mathbf{x}$.

Proof. Let $\rho < 1$ denote the maximum of the absolute values of the eigenvalues of $\mathbf{B}$ (and so of $\mathbf{A}$) and choose $\mu = \frac{\rho+1}{2}$. There are four types of components in $\mathbf{B}$, and the diagonal elements of $\mathbf{Q}$ may be chosen independently in each component. In the above mentioned first case with real eigenvalue λ_j, $q_j = 1$ will be an appropriate choice for the diagonal element of $\mathbf{Q}$. Similarly, in the first case with complex eigenvalue, $(1,1)^T \in \mathbf{R}^2$ will suffice as the diagonal component. In the second case with real eigenvalue the matrix with diagonal component $(\eta,\ldots,\eta,1)^T$ satisfies the inequality for a sufficiently small $\eta > 0$. Similarly, in the remaining case, $(\eta,\ldots,\eta,1,1)^T$ is appropriate with some $\eta > 0$. □

Proof of Theorem 4.1. Since Z_t has full support on the real line and $\sigma(x)$ is bounded away from zero, $\mathbf{Y_t}$ is easily seen to be ψ-irreducible and aperiodic, with ψ being the Lebesgue-measure. Moreover, as the density of Z_t is absolutely continuous, $\mathbf{Y_t}$ is a Feller chain. Thus, by Meyn and Tweedie (1993, Theorem 5.5.7 and 6.0.1), every compact set is small and smallness is equivalent to petiteness. (For the definition and properties of small and petite sets, see Meyn and Tweedie, 1993, Chapter 5.) Then, by Meyn and Tweedie (1993, Theorem 15.0.1) it is enough to find a suitable test function $V \geq 1$, a petite (small) set C, constants $b < \infty$ and $1 > \delta > 0$ such that

$$E(V(\mathbf{Y_1})|\mathbf{Y_0} = \mathbf{y}) \leq (1-\delta)V(\mathbf{y}) + bI_C(\mathbf{y}) \tag{4.9}$$

where I_C denotes the indicator function of the set C. (In other words, the conditional expectation should be bounded on C and be a contraction outside it.) This condition ensures that $\mathbf{Y_t}$ is geometrically ergodic, i.e. denoting by $P^n(\mathbf{y},.)$ the probability measure of the Markov-chain with initial state $\mathbf{y}$ after n steps, there is a unique invariant probability measure π for which $||P^n(\mathbf{y},.) - \pi|| = o(\rho^n)$ in the variation norm with some $0 < \rho < 1$. Moreover, the stationary distribution of $\mathbf{Y_t}$ has finite V-moment (see Meyn and Tweedie, 1993, Theorem 14.0.1).

Assume that $||Z_t||_{L^r}^r = E(|Z_t|^r) < \infty$ for an $r \geq 2$ real number and put

$$V(\mathbf{y}) = 1 + ||\mathbf{QPy}||_r^r.$$

In what follows let us denote the first component of a vector $\mathbf{y}$ by y^1. As $||\mathbf{QPE}_1||_r$ is proportional to $|\varepsilon_1|$ there exists an $s > 0$ such that $||\mathbf{QPE}_1||_r = s|\varepsilon_1| = s\sigma\left(Y_0^1\right)|Z_1|$. The identity $\mathbf{PA} = \mathbf{BP}$, the triangle inequality for the r-norm and inequality (4.8) yield

$$\begin{aligned} E\left(V(\mathbf{Y_1}) \mid \mathbf{Y_0} = \mathbf{y}\right) &= 1 + E\left(||\mathbf{QP}\left(\mathbf{Ay} + \mathbf{E}_1\right)||_r^r \mid \mathbf{Y_0} = \mathbf{y}\right) \\ &= 1 + E\left(||\mathbf{QBPy} + \mathbf{QPE}_1||_r^r \mid \mathbf{Y_0} = \mathbf{y}\right) \\ &\leq 1 + E\left(\left(||\mathbf{QBPy}||_r + ||\mathbf{QPE}_1||_r\right)^r \mid \mathbf{Y_0} = \mathbf{y}\right) \\ &\leq 1 + E\left(\left(||\mu\mathbf{QPy}||_r + ||\mathbf{QPE}_1||_r\right)^r \mid \mathbf{Y_0} = \mathbf{y}\right). \end{aligned}$$

Combined with Minkowski's inequality this gives

$$\begin{aligned} E\left((\mu||\mathbf{QPy}||_r + ||\mathbf{QPE}_1||_r)^r \mid \mathbf{Y_0} = \mathbf{y}\right) = E\left(\left(\mu||\mathbf{QPy}||_r + s\sigma(y^1)|Z_1|\right)^r\right) \leq \\ \leq \left(\mu||\mathbf{QPy}||_r + s\sigma(y^1)||Z_1||_{L_r}\right)^r \leq \left(\mu V^{1/r}(\mathbf{y}) + s\sigma(y^1)||Z_1||_{L^r}\right)^r. \end{aligned}$$

As $\mathbf{Q}$ and $\mathbf{P}$ are nonsingular matrices, $k_1\|\mathbf{y}\|_r^r \leq V(\mathbf{y}) - 1 \leq k_2\|\mathbf{y}\|_r^r$ for some positive constants k_1 and k_2. Furthermore, $\sigma(x) = o(|x|)$ as $x \to \infty$ so the second term in the last expression is dominated by the first one when $||\mathbf{y}||_r \to \infty$. Thus there exist a compact set $C = \{\mathbf{y} : \|\mathbf{y}\|_r \leq M\}$ and $\mu < \mu_2 < 1$ such that for all $\mathbf{y} \notin C$

$$E(V(\mathbf{Y_1})|\mathbf{Y_0} = \mathbf{y}) \leq \mu_2 V(\mathbf{y}).$$

Since $V(\mathbf{y})$ and $\sigma(x)$ are bounded on compact sets, (4.9) is satisfied with $1 - \delta = \mu_2$ and with a suitably chosen b. This means that the stationary distribution exists and has finite V-moments and consequently has finite rth moments. This concludes the proof. □

Remark 4.3. *Assume that $E(|Z_t|^r) < \infty$ and a function g satisfies $g(x) = O(|x|^r)$. Then, by Meyn and Tweedie (1993, Theorem 17.1.7.) the strong law of large numbers holds for the $g(X_t)$ process:*

$$\frac{1}{n}\sum_{t=1}^{n} g(X_t) \to E_\pi(g(X_t)) \text{ a.s.,} \tag{4.10}$$

where E_π denotes the expectation under the stationary distribution.

The results of Theorem 4.1 (and also the previously mentioned results on the special cases) go against the classical ARCH(1) model of quadratic heteroscedasticity (equation (4.2)) where the domain of stationarity depends on the parameters of the $\sigma^2(x)$ function and on the distribution of the noise Z_t. Moreover, not all moments of the stationary distribution are finite in this case. For instance, the simple quadratic ARCH model without AR-term and with normally distributed Z_t has a stationary distribution if and only if $0 \leq \alpha_1 < 2\exp(\gamma) \approx 3.56$ where γ is the Euler-constant, and it has a finite variance only if $\alpha_1 < 1$. (For more details see Borkovec and Klüppelberg, 2001.)

In the sequel, unless otherwise indicated, all probability statements correspond to the stationary distribution of X_t.

4.3 Tail behaviour

Let us now turn to the tail behaviour, i.e. to the decay of the tail of the stationary distribution. The tail behaviour of the standard (quadratic) ARCH processes is much studied in the literature. It has been known for more than 15 years (Goldie, 1991) that in the absence of ARMA-parameters the simple quadratic ARCH process has regularly varying tail even when Z_t is normally distributed. This means that a light-tailed input results in a heavy-tailed output in the case of ARCH-models. More generally, Borkovec and Klüppelberg (2001) proved that the AR(1)-ARCH(1) model also has regularly varying tail for a wide class of noise distributions including the normal or Laplace ones. It follows that the stationary distribution of such processes belongs to the domain of attraction of the Frechet-distribution, or in the peaks-over-threshold framework their tail can be approximated by a GPD with shape parameter $\xi > 0$.

The situation is very different for β-ARCH-type processes. It follows from the previous section that if all moments of the generating noise is finite, all moments of the stationary distribution of X_t will be finite, too. The finiteness of all moments and the infinite support of a distribution imply that the shape parameter of the GPD (if the distribution belongs to the domain of attraction of a GPD at all) is zero. Hence, e.g. a normal or Laplace-distributed noise can only generate a stationary distribution with $\xi = 0$ in the ARMA-β-TARCH model under Assumptions 4.1-4.4.

This finding, however, does not determine the exact tail behaviour: as it was illustrated in section 2.2.1 the Gumbel-domain (the $\xi = 0$ case) contains many different types of distributions. In fact, compared to the quadratic ARCH models, little is known about the precise tail behaviour of ARMA-β-ARCH-type processes: even the case without ARMA-terms (i.e. $p = q = 0$) is not yet settled in its full generality. Therefore, in this section, we focus on the pure β-TARCH case, and examine the model

$$X_t = \left(\alpha_0 + \alpha_{1+}\left(X_{t-1}^{+}\right)^{2\beta} + \alpha_{1-}\left(X_{t-1}^{-}\right)^{2\beta}\right)^{1/2} Z_t. \tag{4.11}$$

Being the tail of X_t very sensitive to the tail of Z_t, more assumptions are needed on the latter than in the previous section. We assume the following:

Assumption 4.5. *Z_t is an i.i.d. sequence and there exist $u_0 > 0$, $\gamma > 0$, $K_1 > 0$ and K_2 such that its probability density satisfies*

$$f_{Z_t}(u) = K_1 |u|^{K_2} \exp\left(-\kappa |u|^{\gamma}\right) \tag{4.12}$$

for every $|u| > u_0$.

Thus Z_t is symmetric and has a Weibull-like tail with exponent γ in the sense of Definition 2.9. The Gaussian $(\gamma = 2)$ or the Laplace $(\gamma = 1)$ distributions are obtained as special cases.

Guegan and Diebolt (1994) showed under the additional assumption $\alpha_{1+} > 0$ and $\alpha_{1-} > 0$ that if $\beta > (\gamma - 1)/\gamma$, X_t has no exponential moment (i.e. it is heavier tailed than the exponential distribution) while if $\beta < (\gamma - 1)/\gamma$, X_t has a moment generating function defined around the neighbourhood of zero. This finding already suggests that X_t may possess (approximately) a Weibull-like tail with exponent $\gamma(1 - \beta)$. Assuming a normally distributed noise (i.e. $\gamma = 2$), $\alpha_{1+} = \alpha_{1-}$ and $1/2 < \beta < 1$, Robert (2000) argued that this is indeed the case: under his assumptions X_t has Weibull-like tail with exponent $2\left(1-\beta\right)$. Although the proof of his findings seems to be incomplete,[2] some of his ideas are useful to prove that X_t has approximately Weibull-like tail even if we allow the more general case, i.e. $\alpha_{1+} \neq \alpha_{1-}$, $\gamma \neq 2$ and $0 < \beta \leq 1/2$.

Theorem 4.4. *Assume* (4.11)*, Assumption 4.5,* $\alpha_0 > 0$, $\alpha_{1+} > 0$, $\alpha_{1-} > 0$ *and* $0 < \beta < 1$. *Then, using the notation* $\alpha_1^{\max} = \max\left(\alpha_{1+}, \alpha_{1-}\right)$,

$$\exp\left(-\frac{\left(\alpha_1^{\max}\right)^{-\gamma/2}\kappa\gamma\beta^{-\frac{\beta}{1-\beta}}}{2}u^{\gamma(1-\beta)} + O\left(u^{\gamma(1-\beta)/2}\right)\right) \leq \bar{F}_{X_t}\left(u\right)$$
$$\leq \exp\left(-\frac{\left(\alpha_1^{\max}+\alpha_0\right)^{-\gamma/2}\kappa\gamma\beta^{-\frac{\beta}{1-\beta}}}{2}u^{\gamma(1-\beta)} + O\left(u^{\gamma(1-\beta)/2}\right)\right). \quad (4.13)$$

Proof. We may assume without loss of generality that $\alpha_1^{\max} = \alpha_{1+} \geq \alpha_{1-}$. Let $Y_t = \log(X_t^2)$, $U_{t,1} = \log\left(\alpha_{1+}Z_t^2\right)$, and $U_{t,2} = \log\left(\alpha_{1-}Z_t^2\right)$. Furthermore, let us introduce the functions

$$h_1\left(y\right) = \log\left(\alpha_0/\alpha_{1+} + \exp\left(\beta y\right)\right),$$
$$h_2\left(y\right) = \log\left(\alpha_0/\alpha_{1-} + \exp\left(\beta y\right)\right)$$

and the random variables $V_{t,i} = h_i\left(Y_{t-1}\right) - \beta Y_{t-1}$ $\left(i = 1, 2\right)$. Then

$$Y_t = h_1\left(Y_{t-1}\right) + U_{t,1} = \beta Y_{t-1} + U_{t,1} + V_{t,1} \qquad \text{if} \quad Z_{t-1} > 0,$$
$$Y_t = h_2\left(Y_{t-1}\right) + U_{t,2} = \beta Y_{t-1} + U_{t,2} + V_{t,2} \qquad \text{if} \quad Z_{t-1} \leq 0.$$

Since $h_i\left(y\right) \geq \beta y$ $\left(i = 1, 2\right)$, $V_{t,i} \geq 0$ a.s. Moreover, as Z_t is a symmetrically distributed i.i.d. sequence, Y_t can be written as

$$Y_t = \beta Y_{t-1} + U_t + V_t$$

where U_t is an independent 1/2-1/2 mixture of $U_{t,1}$ and $U_{t,2}$ and similarly V_t is an independent 1/2-1/2 mixture of $V_{t,1}$ and $V_{t,2}$.

Let us introduce the auxiliary sequence

$$Y_t^* = \beta Y_{t-1}^* + U_t = \sum_{i=0}^{\infty}\beta^{i-1}U_{t-i}.$$

[2] He derives a functional equation for the logarithm of the moment generating function $L_Y(s)$ of Y_t and estimates the tail of Y_t based on the behaviour of $L_Y(s)$ around ∞. During the calculations he assumes (see Appendix 1 of his paper) that if a function g satisfies $g(x) - g(\alpha x) = O(1/x)$ as $x \to \infty$, then $g(x) = O(1/x)$. However, this is not the case: if e.g. $g(x) = \sin\left(2\pi \log x/\log\alpha\right)$ then $g(x) - g(\alpha x) = 0$.

It is clear that $Y_t^* \leq Y_t$, therefore by examining the tail behaviour of Y_t^* we obtain a lower bound for the tail of Y_t as well.

Since U_t is sufficiently light-tailed, Y_t^* lies within the framework of Klüppelberg and Lindner (2005) who examined the tail behaviour of linear moving average processes with light-tailed increments, i.e. of $\sum_{-\infty}^{\infty} c_i W_{t-i}$. They assume that the probability density of the i.i.d. increment sequence W_t satisfies

$$f(u) = \nu(u) \exp\left(-\psi(u)\right), \quad u \geq u_0$$

for some u_0, and $\psi(u)$ is C^2, $\psi'(u_0) = 0$, $\psi'(\infty) = \infty$, ψ'' is strictly positive on $[u_0, \infty]$ and $\phi = 1/\sqrt{\psi''}$ is self-neglecting, i.e.

$$\lim_{u\to\infty} \frac{\phi\left(u + x\phi\left(u\right)\right)}{\phi\left(u\right)} = 1$$

uniformly on bounded x-intervals. The function ν is assumed to be flat for ϕ, i.e.

$$\lim_{u\to\infty} \frac{\nu\left(u + x\phi\left(u\right)\right)}{\nu\left(u\right)} = 1$$

uniformly on bounded x-intervals. Furthermore, define

$$q(\tau) = \psi'^{-1}(\tau), \qquad q_i(\tau) = c_i q(c_i\tau), \qquad Q(\tau) = \sum_{i=-\infty}^{\infty} q_i(\tau),$$

$$S^2(\tau) = q'(\tau), \qquad \sigma_i^2(\tau) = q_i'(\tau), \qquad \sigma_\infty^2(\tau) = \sum_{i=-\infty}^{\infty} \sigma_i^2(\tau).$$

It follows from the conditions that Q is a strictly increasing function. Then, provided that c_i is a summable sequence of non-negative real numbers, not all zero, and assuming that the two conditions below hold:

$$\lim_{m\to\infty} \limsup_{\tau\to\infty} \frac{\sum_{|j|>m} \sigma_j^2\left(\tau\right)}{\sigma_\infty^2\left(\tau\right)} = 0, \tag{4.14}$$

$$\lim_{m\to\infty} \limsup_{\tau\to\infty} \frac{\sum_{|j|>m} \sigma_j\left(\tau\right)}{\sigma_\infty^2\left(\tau\right)} = 0, \tag{4.15}$$

the following theorem is true:

Theorem 4.5. *(Klüppelberg and Lindner, 2005) Under the above conditions, as $u \to \infty$,*

$$P\left(\sum_{i=-\infty}^{\infty} c_i W_{t-i} > u\right)$$
$$\sim \frac{1/\sqrt{2\pi}}{Q^{-1}\left(u\right)\sigma_\infty\left(Q^{-1}\left(u\right)\right)} \exp\left(-\int_{u_0\sum c_i}^{u} \left(Q^{-1}\left(v\right) + \rho\left(Q^{-1}\left(v\right)\right)\right)dv\right)$$

where $\rho\left(\tau\right) = o\left(1/\sigma_\infty\left(\tau\right)\right)$. It is also true that $1/\sigma_\infty\left(\tau\right) = o\left(\tau\right)$ so the first term in the integral is the leading term.

In our case, $U_t = W_t$ and

$$f_{U_t}(u) = \frac{1}{2}\left(K_1 \exp\left(K_2 u\right) \exp\left(-\kappa\left(\alpha_{1+}\right)^{-\gamma/2} e^{\frac{\gamma u}{2}}\right)\right)$$
$$+\frac{1}{2}\left(K_3 \exp\left(K_4 u\right) \exp\left(-\kappa\left(\alpha_{1-}\right)^{-\gamma/2} e^{\frac{\gamma u}{2}}\right)\right) = \nu(u)\exp\left(-\psi(u)\right)$$

where – in order to satisfy the necessary assumptions – $\psi(u)$ can be defined as

$$\begin{aligned} \psi(u) &= \kappa\left(\alpha_{1+}\right)^{-\gamma/2} e^{\frac{\gamma u}{2}} - e\kappa\left(\alpha_{1+}\right)^{-\gamma/2}/2 && \text{if} \quad u \geq 2/\gamma, \\ \psi(u) &= e\kappa\left(\alpha_{1+}\right)^{-\gamma/2}\left(\gamma/2\right)^2 u^2/2 && \text{if} \quad u < 2/\gamma. \end{aligned}$$

Then it is a routine matter to check that the resulting $\nu(u)$ function is flat for $\psi(u)$ and $\phi(u)$ is self-neglecting (see also Example 2.4. (c) in Klüppelberg and Lindner (2005)), so the tail of Y_t^* can be approximated in principle using $c_i = \beta^{i-1}$ for $i \geq 0$ and $c_i = 0$ for $i < 0$. (Conditions (4.14)-(4.15) should also be checked, see below.) Using the notation $\tau_0 = e\kappa\left(\alpha_{1+}\right)^{-\gamma/2}\gamma/2$, we obtain

$$\begin{aligned} \psi'(u) &= \kappa\left(\alpha_{1+}\right)^{-\gamma/2}\left(\gamma/2\right) e^{\frac{\gamma u}{2}} = e^{-1}\tau_0 e^{\frac{\gamma u}{2}} && \text{if} \quad u \geq 2/\gamma, \\ \psi'(u) &= e\kappa\left(\alpha_{1+}\right)^{-\gamma/2}\left(\gamma/2\right)^2 u = \tau_0\left(\gamma/2\right) u && \text{if} \quad u < 2/\gamma \end{aligned}$$

and hence

$$\begin{aligned} q(\tau) &= 2\gamma^{-1}\tau/\tau_0 && \text{if} \quad \tau < \tau_0 \\ q(\tau) &= 2\gamma^{-1}\log\left(e\tau/\tau_0\right) && \text{if} \quad \tau \geq \tau_0. \end{aligned}$$

Then

$$Q(\tau) = \sum_{j=0}^{\infty}\beta^j q\left(\beta^j\tau\right) = 2\gamma^{-1}\sum_{j=0}^{\infty}\beta^j\left(\log\left(\beta^j\tau\right) - \log\tau_0 + 1\right)$$
$$+ 2\gamma^{-1}\sum_{j:\ \beta^j\tau<\tau_0}\beta^j\left(\beta^j\tau/\tau_0 - \log\left(\beta^j e\tau/\tau_0\right)\right). \quad (4.16)$$

Nevertheless, for any $0 < \theta < 1$, the second term can be written as

$$\sum_{j:\ \beta^j\tau<\tau_0}\beta^j\left(\beta^j\tau/\tau_0 - \log\left(\beta^j e\tau/\tau_0\right)\right)$$
$$= \left(e\tau/\tau_0\right)^{-\theta}\sum_{j:\ \beta^j\tau<\tau_0}\left(\beta^{1-\theta}\right)^j\left(e^\theta\left(\beta^j\tau/\tau_0\right)^{1+\theta} - \left(\beta^j e\tau/\tau_0\right)^\theta\log\left(\beta^j e\tau/\tau_0\right)\right). \quad (4.17)$$

Since the function $f(x) = x^\theta\log x \to 0$ as $x \to 0$ for $0 < \theta < 1$, $f(x)$ is bounded on $(0, e]$. Hence $\left(\beta^j e\tau/\tau_0\right)^\theta\log\left(\beta^j e\tau/\tau_0\right)$ and trivially $e^\theta\left(\beta^j\tau/\tau_0\right)^{1+\theta}$ are bounded if $\beta^j\tau/\tau_0 < 1$, thus

$$\sum_{j:\ \beta^j\tau<\tau_0}\left(\beta^{1-\theta}\right)^j\left(e^\theta\left(\beta^j\tau/\tau_0\right)^{1+\theta} - \left(\beta^j e\tau/\tau_0\right)^\theta\log\left(\beta^j e\tau/\tau_0\right)\right) = o(1)$$

as $\tau \to \infty$. Therefore the sum in (4.17) (and so the second term in (4.16)) is $o\left(\tau^{-\theta}\right)$, thus

$$Q(\tau) = 2\gamma^{-1}(1-\beta)^{-1}\left(\log\tau + \beta(1-\beta)^{-1}\log\beta + 1 - \log\tau_0\right) + o\left(\tau^{-\theta}\right)$$

as $\tau \to \infty$, which is of the form $Q(\tau) = A(\log\tau + B) + o\left(\tau^{-\theta}\right)$. Trivially, $Q^{-1}(u) \to \infty$ as $u \to \infty$, hence

$$Q^{-1}(u) = \exp\left(A^{-1}u - B\right)\exp\left(-o\left(\left(Q^{-1}(u)\right)^{-\theta}\right)\right) = \exp\left(A^{-1}u - B\right)(1 + o(1)).$$

Using this we obtain a better estimate for $Q^{-1}(u)$ in the second round:

$$\begin{aligned} Q^{-1}(u) &= \exp\left(A^{-1}u - B\right)\left(1 + o\left(\exp\left(-\theta A^{-1}u\right)\right)\right) \\ &= \frac{\kappa\gamma(\alpha_{1+})^{-\gamma/2}\beta^{-\frac{\beta}{1-\beta}}}{2}\exp\left(\frac{\gamma(1-\beta)}{2}u\right) + o\left(\exp\left(\frac{\gamma(1-\beta)(1-\theta)}{2}u\right)\right). \end{aligned}$$

Moreover, $q'(\tau) = 2\gamma^{-1}/\tau_0$ if $\tau < \tau_0$ and $q'(\tau) = 2\gamma^{-1}/\tau$ if $\tau \geq \tau_0$, hence $\sigma_i^2(\tau) = 2\gamma^{-1}\beta^{2i}/\tau_0$ if $\beta^i\tau < \tau_0$ and $\sigma_i^2(\tau) = 2\gamma^{-1}\beta^i/\tau$ if $\beta^i\tau \geq \tau_0$. Thus

$$\sigma_\infty^2(\tau) = 2\gamma^{-1}\left(\sum_{j:\ \beta^j\tau\geq\tau_0}\beta^j/\tau + \sum_{j:\ \beta^j\tau<\tau_0}\beta^{2j}/\tau_0\right) \sim 2\gamma^{-1}(1-\beta)^{-1}/\tau,$$

so (4.14)-(4.15) hold for the $c_i = \beta^{i-1}$ sequence, thus Theorem 4.5 can be applied. It also follows that $\rho(\tau) = O\left(\tau^{1/2}\right)$ and $u_0 = 0$ in that Theorem. If we choose $\theta > 1/2$, we obtain

$$\begin{aligned} \bar{F}_{Y_t^*}(u) &= \exp\left(-\int_0^u\left(Q^{-1}(v) + O\left(e^{\frac{\gamma(1-\beta)}{4}v}\right)dv\right)\right) \\ &= \exp\left(-\frac{\kappa(\alpha_{1+})^{-\gamma/2}\beta^{-\frac{\beta}{1-\beta}}}{1-\beta}e^{\frac{\gamma(1-\beta)}{2}u} + O\left(e^{\frac{\gamma(1-\beta)}{4}u}\right)\right). \quad (4.18) \end{aligned}$$

Taking into account that $Y_t^* \leq Y_t = \log\left(X_t^2\right)$, the lower bound is obtained for $\bar{F}_{X_t}(u)$ in (4.13).

To show the upper bound for the tail, let us first observe that, trivially, the increase of either α_{1+} or α_{1-} does not make the tail of Y_t lighter. Therefore, we can assume that $\alpha_{1+} = \alpha_{1-}$ and get an upper bound for the tail of this restricted model. In this case, let us introduce for each t a random variable $U_t^{**} \geq 0$ such that $U_t \leq U_t^{**}$ a.s. and $f_{U_t^{**}}(u) = Kf_{U_t}(u)$ for all $u > 0$ with an appropriate $K > 0$. (Such a variable can be easily constructed.) Define also $h(y) = \beta y\chi_{\{y\geq 0\}} + \log(1 + \alpha_0/\alpha_{1+})$. It follows from $\alpha_{1+} = \alpha_{1-}$ that $h_i(y) \leq h(y)$ $(i = 1, 2)$ and thus it can be shown straightforwardly that $\bar{F}_{Y_t^{**}}(u) \geq \bar{F}_{Y_t}(u)$ holds for the stationary distribution of the model defined by

$$Y_t^{**} = h\left(Y_{t-1}^{**}\right) + U_t^{**}.$$

Indeed, let $\hat{Y}_0 = Y_0$ and define $\hat{Y}_t$ recursively as

$$\hat{Y}_t = h\left(\hat{Y}_{t-1}\right) + U_t^{**}.$$

Using $U_t \leq U_t^{**}$ we can prove by induction that $Y_t \leq \hat{Y}_t$:

$$Y_t \leq h\left(Y_{t-1}\right) + U_t \leq h\left(\hat{Y}_{t-1}\right) + U_t \leq h\left(\hat{Y}_{t-1}\right) + U_t^{**} = \hat{Y}_t.$$

Since the distribution of $\hat{Y}_t$ tends to the stationary distribution of Y_t^{**} as $t \to \infty$, Y_t is indeed stochastically smaller than Y_t^{**}.

As $h(y) \geq 0$ for all y and $U_t^{**} \geq 0$ a.s., an alternative definition for Y_t^{**} is

$$Y_t^{**} = \beta Y_{t-1}^{**} + U_t^{**} + \log\left(1 + \alpha_0/\alpha_{1+}\right) = \sum_{i=0}^{\infty} \beta^{i-1} U_{t-i}^{**} + \frac{\log\left(1 + \alpha_0/\alpha_{1+}\right)}{1-\beta}.$$

The result of Klüppelberg and Lindner (2005) again gives that the tail of $\sum_{i=0}^{\infty} \beta^{i-1} U_{t-i}^{**}$ has the same form as the tail of Y_t^* (equation (4.18)). Then it is easy to derive the upper bound for the tail of X_t in (4.13). □

Theorem 4.4 falls short of stating that X_t has Weibull-like tail as defined in Definition 2.9 because it does not give a precise asymptotics but only bounds the tail above and below by Weibull-like distributions with the same exponent.

What happens when certain restrictions of the theorem are relaxed? If $\alpha_{1-} = 0$ is allowed, then the upper bound in (4.13) certainly holds and a slightly weaker lower bound can also be proven easily:

Proposition 4.6. *Assume the assumptions of Theorem 4.4 but allow $\alpha_{1-} = 0$. Then for every $\delta > 0$ there exists a $K > 0$ such that*

$$\exp\left(-Ku^{(1+\delta)\gamma(1-\beta)}\right) \leq \bar{F}_{X_t}(u). \tag{4.19}$$

Proof. Let us use the same notations as in the proof of Theorem 4.4 and let Y_t^{***} be defined by

$$\begin{aligned} Y_t^{***} &= \beta Y_{t-1}^{***} + U_t^{+} && \text{if} \quad Z_t \geq 0 \\ Y_t^{***} &= U_t^{0} && \text{if} \quad Z_t < 0 \end{aligned}$$

where $U_t^0 = \log\left(\alpha_0 Z_t^2\right)$. It is easily shown that $Y_t^{***} \leq Y_t$ stochastically in this case. Moreover, as Z_t is symmetrically distributed, for every $n \in \mathrm{Z}^+$ with probability 2^{-n}

$$Y_t^{***} = \sum_{i=0}^{n-1} \beta^i U_{t-i}^{+} + \beta^n Y_{t-n}^{***}.$$

Therefore, using the notations $q = \bar{F}_{Y^{***}_{t-n}}(0)$ and $Y_{t,n} = \sum_{i=0}^{n-1} \beta^i U^+_{t-i}$,

$$\bar{F}_{Y^{***}_t}(u) \geq q 2^{-n} \bar{F}_{Y_{t,n}}(u).$$

Moreover, similarly to the derivation of the tail of Y^*_t, it follows again from Theorem 4.5 that

$$\bar{F}_{Y_{t,n}}(u) = \exp\left(-Ke^{\frac{\gamma(1-\beta)}{2(1-\beta^n)}} + O\left(e^{\frac{\gamma(1-\beta)}{4(1-\beta^n)}}\right)\right)$$

with a suitable $K > 0$. Choosing n such that $1/(1-\beta^n) < 1+\delta$ we obtain

$$\bar{F}_{Y^{***}_t}(u) \geq \exp\left(-Ke^{\frac{(1+\delta)\gamma(1-\beta)}{2}}\right)$$

with a possibly different $K > 0$, and transforming Y_t to X_t gives the statement of the proposition. □

It is a natural question to ask how the tail behaviour is modified when AR- or MA-terms are added to the simple uncorrelated model. Unfortunately, this question is not yet settled but we conjecture that the more general ARMA-β-TARCH model has approximately a Weibull-like tail, too.

4.4 Extremal clustering behaviour

The next question regards the extremal clustering behaviour of the ARMA-β-TARCH model. It is a well-known fact that, regardless of the parameters, the extremal index of a simple quadratic ARCH model is strictly less than one, and this is the case for the quadratic AR-ARCH model as well (Borkovec, 2000). Things are changing, however, when $0 < \beta < 1$. It is conjectured that an AR-β-(T)ARCH model has extremal index equal to one, i.e. it does not exhibit clustering at extreme levels.

To illustrate this conjecture, Figure 4.1 shows the extremal index estimated by the method of Ferro and Segers (see section 2.2.2) for three different processes. The first process is a linear AR(1) model, the second is an AR(1)-β-ARCH(1) model with $\beta = 1/2$ and the third is an AR(1)-ARCH(1) model (i.e. $\beta = 1$). The AR-term is chosen as 0.3 for all cases. As the threshold is increasing, the extremal index estimate tends to one (apart from random fluctuation) for the first and the second model, indicating the absence of extremal clustering, while it stays well below one for the third process, in line with theory.

4.5 Model estimation

In this section we address the issue of parameter estimation. Besides treating the model orders p and q as given, we also keep the parameters m and β fixed. Since the partial

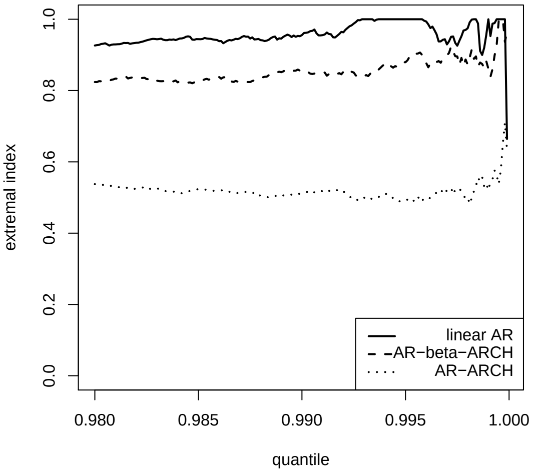


Figure 4.1: The extremal index estimate as a function of the threshold as a quantile for the linear AR, the AR-β-ARCH and the AR-ARCH models

derivative of $\sigma(x)$ with respect to m and β does not exist at every point, the estimation of these parameters is not possible within the standard likelihood framework but would require e.g. the concept of Frechet derivatives (see Hwang and Basawa (2003)). Details of identification of p, q, m and β are described in section 4.6 for the particular hydrological application considered in this dissertation.

Let us denote the set of the linear filter and variance parameters to be estimated by $\boldsymbol{\theta}^0 = (a_1, \ldots, a_p, b_1, \ldots, b_q)$ and by $\boldsymbol{\alpha}^0 = (\alpha_0, \alpha_{1+}, \alpha_{1-})$. (As we are concerned with asymptotic properties, and these are not affected by replacing the expectation by the empirical mean, in the following we make the technical assumption that the mean parameter c is also known.) In order to achieve a fast algorithm for large sample sizes, a two stage estimation procedure is applied instead of full maximum likelihood. We will prove that no efficiency is lost by this method compared to the full ML procedure in the estimation of $\boldsymbol{\alpha}^0$. (Some efficiency may be lost in the estimation of $\boldsymbol{\theta}^0$ but – compared to $\boldsymbol{\alpha}^0$ – it is estimated quite accurately regardless of the particular method chosen.)

As usual in the estimation literature, we impose some compactness restrictions on the parameter space. We assume that the roots of the AR and MA characteristic polynomials lie in a closed subset disjoint from the closed unit disk. Furthermore, the parameter space for the variance equation is compact, the true variance parameter vector lies within its interior and the set of possible constant terms in the variance equation is separated from zero:

Assumption 4.6. *There exists a $\delta > 0$ such that $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}_\delta$ where*

$$\boldsymbol{\Theta}_\delta = \{\boldsymbol{\theta} \in \mathbf{R}^{p+q} : \textit{the roots of } \Phi_{\boldsymbol{\theta}}(x) \textit{ and } \Psi_{\boldsymbol{\theta}}(x) \textit{ have moduli } \geq 1 + \delta\}.$$

Moreover, $\boldsymbol{\alpha}^0 \in \text{int}(\mathbf{K})$*, where* $\mathbf{K}$ *is a compact subset of* $\mathbf{R}_{++} \times \mathbf{R}_{+} \times \mathbf{R}_{+}$.

The estimation method is as follows. First, based on the weak ARMA representation of X_t the least squares estimator of $\boldsymbol{\theta}^0$ is obtained. Repeating the expressions in section 2.1, the estimator $\hat{\boldsymbol{\theta}}_n$ minimizes

$$\hat{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} e_t^2(\boldsymbol{\theta})$$

where $e_t(\boldsymbol{\theta})$ is recursively defined as

$$e_t(\boldsymbol{\theta}) = X_t - c - \sum_{i=1}^{p} \theta_i(X_{t-i} - c) - \sum_{j=1}^{q} \theta_{p+j} e_{t-j}(\boldsymbol{\theta}).$$

The unknown starting values are set to zero and $\boldsymbol{\theta}$ denotes $(\theta_1, \ldots, \theta_{p+q})$. Hereinafter, we use the notation $\hat{\varepsilon}_t$ for the fitted innovations at $\hat{\boldsymbol{\theta}}_n$, i.e. $\hat{\varepsilon}_t = e_t(\hat{\boldsymbol{\theta}}_n)$.

It follows e.g. from Theorem 1 in Francq and Zakoian (1998) that the least squares estimator of a weak ARMA process is consistent provided that all roots of the AR and MA characteristic polynomials lie in a closed subset disjoint from the closed unit disk, they have no common zeros and the process belongs to L^2. Hence, Assumptions 4.1-4.3 and 4.6 together imply that $\hat{\boldsymbol{\theta}}_n$ is consistent, i.e. $\hat{\boldsymbol{\theta}}_n \rightarrow \boldsymbol{\theta}^0$ a.s. For asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ an additional condition is needed:

Assumption 4.7. $E(|Z_t|^{4+2\eta}) < \infty$ *holds for some* $\eta > 0$.

By Theorem 4.1 this assumption implies that $E(|X_t|^{4+2\eta}) < \infty$. Using Theorem 2 in Francq and Zakoian (1998) the $n^{1/2}$-consistency and asymptotic normality of the least squares estimator follows.[3]

The estimation of the parameter vector $\boldsymbol{\alpha}^0 = (\alpha_0, \alpha_{1+}, \alpha_{1-})$ in the variance equation can be carried out by Gaussian quasi maximum likelihood (QML). (The term "quasi" refers to the fact that although the likelihood is defined as if Z_t were normally distributed, much weaker distributional assumptions are needed for consistency and asymptotic normality.) Denoting the sample size by n, this means the maximization of the following term:

$$\hat{L}_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{t=1}^{n} l(\hat{\varepsilon}_t, X_{t-1}, \boldsymbol{\alpha}), \tag{4.20}$$

[3] According to Theorem 2 in Francq and Zakoian (1998) asymptotic normality of the least squares estimator of a weak ARMA representation holds if, in addition to the conditions of Theorem 1, the $(4 + 2\eta)$th moment condition of the X_t process and an additional summability condition of the mixing coefficients defined in Definition 2.10 are satisfied. However, the summability condition trivially holds in our case because of the geometric ergodicity of X_t, which follows from Theorem 4.1.

where

$$l(y,x,\boldsymbol{\alpha}) = -\log\sqrt{2\pi} - \frac{1}{2}\log(\sigma^2(x,\boldsymbol{\alpha})) - \frac{1}{2}\frac{y^2}{\sigma^2(x,\boldsymbol{\alpha})} \tag{4.21}$$

is the log-likelihood contribution of the observation y coming from a Gaussian distribution with zero mean and $\sigma^2(x,\boldsymbol{\alpha})$ variance as defined in (4.6). (With this notation we emphasize that σ depends on the parameter vector $\boldsymbol{\alpha}$, too.)

Then the following holds.

Theorem 4.7. *Under Assumptions 4.1-4.3 and 4.6 the QML estimator is consistent, i.e.*

$$\hat{\boldsymbol{\alpha}}_n \to \boldsymbol{\alpha}^0 \quad \textit{a.s.} \tag{4.22}$$

If in addition Assumption 4.7 holds, the resulting estimator is asymptotically normally distributed, i.e.

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0) \to_d N\left(\mathbf{0}, \mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\mathbf{V}\left(\boldsymbol{\alpha}^0\right)\mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\right) \tag{4.23}$$

where

$$\begin{aligned} \mathbf{V}(\boldsymbol{\alpha}) &= E_\pi\left(\frac{\partial l(\varepsilon_t, X_{t-1},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\frac{\partial l(\varepsilon_t, X_{t-1},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}^T\right) &(4.24)\\ \mathbf{H}(\boldsymbol{\alpha}) &= E_\pi\left(-\frac{\partial^2 l(\varepsilon_t, X_{t-1},\boldsymbol{\alpha})}{\partial^2\boldsymbol{\alpha}}\right). &(4.25)\end{aligned}$$

Moreover, the $\mathbf{H}(\boldsymbol{\alpha}^0)$ *and* $\mathbf{V}(\boldsymbol{\alpha}^0)$ *matrices can be consistently estimated by the empirical counterparts of* $\mathbf{H}(\hat{\boldsymbol{\alpha}}_n)$ *and* $\mathbf{V}(\hat{\boldsymbol{\alpha}}_n)$, *where expectations are replaced by sample averages.*

Proof. Let us first assume that the true ε_t innovations are observed (i.e. that $\boldsymbol{\theta}^0$ is known a priori and so $\hat{\varepsilon}_t = \varepsilon_t$ holds for all t). Then the maximum likelihood estimator maximizes

$$L_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{t=1}^{n} l(\varepsilon_t, X_{t-1},\boldsymbol{\alpha}). \tag{4.26}$$

The proof of consistency in this case goes directly along the lines of Kristensen and Rahbek (2005). They consider the ARCH model

$$\begin{aligned}\varepsilon_t &= \sigma_t Z_t\\ \sigma_t^2 &= \alpha_0 + \boldsymbol{\alpha_1}'\mathbf{U}_t,\end{aligned}$$

where $\mathbf{U}_t$ is an m-dimensional process and $\{(\epsilon_t, \mathbf{U}_t)\}$ is observed. Clearly, our model fits into this framework by choosing $\boldsymbol{\alpha_1} = (\alpha_{1+}, \alpha_{1-})$ and $\mathbf{U}_t = \left((X_t - m)^+, (X_t - m)^-\right)$.

For the consistency, conditions C.1.-C.4. of Kristensen and Rahbek (2005) are needed. However, they note that if the the parameter space is compact (our Assumption 4.6), condition C.1. is not required and C.4. (ii) can be weakened to the $E\left[|\log\sigma_t^2|\right]<\infty$ assumption, which is trivially satisfied in our case because X_t has finite variance (by Theorem 4.1) and $\sigma_t^2 \geq \alpha_0$. Condition C.2. is about the parameter space, and is actually weaker than our Assumption 4.6. Finally, C.3. assumes that $\mathbf{U}_t \in \mathbf{R}_+^m$, there does not exist a $\mathbf{v}\in\mathbf{R}^m$ and $c\in\mathbf{R}$ such that $P\left(\mathbf{v}'\mathbf{U}_t = c\right) = 1$, and $\mathbf{U}_t$ is measurable with respect to the σ-algebra generated by $(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-r})$ for an $r \geq 1$. Among these conditions, only the last one is not satisfied by our model because here X_t depends on the infinite past of $\{\epsilon_t\}$. Nevertheless, the last condition can be replaced by the assumption that $(\varepsilon_t, \mathbf{U}_{t-1})$ is embedded into a multidimensional Markov chain.[4] In our case, $(\varepsilon_t, \ldots, \varepsilon_{t-q}, \mathbf{U}_{t-1}, \ldots, \mathbf{U}_{t-p})$ is a Markov chain, hence all conditions of consistency hold when ε_t is properly observed.

For later reference, let us summarize the main steps of the proof of consistency in Kristensen and Rahbek (2005). According to the ergodic theorem

$$L_n(\boldsymbol{\alpha}) \to L(\boldsymbol{\alpha}) = E_\pi l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha}) \quad \text{a.s.}$$

and it can be proven that $L(\boldsymbol{\alpha})$ obtains its maximum at the true parameter value $\boldsymbol{\alpha}^0$ (Kristensen and Rahbek, 2005, Lemma 3). Thus, after some technical details, it is not surprising that the maximisation of $L_n(\boldsymbol{\alpha})$ gives a consistent estimator.

Asymptotic normality in the case of no estimation error follows from a standard Taylor-expansion. There exists an $\boldsymbol{\alpha}_n^*$ lying between $\hat{\boldsymbol{\alpha}}_n$ and $\boldsymbol{\alpha}^0$ such that

$$\mathbf{0} = \mathbf{S}_n(\hat{\boldsymbol{\alpha}}_n) = \mathbf{S}_n(\boldsymbol{\alpha}^0) - \mathbf{H}_n(\boldsymbol{\alpha}_n^*)(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0) \tag{4.27}$$

where

$$\begin{aligned}
\mathbf{S}_n(\boldsymbol{\alpha}) &= \frac{1}{n}\sum_{t=1}^n \frac{\partial l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\
\mathbf{H}_n(\boldsymbol{\alpha}) &= -\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial^2 \boldsymbol{\alpha}}.
\end{aligned}$$

It can be proven easily that $\sqrt{n}\mathbf{S}_n(\boldsymbol{\alpha}^0) \to_d N\left(\mathbf{0}, \mathbf{V}\left(\boldsymbol{\alpha}^0\right)\right)$ and $\mathbf{H}_n(\boldsymbol{\alpha}_n^*) \to \mathbf{H}(\boldsymbol{\alpha}^0)$ a.s. (for details see Kristensen and Rahbek, 2005), hence

$$\sqrt{n}\left(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0\right) = \mathbf{H}_n(\boldsymbol{\alpha}_n^*)^{-1}\sqrt{n}\mathbf{S}_n(\boldsymbol{\alpha}^0) \to_d N\left(\mathbf{0}, \mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\mathbf{V}\left(\boldsymbol{\alpha}^0\right)\mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\right),$$

which is just (4.23).

[4] The proofs do not change by relaxing the assumptions this way, which has also been confirmed by private communication with Dennis Kristensen.

To deal with the case when estimation error is present in the ARMA-parameters, the following lemma is needed.

Lemma 4.8. *If $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}^0$,*

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{\varepsilon}_t^2 - \varepsilon_t^2) \to 0 \quad \text{a.s.} \tag{4.28}$$

If, moreover, $\hat{\boldsymbol{\theta}}_n$ is the least squares estimator of $\boldsymbol{\theta}^0$, then norming by $1/\sqrt{n}$ is sufficient:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\varepsilon}_t^2 - \varepsilon_t^2) \to 0 \quad \text{a.s.} \tag{4.29}$$

Proof. It is easy to show that $Q_n(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ (in fact, it is differentiable, see e.g. the lemmas in Francq and Zakoian, 1998) so the first statement follows. To prove the second statement, a Taylor-expansion is used:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\hat{\varepsilon}_t^2 - \varepsilon_t^2\right) = \sqrt{n}\left[Q_n\left(\hat{\boldsymbol{\theta}}_n\right) - Q_n\left(\boldsymbol{\theta}^0\right)\right] = \left(\frac{\partial Q_n\left(\boldsymbol{\theta}_n^*\right)}{\partial \boldsymbol{\theta}}\right)^T \sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\right)$$

where $\boldsymbol{\theta}_n^*$ lies between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}^0$. Using the facts that $\partial Q_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is continuous (see again the lemmas in Francq and Zakoian, 1998), $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^0$ almost surely and by the definition of the least squares estimator $\partial Q_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta} = \mathbf{0}$, we obtain that $\partial Q_n(\boldsymbol{\theta}_n^*)/\partial\boldsymbol{\theta} \to 0$ almost surely. As $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ is asymptotically normally distributed, the statement follows. □

Returning to the proof of Theorem 4.7 we can compare the log-likelihood $\hat{L}_n(\boldsymbol{\alpha})$ calculated from the $\hat{\varepsilon}_t$ series and the log-likelihood $L_n(\boldsymbol{\alpha})$ calculated from the true ε_t series:

$$\hat{L}_n(\boldsymbol{\alpha}) - L_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2 - \hat{\varepsilon}_t^2}{\sigma^2(X_{t-1},\boldsymbol{\alpha})}.$$

By (4.28) this almost surely converges to zero uniformly on $\mathbf{K}$. As the maximisation of $L_n(\boldsymbol{\alpha})$ leads to a consistent estimator and $L_n(\boldsymbol{\alpha}) \to L(\boldsymbol{\alpha})$ and $L(\boldsymbol{\alpha})$ obtains its maximum at the true parameter value, the maximisation of $\hat{L}_n(\boldsymbol{\alpha})$ also provides a consistent estimator of $\boldsymbol{\alpha}^0$.

To prove asymptotic normality when there is estimation error in the ARMA-parameters let us introduce the versions of $\mathbf{S}_n$ and $\mathbf{H}_n$ adapted to this case:

$$\begin{aligned}\hat{\mathbf{S}}_n(\boldsymbol{\alpha}) &= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l(\hat{\varepsilon}_t, X_{t-1},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\\ \hat{\mathbf{H}}_n(\boldsymbol{\alpha}) &= -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l(\hat{\varepsilon}_t, X_{t-1},\boldsymbol{\alpha})}{\partial^2\boldsymbol{\alpha}}.\end{aligned}$$

Equations (4.21) and (4.29) imply that $n^{1/2}[\hat{\mathbf{S}}_n(\boldsymbol{\alpha}) - \mathbf{S}_n(\boldsymbol{\alpha})] \to \mathbf{0}$ and $\hat{\mathbf{H}}_n(\boldsymbol{\alpha}) - \mathbf{H}_n(\boldsymbol{\alpha}) \to \mathbf{0}$ almost surely uniformly on $\mathbf{K}$. Thus all arguments regarding the above Taylor-expansion (4.27) remain valid and so asymptotic normality of the parameters in the variance equation holds in this case, too. □

4.6 Application to water discharge data

To our knowledge, no attempt had been made to model the nonlinear structure of river flow series by ARCH-type processes before the article of Elek and Márkus (2004). Since then, the idea has received attention in the hydrological literature and has been incorporated into other models as well. For instance, motivated by our research, Szilágyi et al. (2006) introduce an ARCH-type specification of the innovation in their detailed regime switching river flow model.

In Elek and Márkus (2004) we developed a FARIMA model driven by a GARCH-type innovation ϵ_t defined as

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \left(1 - \exp\left(-s\epsilon_{t-1}\right)\right) + \beta\sigma_{t-1}^2. \end{aligned}$$

Thus the equation approximately yields a standard FARIMA-GARCH model if ϵ_{t-1} is close to zero: $\sigma_t^2 \approx \alpha_0 + \alpha_1 s \epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ but the conditional variance is roughly an autoregression when $|\epsilon_{t-1}|$ is large: $\sigma_t^2 \approx \alpha_0 + \alpha_1 + \beta\sigma_{t-1}^2$. Therefore the model is lighter-tailed than a standard GARCH model.

The model gives reasonable results in hydrological simulations but the performance of the newer model presented in Elek and Márkus (2008) is superior to it. Therefore, in the following we present only the hydrological application of the latter, ARMA-ARCH-type model and do not go into the details of the previous specification.

4.6.1 Model identification and estimation

The starting point of the analysis is to fit ARMA processes to the series as described in section 3.1.1. It turns out that $p = 3$ (or $p = 2$) and $q = 1$ are appropriate at all sites to model the linear dependence structure. It remains to determine m and β.

We use a simplified version of the nonparametric procedure described in Bühlmann and McNeil (2002) to identify the parametric form of the conditional variance function $\sigma^2(x)$. The discharges (the X_t values) are grouped into 50 groups according to their magnitude and the fitted innovations of the ARMA model (the $\hat{\varepsilon}_t$-s) are classified based on the group

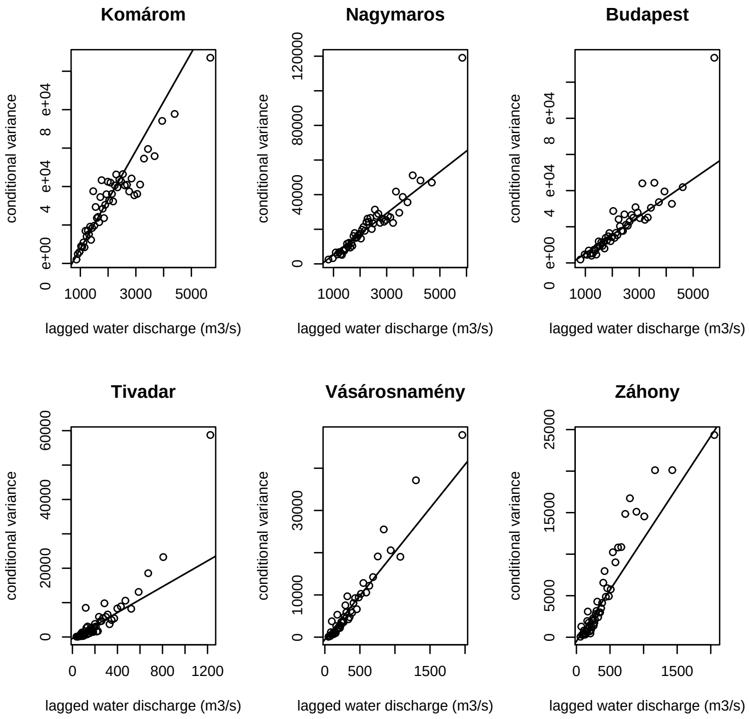


Figure 4.2: Conditional variance as a function of lagged water discharge at six stations, the lines showing the fitted relationships with maximum likelihood estimates α_0 and α_{1+}

of X_{t-1}. Thereafter, the variance of the fitted innovations and the mean of the discharges in each group are calculated and plotted against each other.

Figure 4.2 displays these plots (the estimated conditional variance as a function of the lagged water discharge) for the three monitoring stations at river Danube (upper row) and for the three monitoring stations at river Tisza (lower row). Apart from one outlier at Nagymaros, Budapest and Tivadar respectively,[5] the empirical relationship is close to linear for all stations. This fact makes the choice of the simplest model for $\sigma^2(x)$, the linear, sensible, which leads to the $\beta = 1/2$ parameter choice. Note that there would be no apparent pattern in the figures if a strong ARMA model were appropriate.

Since the empirical variance function does not have a negative slope anywhere, $\alpha_{1-} = 0$ is also a reasonable parameter restriction and the break point m should lie very close to (or perhaps even lower than) the observed minimum of the water discharge series. After a pilot

[5] At Nagymaros and Budapest the outlier is caused by the spring flood of 1940, which corresponds to the largest calculated water discharge at Budapest and to the second largest at Nagymaros in the examined period. Although we did not find a convincing reason to leave out these values, it should be noted that estimation of large water discharges from level data contains more uncertainty for the distant than for the recent past. At Tivadar more than one flood events contribute to the highest point in the discharge-variance plot.

study, we have fixed m at the minimum at each site. Different choices had no substantial effect on the properties of the simulated series from the models as long as m was not chosen too large, although the estimate of α_0 obviously varied with different values of m. Note also that different models with m lower than the minimum discharge value cannot be distinguished from each other since the conditional variance applies for the whole observed series in these cases. In practice, however, this is not a crucial problem as they produce only slightly different simulation outputs.

Figure 4.2 also displays the fitted lines $y = \alpha_0 + \alpha_{1+}(x - m)$ with QML estimates α_0 and α_{1+} (see below) at each site. Apart from Tivadar, the lines are generally close to the estimated variance points, supporting the decision for the form of $\sigma^2(x)$.

After the above parameter restrictions, Table 4.1 shows the QML estimates of the α_0 and α_{1+} parameters (along with the fixed value of m) for the six stations.[6] Asymptotic standard errors from Theorem 4.7 are provided in parentheses. The α_{1+} parameter is significant at all reasonable significance levels at all stations, indicating the presence of the ARCH-effect in the river flow series.

Table 4.1: Parameter estimates with standard errors in parentheses (measurement unit is m^3/s)

Monitoring station	m	α_0	α_{1+}
Komárom	789	1807.8 (2009.8)	26.06 (2.22)
Nagymaros	586	544.7 (154.1)	11.95 (0.57)
Budapest	580	907.1 (314.0)	10.29 (0.55)
Tivadar	23	24.49 (5.95)	18.80 (1.13)
Vásárosnamény	30	82.45 (32.82)	20.71 (0.51)
Záhony	45	67.04 (17.31)	12.37 (1.19)

Based on the estimated parameters, the fitted noise sequence $(\hat{Z}_t = \hat{\varepsilon}_t / \sigma(X_{t-1}))$ can be calculated easily. Figure 4.3 shows its probability density, its autocorrelation function as well as the autocorrelation function of its square and of its absolute value at Nagymaros. Similarly to the innovations of the ARMA model (Figure 3.2), the noises are highly non-Gaussian and highly peaked. However, they are much closer to independence than the innovations as not only themselves, but also their squares can be regarded as uncorrelated (with some remaining autocorrelation in the absolute valued sequence).

[6] Strictly speaking Assumption 4.6 does not hold in this case because $\alpha_{1-} = 0$. However, Theorem 4.7 remains valid if α_{1-} is fixed and $(\alpha_0, \alpha_{1+}) \in \mathrm{int}(\mathbf{K})$ where $\mathbf{K}$ is a compact subset of $\mathbf{R}_{++} \times \mathbf{R}_{+}$.

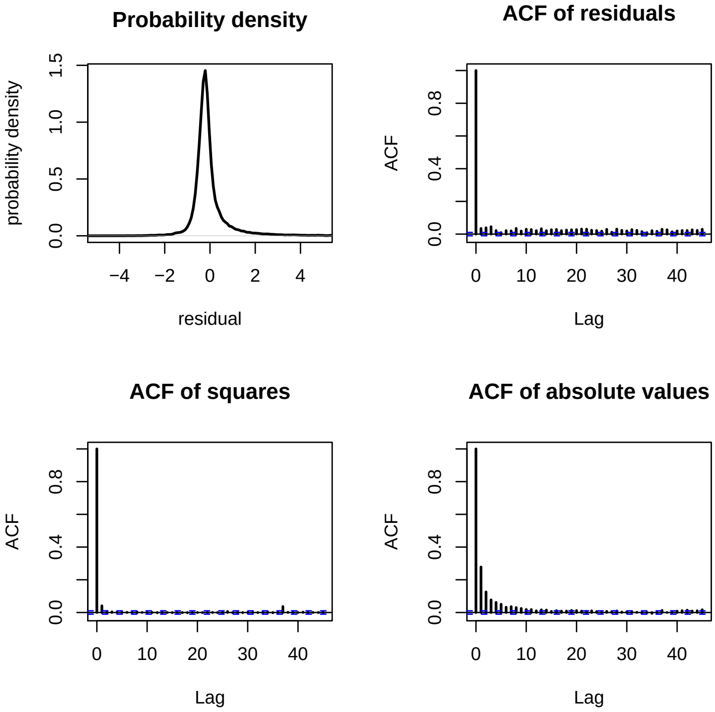


Figure 4.3: Probability density of the noise, autocorrelation function of the noise, of its square and of its absolute value at Nagymaros

4.6.2 Simulations

Simulation of synthetic water discharge series from the fitted model goes as follows. First, the same seasonal resampling procedure that was used in section 3.1.1 to generate the ε_t innovations is used to simulate the independent-valued noise sequence Z_t. (The simulated values in a given month are drawn from the set of empirical values in the same month.) Since the sample is substantially large (15000-36000 observations at each site), this method provides a reasonable approximation to the distribution of the noise, at least at not very large quantiles. Having generated Z_t, the synthetic water discharge series are then simulated by applying the estimated nonlinear ARCH-filter and the linear filter, and finally adding back the seasonal component c_t which was estimated with a local polynomial smoothing (LOESS) procedure. To tackle the substantial parameter uncertainty in the variance equation, the parameters α_0 and α_{1+} in the ARCH-filter are drawn from a bivariate normal distribution with mean vector and covariance matrix estimated from Theorem 4.7.

Figure 4.4 shows the goodness of fit of the model in terms of approximating the probability density of the observed discharge series. (The simulated probability density is cal-

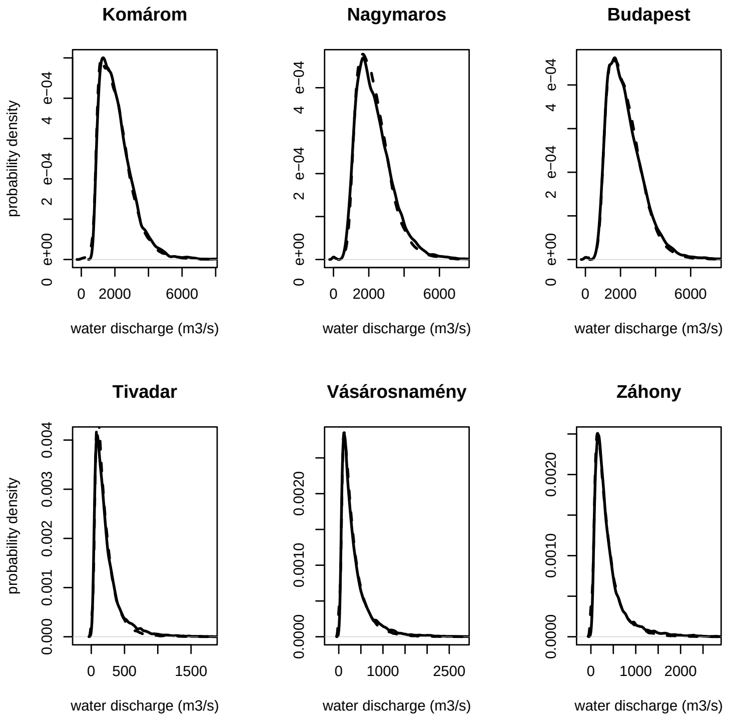


Figure 4.4: Probability density of the empirical (continuous) and simulated (dashed) series at the monitoring stations. The almost perfect fit makes the lines almost indistinguishable.

culated from a 1000-year long synthetic series at each site.) Figure 4.5 compares various high quantiles of the empirical discharge series to the corresponding quantities of synthetic series of the same length. The quantiles fit generally well at sites of river Danube (upper row), while some quantiles are slightly underestimated at sites of river Tisza. However, the fit of the probability density is adequate at the latter sites, too, and the model performs altogether much better than the ARMA model (Figure 3.3).

Let us compare specifically the extremal behaviours, i.e. the tails and the extremal clustering tendencies of the simulated and original series. Figure 4.6 shows the histogram of the estimated shape, scale parameters and the expected values of the GPDs fitted to the upper tails of 500 simulated series at Nagymaros, the vertical lines indicating the results for the observed series. (The threshold is equal to 4300 m^3/s and a declustering period of 15 days is applied. The lengths of the observed and simulated series are the same.) The median and mean of the simulated shape parameters are a bit higher than zero (0.04), which indicates that the simulated series is a bit heavier tailed at this threshold than the exponential distribution. This fact is not surprising for the following reason. Since the fitted ARMA-β-TARCH model exhibits strong linear dependence (with first order auto-

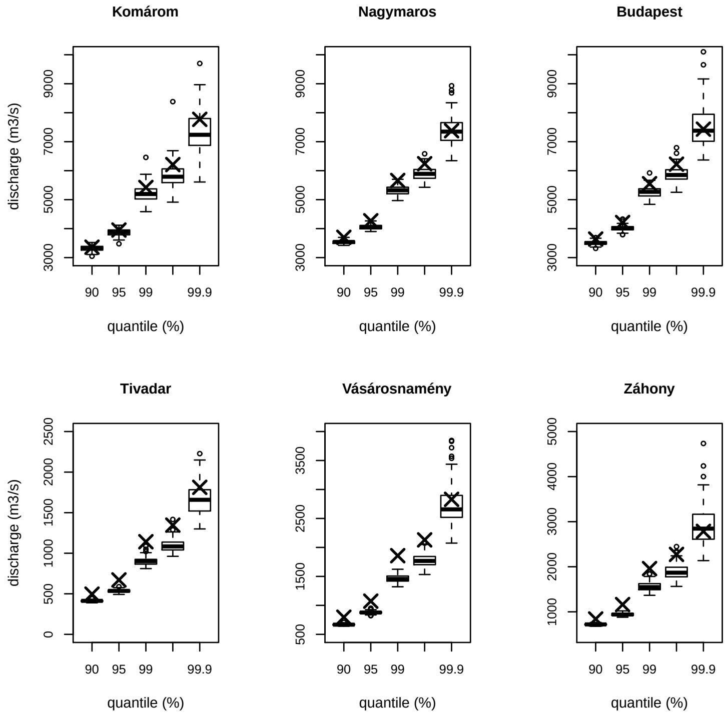


Figure 4.5: High quantiles of empirical (crosses) and simulated (boxplots) series at the monitoring stations. The boxplots of quantiles are constructed from 100 simulated series of the same length as the original ones. The boxes show the middle 50% of the distribution.

correlation around 0.95), it is at least as heavy tailed than the corresponding β-TARCH model without ARMA terms, and Proposition 4.6 yields that the latter is heavier tailed than any Weibull distribution with exponent larger than $\gamma\,(1-\beta)$. A careful examination of the fitted noise sequence reveals that Z_t is heavier tailed than the normal distribution, hence $\gamma < 2$ and thus $\gamma\,(1-\beta) < 1,$ yielding positive shape parameter estimates for finite thresholds.

Although the observed shape parameter of the original series (-0.10) at threshold 4300 m^3/s is below the mean of the simulated shape parameters (0.04), it is still within the range acceptable for model fit since around 10% of the simulated shape parameters lie below -0.10. At the same time, the simulated scale parameters are generally lower than the observed one, with only around 5% of them exceeding it. As a result of these two effects, the expected values of the GPDs fitted to the simulated exceedances approximate well the corresponding quantity of the observed threshold exceedances. (This quantity estimates the average height of an exceedance above the chosen threshold and is also an important hydrological parameter.)

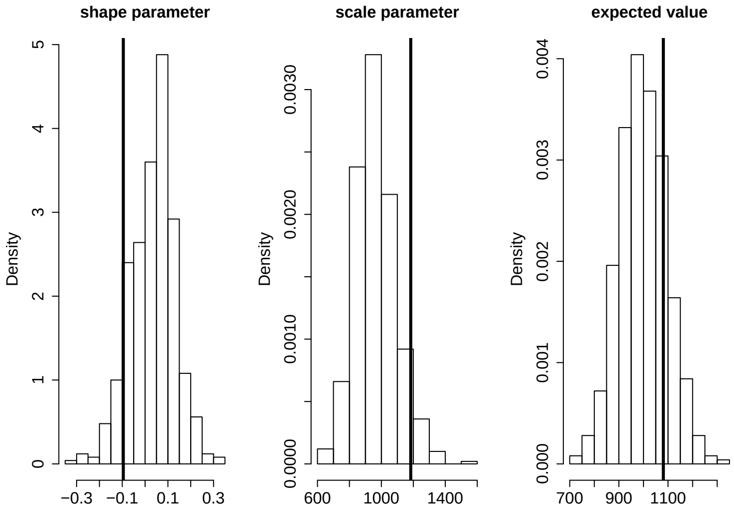


Figure 4.6: Histograms of shape and scale parameters and expected values of the GPDs fitted to the upper tails of 500 simulated series at Nagymaros. Vertical lines indicate the estimates for the observed series, the threshold is chosen as 4300 m^3/s.

Figure 4.7 displays the extremal index estimates using the Ferro-Segers method for the observed and simulated series at two selected sites. Threshold values are determined as various quantiles (from 90% to 99.9%) of the observed water discharge distribution. Estimates for the observed series lie in the range of the estimates for simulated series except for the highest quantiles (99.9% quantile for Nagymaros and 99.5% and 99.9% quantiles for Tivadar). The fact that the estimated extremal index is small for the simulated data does not contradict the conjecture of section 4.4 about a unit extremal index because the large first order autocorrelations already lead to strong clustering at subasymptotic levels, which is slightly further increased by the seasonal component c_t.

It should also be noted that other extremal cluster characteristics such as the distribution of aggregate excesses or the seasonality of high-level exceedances are reasonably well captured by the model, too.

4.7 Summary

In this chapter we examined the probabilistic properties, extreme value features and statistical estimation of the ARMA-β-TARCH model and fitted it to water discharge data. The model approximates the important extremal features of river flow series reasonably well but tends to possess slightly heavier tail and stronger high-level clustering above the

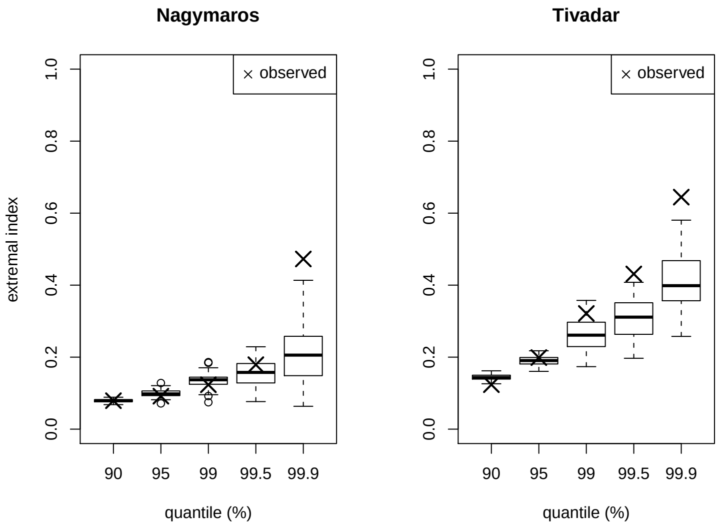


Figure 4.7: Extremal index estimates for the original and simulated series at Nagymaros and Tivadar. The boxplots are constructed from 50 simulated series of the same length as the original ones. The boxes show the middle 50% of the distribution.

thresholds of practical interest than the observed data. Moreover, although the model specification is straightforward and statistically appealing, it is not easy to explain – apart from the feedback argument at the beginning of the section – what physical reasons lead to the exact conditional variance specification.

To the contrary, the next chapter will deal with more physically motivated models, which have asymptotically exponential tail and are more able to describe certain empirical features (such as the pulsatile nature) of the river flow series than the conditionally heteroscedastic model.

Chapter 5

Markov-switching autoregressive models

5.1 The model and basic results

Markov-switching autoregressive (MS-AR) processes are governed by a latent Markov chain with a finite number of states (which are called regimes), and they behave as an autoregressive model in each regime. Because of their easy interpretability and flexible structure, they are widely used in economics (see the sequence of articles originating from Hamilton (1990)), engineering or hydrology (Lu and Berliner (1999), Szilágyi et al. (2006), Vasas et al. (2007)).

In this chapter we first explore the tail behaviour and extremal clustering of a class of two-state Markov-switching autoregressive processes. Then we show that the extremal behaviour of certain more general Markov-switching models can be approximated by these MS-AR processes. This gives the idea to develop a method of fitting MS-AR models when only high level exceedances of a process are known. Finally, we use this method to analyse extremal clustering behaviour of the river flow series.

Define the X_t process as a two-state Markov-switching AR(1) model:

$$X_t = a_1 X_{t-1} + \varepsilon_{1,t} \quad \text{if} \quad I_t = 1, \tag{5.1}$$

$$X_t = a_0 X_{t-1} + \varepsilon_{0,t} \quad \text{if} \quad I_t = 0, \tag{5.2}$$

where I_t is a two-state Markov chain with transition probabilities

$$p_1 = P\left(I_t = 0 | I_{t-1} = 1\right), \tag{5.3}$$

$$p_0 = P\left(I_t = 1 | I_{t-1} = 0\right). \tag{5.4}$$

We assume that $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{0,t}\}$ are both independent, identically distributed noise sequences (but the two distributions need not be the same), independent from each other and from the $\{I_t\}$ sequence as well.

The Markov structure of I_t implies that regime durations are independent geometrically distributed random variables: the duration of staying in regime j is distributed as Geom(p_j) $(j = 0, 1)$. The model can be written as a random coefficient autoregression $X_t = A_t X_{t-1} + B_t$ where $A_t = \sum_{j=0}^{1} \chi_{\{I_t=j\}} a_j$ and $B_t = \sum_{j=0}^{1} \chi_{\{I_t=j\}} \varepsilon_{j,t}$. According to Brandt (1986) such a model has a unique stationary solution if

$$E\left(\log|A_t|\right) < 0 \quad \text{and} \quad E\left(|B_t|\right) < \infty.$$

Here the second condition is automatically satisfied for a very wide range of distributions (in fact, Assumptions 5.2-5.3 defined below are much stronger), and the first condition holds in the case of the MS-AR model if

$$p_1 \log|a_0| + p_0 \log|a_1| < 0. \tag{5.5}$$

Hence, local stationarity (i.e. that $|a_0| < 1$ and $|a_1| < 1$) is a sufficient but not necessary condition for the "global" stationarity of the Markov-switching autoregressive model. Further probabilistic properties of MS-AR models are given e.g. in Yao and Attali (2000).

Let us assume that $|a_1| \geq |a_0|$. The extremal properties of the model depend substantially on the stability of the dynamics in the particular regimes. If in both regimes the parameters lie within the open interval (-1,1) (i.e. $|a_0| \leq |a_1| < 1$) and the generating noise is light-tailed, the precise form of the stationary distribution depends very much on the generating noise, but it is certainly light-tailed and there is no extremal clustering. On the other hand, if in one of the regimes the AR(1)-parameter is greater than one ($|a_1| > 1$) but (5.5) still holds, it is easy to show that not all moments of X_t exist even when $\varepsilon_{1,t}$ is sufficiently light-tailed, e.g. normally distributed. In fact, under some additional conditions, it follows from Saporta (2005) that for $|a_1| > 1$, there exist a $K > 0$ and $\lambda > 0$ (which certainly depend on a_1, a_0 and the distribution of the noises) such that $P(X_t > u) \sim K u^{-\lambda}$. Hence, the tail is typically regularly varying in this case. There remains the case when one of the parameters is exactly one, i.e. when the process behaves as a random walk in one regime. This parameter choice – whose extremal properties have not been studied yet in the literature – is qualitatively different from the above mentioned ones thus our main focus is the following:

Assumption 5.1. $a_1 = 1$ *and* $0 \leq a_0 < 1$.

This assumption implies that the process behaves like a random walk in the first regime, while it is a stationary autoregression in the second one. (The $0 \leq a_0 < 1$ assumption can be weakened, see Remark 5.16.) The stationary solution for X_t always exists by (5.5), and unless otherwise indicated, all probability statements in the sequel correspond to this unique stationary distribution. As far as extremes are concerned this parameter choice is on the border of the two previously mentioned cases because – as it will turn out in sections 5.2 and 5.3 – if the noise is light-tailed the stationary solution is light-tailed (as when $|a_1| < 1$), but there is asymptotic clustering of high values (as when $a_1 > 1$).

Throughout the chapter, the following assumptions on the noise sequences will be applied. Assumption 5.2 essentially states that the upper tail of $\varepsilon_{1,t}$ is light (excluding e.g. subexponential noises), while Assumption 5.3 implies that $\varepsilon_{0,t}$ is not much heavier tailed than the positive part of $\varepsilon_{1,t}$. As usual in the dissertation, $L_X(s) = E(\exp(sX))$ denotes the moment generating function and $c_X(s) = \log L_X(s)$ the cumulant generating function.

Assumption 5.2. *The distribution of $\varepsilon_{1,t}$ is absolutely continuous w.r.t. the Lebesgue-measure and $E\varepsilon_{1,t}^2 < \infty$. Moreover, there exists a $\kappa > 0$ such that*

$$(1 - p_1) L_{\varepsilon_{1,t}}(\kappa) = 1, \tag{5.6}$$

and $L'_{\varepsilon_{1,t}}(\kappa) < \infty$.

Assumption 5.3. *The distribution of $\varepsilon_{0,t}$ is absolutely continuous w.r.t. the Lebesgue-measure and its support is the whole real line. With κ defined in* (5.6)*, there exists an $s_0 > \kappa$ such that $L_{|\varepsilon_{0,t}|}(s_0) < \infty$.*

Assumption 5.2 is satisfied for a wide range of distributions, Examples 5.1–5.3 state a few practically important ones. (Assumption 5.3 is even weaker.) Example 5.3 is particularly interesting because it relates to the river flow model of Vasas et al. (2007), which will be discussed in section 5.6. The derivative condition of Assumption 5.2 is satisfied in all the cases below.

Example 5.1. *If $\varepsilon_{1,t}$ is normally distributed with mean μ and variance σ^2,*

$$\kappa = \frac{\left(\mu^2 - 2\sigma^2 \log(1 - p_1)\right)^{1/2} - \mu}{\sigma^2}$$

is the positive solution of equation (5.6).

Example 5.2. *Let $\varepsilon_{1,t}$ be distributed as skewed Laplace with parameters c, λ_L and λ_U, i.e. have probability density*

$$\begin{aligned} f(x) &= \tfrac{\lambda_L \lambda_U}{\lambda_L + \lambda_U} \exp(\lambda_L (x - c)) \quad \text{if} \quad x < c, \\ f(x) &= \tfrac{\lambda_L \lambda_U}{\lambda_L + \lambda_U} \exp(-\lambda_U (x - c)) \quad \text{if} \quad x \geq c. \end{aligned}$$

Then, κ is the solution of the equation

$$(1 - p_1) \exp(c\kappa) = \frac{(\kappa + \lambda_L)(\lambda_U - \kappa)}{\lambda_L \lambda_U}. \tag{5.7}$$

If $c = 0$ and $\lambda_L = \lambda_U = \lambda$, then $\kappa = p_1^{1/2} \lambda$.

Example 5.3. *Let $\varepsilon_{1,t}$ be distributed as $\Gamma\left(\alpha,\lambda\right)$, then $\kappa=\lambda\left(1-\left(1-p_1\right)^{1/\alpha}\right)$.*

Before showing the importance of Assumption 5.2, the following simple observations are given, which will be used throughout the chapter.

Lemma 5.4. *Let Q_1 be a random variable with $P\left(Q_1>u\right)\sim K_1\exp\left(-\kappa u\right)$, and let Q_2 be an independent variable with $L_{Q_2}\left(s\right)<\infty$ for an $s>\kappa$. Then,*

$$P\left(Q_1+Q_2>u\right)\sim K_1L_{Q_2}\left(\kappa\right)\exp\left(-\kappa u\right).$$

Proof. According to Breiman's theorem (Breiman, 1965), if X and Y are two independent nonnegative random variables such that the tail of X is regularly varying with index $-\delta$ $(\delta>0)$ and $E\left(Y^{\delta+\eta}\right)<\infty$ for some $\eta>0$, then $P\left(XY>v\right)\sim E\left(Y^{\delta}\right)P\left(X>v\right)$ as $v\rightarrow\infty$. Thus the statement follows with the choice $X=\exp\left(Q_1\right),Y=\exp\left(Q_2\right),\delta=\kappa$ and $\eta=s-\kappa>0$. □

Lemma 5.5. *Let Q_1 and Q_2 be two independent random variables with tails $P\left(Q_i>u\right)\sim K_i\exp\left(-\kappa u\right)$ $(i=1,2)$ and let $|a|<1$. Then for every fixed $v\geq 0$, as $u\rightarrow\infty$,*

$$P\left(aQ_1+Q_2>u+v|Q_1>u\right)\sim K_2\left(1-a\right)^{-1}\exp\left(-\left(1-a\right)\kappa u\right)\exp\left(-\kappa v\right).$$

Proof. For $0<a<1$, let $\eta=\left(2-a+1/a\right)/2$ and for $-1<a\leq 0$, let $\eta>2-a$ arbitrary. Then $\eta>1$ and $a\eta<1$, hence for $u\rightarrow\infty$

$$\begin{aligned}P\left(aQ_1+Q_2>u+v,u<Q_1\leq\eta u\right) &= -\int_u^{\eta u}\bar{F}_{Q_2}\left(u+v-ax\right)d\bar{F}_{Q_1}\left(x\right)\\ &\sim -\int_u^{\eta u}K_2\exp\left(-\kappa\left(u+v-ax\right)\right)d\bar{F}_{Q_1}\left(x\right).\end{aligned}$$

By partial integration, this is equal to

$$\begin{aligned}&= -\left[K_2\exp\left(-\kappa\left(u+v-ax\right)\right)\bar{F}_{Q_1}\left(x\right)\right]_{x=u}^{x=\eta u}\\ &\quad+\int_u^{\eta u}K_2\kappa a\exp\left(-\kappa\left(u+v-ax\right)\right)\bar{F}_{Q_1}\left(x\right)dx\\ &\sim -\left[K_2\exp\left(-\kappa\left(u+v-ax\right)\right)K_1\exp\left(-\kappa x\right)\right]_{x=u}^{x=\eta u}\\ &\quad+\int_u^{\eta u}K_2\kappa a\exp\left(-\kappa\left(u+v-ax\right)\right)K_1\exp\left(-\kappa x\right)dx\\ &= -\left[K_2K_1\left(1-a\right)^{-1}\exp\left(-\kappa\left(u+v+\left(1-a\right)x\right)\right)\right]_{x=u}^{x=\eta u}\\ &\sim K_2K_1\left(1-a\right)^{-1}\exp\left(-\kappa\left(u+v+\left(1-a\right)u\right)\right)\\ &\sim P\left(Q_1>u\right)K_2\left(1-a\right)^{-1}\exp\left(-\kappa\left(1-a\right)u\right)\exp\left(-\kappa v\right).\end{aligned}$$

Moreover, as $u\rightarrow\infty$, $P\left(Q_1>\eta u\right)\sim P\left(Q_1>u\right)\exp\left(-\left(\eta-1\right)\kappa u\right)$ is negligible compared to the above term for every fixed $v\geq 0$ because $\eta-1>1-a$. Hence the statement follows. □

The importance of Assumption 5.2 comes from its connection with the theory of random walks, which we will explore in a slightly unusual setting. Let

$$S_0 = 0, \qquad S_n = S_{n-1} + \varepsilon_n \quad (n = 1, 2, \dots) \tag{5.8}$$

be a random walk and define $\tau_u = \min\{n : S_n > u\}$ the crossing time of level u and $B_u = (S_{\tau_u} - u)$ on $(\tau_u < \infty)$ the overshoot of u (when it exists). First, the following well-known result holds.

Lemma 5.6. *If $E\varepsilon_n \geq 0$, $P(\tau_u < \infty) = 1$ for all $u \geq 0$. Furthermore, let us assume that the distribution is non-lattice.*[1] *Under this condition if $0 < E\varepsilon_n < \infty$, or $E\varepsilon_n = 0$* and *$E\varepsilon_n^2 < \infty$ then $E(B_0) < \infty$ and $B_u \to_d B_\infty$ as $u \to \infty$, where B_∞ has probability density function*

$$f_{B_\infty}(y) = \bar{F}_{B_0}(y) / E(B_0) \tag{5.9}$$

for $y > 0$.

Proof. By Asmussen (1987, Chapter VII., Thm. 2.4.) $E\varepsilon_n \geq 0$ implies that $P(\tau_0 < \infty) = 1$, hence $P(\tau_u < \infty) = 1$ holds as a consequence of Asmussen (1987, Chapter VII., Thm. 2.1.) so the first part is proven.

Spitzer (1960, Thm. 3.4.) states that $0 < E\varepsilon_n < \infty$ implies $E(B_0) < \infty$, and the same is true if $E\varepsilon_n = 0$ and $E\varepsilon_n^2 < \infty$ together hold. According to Asmussen (1987, Chapter VII., Thm. 2.1.) $P(\tau_0 < \infty) = 1$, $E(B_0) < \infty$ and the non-latticeness of the distribution of $\{\varepsilon_n\}$ together yield that $B_u \to_d B_\infty$ and (5.9) holds. □

Now let ε_n be distributed as $\varepsilon_{1,t}$ in definition (5.8) of the random walk. Let T be a Geom(p_1)-distributed random variable, i.e. $P(T \geq k) = (1 - p_1)^{k-1}$, independent of S_n. Define the maximum of the stopped random walk

$$M^S_{k,T-1} = \max\{S_i : k \leq i \leq T - 1\} \tag{5.10}$$

where $M^S_{k,T-1} = -\infty$ if $k > T - 1$. (We will also use the notation $M^S_{k,m}$ for the maximum of $\{S_n\}$ between time points k and m.) Then

Proposition 5.7. *Under Assumption 5.2 there exists a $K > 0$ such that*

$$P\left(M^S_{0,T-1} > u\right) \sim K \exp(-\kappa u). \tag{5.11}$$

[1] A distribution is called lattice if it concentrates on the set of points $\{a + i\lambda : i \in \mathbf{Z}\}$ with some a and λ.

Proof. The statement is very similar to the Cramér-Lundberg (C-L) approximation for random walks, which gives the tail behaviour of the distribution of the maximum of a random walk with negative drift if the increments are light-tailed (Asmussen, 1987, Chapter XII., Thms 5.2. and 5.3.). If $S'_0 = 0$, $S'_n = S'_{n-1} + \varepsilon'_n$ $(n = 1, 2, \dots)$, $E\varepsilon'_n < 0$, ε'_n has a non-lattice distribution and there exists a κ' such that $L_{\varepsilon'_n}(\kappa') = 1$ and $L'_{\varepsilon'_n}(\kappa') < \infty$, then the C-L approximation states that for $M' = \max\{S'_i : 0 \leq i\}$ (which exists a.s.) $P(M' > u) \sim K \exp(-\kappa' u)$ with some $K > 0$.

Heuristically, the random walk in our lemma has a defective step distribution. Let $\varepsilon'_n = -\infty$ with probability p_1 and be equal to ε_n with probability $1 - p_1$. Then $M^S_{0,T-1} = M'$, and the exponent κ of its tail comes from the equation

$$1 = L_{\varepsilon'_n}(\kappa) = (1 - p_1) L_{\varepsilon_{1,t}}(\kappa),$$

which is just the statement of the lemma.

The formal proof goes similarly to the proof of the original C-L approximation. Let F_0 be the distribution function of the i.i.d. random variables $\{\varepsilon_n\}$ and define $\Theta = \{s : L_{\varepsilon_n}(s) < \infty\}$. For an $s \in \Theta$

$$dF_s(x) = \exp(sx - c(s)) \, dF_0(x)$$

is a proper distribution with the cumulant generating function $c(s) = \log L_{\varepsilon_n}(s)$, and $\{F_s\}$ is called the conjugate family of distributions. Now, for a random walk $\{S_n\}$ with increments $\{\varepsilon_n\}$, we can examine the corresponding events not only under the assumption that ε_n is distributed as F_0 but for any other distribution F_s from the conjugate family. The resulting probabilities and expectations will be denoted by P_s and E_s, respectively. Then $P = P_0$ and $E = E_0$.

Let τ be a stopping time of the random walk $\{S_n\}$ and let $G \in \mathcal{F}_\tau$, $G \subset \{\tau < \infty\}$. Then a form of Wald's fundamental identity (Asmussen, 1987, Chapter XII., Thm. 4.1.) states that for any $s \in \Theta$

$$P(G) = E_s \left[\exp(-sS_\tau - \tau c(s)) \chi_{\{G\}}\right].$$

In our case let $s = \kappa$ as defined in Assumption 5.2, let $\tau = \tau_u$ be the crossing time of level u and $G_n = \{\tau = n\}$. (Also, let $\{\varepsilon_n\}$ be distributed as $\{\varepsilon_{1,t}\}$.) Since $\log(1 - p_1) + c(\kappa) = 0$ and $S_{\tau_u} = u + B_u$ on $(\tau_u < \infty)$, the formula gives

$$P(\tau_u = n) = \exp(-\kappa u)(1 - p_1)^{-n} E_\kappa [\exp(-\kappa B_u) | \tau_u = n] P_\kappa(\tau_u = n).$$

However, $P\left(M^S_{0,T-1} > u\right) = P(\tau_u < T)$ for the independent geometrically distributed variable T, hence summing up the previous formula with the $P(T > n)$ probabilities we obtain

$$\begin{aligned} P\left(\tau_u < T\right) &= \exp\left(-\kappa u\right) \sum_{n=0}^{\infty} P\left(T > n\right) \left(1 - p_1\right)^{-n} E_\kappa \left[\exp\left(-\kappa B_u\right) |\tau_u = n\right] P_\kappa \left(\tau_u = n\right) \\ &= \exp\left(-\kappa u\right) E_\kappa \left[\exp\left(-\kappa B_u\right) |\tau_u < \infty\right] P_\kappa \left(\tau < \infty\right). \end{aligned} \tag{5.12}$$

Since $c\left(s\right)$ is convex, $c(0) = 0$ and $c\left(\kappa\right) = -\log\left(1 - p_1\right) > 0$, we obtain that $c'\left(\kappa\right) > 0$ (which exists by Assumption 5.2). However, $E\left(\varepsilon_{1,t}\right) = c'(0)$ from a well-known property of the cumulant generating function, and similarly $E_s\left(\varepsilon_{1,t}\right) = c'(s)$ if the derivative exists. Thus our random walk has a positive drift under the F_κ distribution hence $P_\kappa\left(\tau_u < \infty\right) = 1$ and $B_u \to_d B_\infty$ under the P_κ measure by Lemma 5.6. This last consequence also provides that

$$E_\kappa \left[\exp\left(-\kappa B_u\right)\right] \to E_\kappa \left[\exp\left(-\kappa B_\infty\right)\right] = K > 0$$

as $u \to \infty$ so (5.12) yields

$$P\left(M^S_{0,T-1} > u\right) = P\left(\tau_u < T\right) \to K \exp\left(-\kappa u\right).$$

□

A more important consequence from our point of view is that Assumption 5.2 also ensures that S_T, which is a geometric sum of i.i.d. variables distributed as $\{\varepsilon_{1,t}\}$ in this case, has an exponential tail.

Proposition 5.8. *Under Assumption 5.2 there exists a* $K > 0$ *such that*

$$P\left(S_T > u\right) \sim K \exp\left(-\kappa u\right).$$

Proof. Theorem 1 in Greenwood (1976) states that

$$S_{T-1} = \max\{S_n : n \leq T-1\} + \min\{S_n : n \leq T-1\}$$

if the terms on the right are added *independently*. Here the first term is just $M^S_{0,T-1}$ defined in (5.10), which has approximately Exp$\left(\kappa\right)$-tail by Proposition 5.7. As the second term in the sum is nonpositive, Lemma 5.4 yields that S_{T-1} has the same tail, too. But $S_T = S_{T-1} + \varepsilon_1$, and $L_{\varepsilon_1}\left(s\right) < \infty$ for an $s > \kappa$ by Assumption 5.2. Hence Lemma 5.4 can be applied again to obtain the tail of S_T. □

Remark 5.9. *In the special case when* $\varepsilon_{1,t} \geq 0$ *a.s., Proposition 5.8 follows easily from the renewal theorem.*

Proof. By the constant hazard property of the geometric distribution

$$\bar{F}_{S_T}(u) = \bar{F}_{\varepsilon_{1,t}}(u) + (1-p_1)\int_0^u f_{\varepsilon_{1,t}}(z)\,\bar{F}_{S_T}(u-z)\,dz.$$

Multiplying both sides by $\exp(\kappa u)$ we obtain, using the notations of the Proposition and the notation $h(u) = \exp(\kappa u)\,\bar{F}_{S_T}(u)$ that

$$\begin{aligned} h(u) &= \exp(\kappa u)\,\bar{F}_{\varepsilon_{1,t}}(u) + \int_0^u (1-p_1)\exp(\kappa z)\,f_{\varepsilon_{1,t}}(z)\,h(u-z)\,dz \\ &= \exp(\kappa u)\,\bar{F}_{\varepsilon_{1,t}}(u) + \int_0^u h(u-z)\,dF_\kappa(z) \end{aligned}$$

as a consequence of Assumption 5.2. This is a renewal equation hence

$$\lim_{u\to\infty} h(u) = \frac{\int_0^\infty \exp(\kappa z)\,\bar{F}_{\varepsilon_{1,t}}(z)\,dz}{E_\kappa(\varepsilon_{1,t})} = K > 0$$

follows from the key renewal theorem, thus the Proposition is proven in this special case. □

Remark 5.10. *The Proposition can be proven directly even in some cases when the $\varepsilon_{1,t} \geq 0$ a.s. assumption does not hold. For instance, if $\varepsilon_{1,t}$ is normally distributed, f_{S_T} is given straightforwardly as an infinite sum, whose asymptotic behaviour can be determined by Laplace's method for sums (by similar techniques as used in the proof of Lemma 5.25).*

5.2 Tail behaviour

As already noted, we examine the $a_1 = 1$ case. In the sequel the $I_t = 1$ regime will be called the "random walk" or nonstationary regime, while the $I_t = 0$ regime the stationary one. Since I_t is a Markov chain, regime durations are independent and geometrically distributed. To examine the extremal behaviour of the model, let us introduce a few auxiliary processes. Let ξ_m and ζ_m, respectively, denote the series of time points when the $I_t = 1$ and $I_t = 0$ regimes end. (The indexing is chosen to ensure that $\xi_{m-1} < \zeta_m < \xi_m < \zeta_{m+1}$.) For later reference, let $\gamma_m(u) = \min\{\zeta_m + 1 \leq t \leq \xi_m : X_t > u\}$ be the time of first reaching a threshold u in a nonstationary regime (and $\gamma_m(u)$ remains undefined if there is no such t). The notations $N_{1m} = \xi_m - \zeta_m$ and $N_{0m} = \zeta_m - \xi_{m-1}$ are used. Finally, let $B(t) = \max(\xi_m | \xi_m \leq t)$ be the end of the last nonstationary regime up to time t, and similarly let $D(t) = \max(\zeta_m | \zeta_m \leq t)$ the end of the last stationary regime.

Then $Y_m = X_{\xi_m}$ (the sequence of last values in the nonstationary regimes) is a Markov chain, and one can expect that its tail behaviour characterises the tail of X_t. Similarly,

$Z_m = X_{\zeta_m}$ (the series of last values in the stationary regimes) is a Markov chain as well. For later reference, let $M^{(m)} = \max\{X_t : \zeta_m + 1 \leq t \leq \xi_m\}$ be the maximum in a nonstationary regime.

The sequence Y_m can be written as

$$Y_m = Z_m + V_m = A_{0m} Y_{m-1} + U_m + V_m. \tag{5.13}$$

Here, $A_{0m} = a_0^{N_{0m}}$, $V_m = \sum_{t=\zeta_m+1}^{\xi_m} \varepsilon_{1,t}$ is a geometric random sum of i.i.d. variables (because the time spent in a regime is geometrically distributed), thus as a direct consequence of Proposition 5.8 there exists a $K > 0$ such that

$$P\left(V_m > u\right) \sim K \exp\left(-\kappa u\right). \tag{5.14}$$

$U_m = \sum_{t=\xi_{m-1}+1}^{\zeta_m} a_0^{\zeta_m - t} \varepsilon_{0,t}$ is a geometric random weighted sum of i.i.d. variables. If $m \neq k$, (A_{0m}, U_m, V_m) is independent of (A_{0k}, U_k, V_k) , but A_{0m} is not independent of $U_m.$ Hence, standard results on the solutions of stochastic difference equations are not directly applicable.

Based on the asymptotically exponential tail of $V_m,$ the following Theorem states that X_t also has such a tail.

Theorem 5.11. *If Assumptions 5.1–5.3 hold there is a constant $K > 0$ such that*

$$P\left(X_t > u\right) \sim K \exp\left(-\kappa u\right).$$

Proof. Let $L_0\left(s\right) = L_{|\varepsilon_{0,t}|}\left(s\right)$. By Jensen's inequality, $L_0\left(as\right) \leq \left(L_0\left(s\right)\right)^a$ for all $0 \leq a < 1.$ According to Assumption 5.3, for all $s \leq s_0$ given there,

$$\log L_{|U_m|}\left(s\right) \leq \sum_{k=0}^{\infty} \log L_0\left(|a_0|^k s\right) \leq \sum_{k=0}^{\infty} |a_0|^k \log L_0\left(s\right) < \infty. \tag{5.15}$$

To determine the tail behaviour of $X_t,$ we first use the drift condition for the stability of Markov chains (c.f. Meyn and Tweedie (1993)) to prove that for all $0 < s < \kappa$

$$L_{Y_m}\left(s\right) < \infty. \tag{5.16}$$

Clearly, Y_m is a ψ-irreducible and aperiodic Feller chain, with ψ being the Lebesgue-measure. Thus, by Meyn and Tweedie (1993, Thms 5.5.7 and 6.0.1), every compact set is small and smallness is equivalent to petiteness. (For the definition and properties of small and petite sets, see Meyn and Tweedie (1993, Chapter 5).) Following Meyn and Tweedie (1993, Thms 14.0.1 and 15.0.1), for every $0 < s < \kappa$ it is enough to find a suitable test function $h \geq 1$ which satisfies $h(y) \geq \exp(sy)$ for $y \geq 0$, a compact set C and constants d and $0 < \beta < 1$ such that

$$E\left(h\left(Y_m\right)|Y_{m-1} = y\right) \leq (1-\beta)\, h(y) + d\chi_C(y)$$

where χ_C denotes the indicator function of the set C. Then $E\left(h\left(Y_m\right)\right) < \infty$, and thus (5.16) also holds.

In our case we can choose $h(y) = y^- + \exp\left(sy^+\right)$. By (5.13), $Y_n^+ \leq a_0 y^+ + U_n^+ + V_n^+$. Therefore, (5.14) and (5.15) yield

$$E\left(\exp\left(sY_m^+\right)|Y_{m-1} = y\right) \leq \exp\left(sa_0 y^+\right) L_{U_m^+}(s)\, L_{V_m^+}(s) \leq K(s) \exp\left(sa_0 y^+\right)$$

for all $0 < s < \kappa$. Furthermore, $\exp\left(sa_0 y\right) / \exp\left(sy\right) \to 0$ as $(y \to \infty)$ and

$$E\left(Y_m^- | Y_{m-1} = y\right) \leq a_0 y^- + E\left(U_m^-\right) + E\left(V_m^-\right).$$

Thus, there exists a $C = \{y : |y| \leq m\}$ compact set and $\beta < 1 - a_0$ for which

$$E\left(Y_m^- + \exp\left(sY_m^+\right)|Y_{m-1} = y\right) \leq (1-\beta)\left(y^- + \exp\left(sy^+\right)\right) + d\chi_C(y),$$

hence (5.16) is proven.

It follows that

$$E\left(\exp\left(rZ_m^+\right)\right) \leq E\left(\exp\left(ra_0 Y_{m-1}^+\right)\right) E\left(\exp\left(rU_m^+\right)\right) < \infty$$

if $0 < r < \min\left(\kappa/a_0, s_0\right) > \kappa$. Since Y_m is the independent sum of Z_m and V_m, Lemma 5.4 with the choice $Q_1 = V_m$ and $Q_2 = Z_m$ immediately implies that Y_m has Exp (κ) upper tail.

Finally, it is easy to show by the constant hazard property of the geometric distribution that the same asymptotic results hold for the tail of X_t in the whole nonstationary period (i.e. $X_t|(I_t = 1)$), not just of Y_m. On the other hand, Assumption 5.3 ensures that

$$E\left(\exp\left(sX_t\right)|I_t = 0\right) < \infty$$

for all $0 < s < \min\left(s_0, \kappa/a_0\right) > \kappa$, hence the tail of X_t is completely determined by the $I_t = 1$ regime. Thus the theorem is proven. □

Corollary 5.12. *Under Assumptions 5.1–5.3 the stationary distribution of X_t belongs to the domain of attraction of the GPD with $\xi = 0$ (with the choice $a(u) = \kappa^{-1}$ in equation* (2.7)*) and to the max-domain of attraction of the Gumbel distribution.*

Remark 5.13. *It follows from the proof that $P\left(X_t > u|I_t = 0\right)/P\left(X_t > u|I_t = 1\right) \to 0$ as $u \to \infty$, hence $P\left(I_t = 1|X_t > u\right) \to 1$ as $u \to \infty$.*

Remark 5.14. *Since there exists an $s > \kappa$ with $L_{\varepsilon_{1,t}}(s) < \infty$, a related consequence is that $(\gamma_m(u) - \zeta_m) \mid \left(M^{(m)} > u\right) \to \infty$ as $u \to \infty$, i.e. the time necessary to reach a high threshold u in a nonstationary regime goes to infinity.*

Remark 5.15. *By Proposition 5.7, $M^{(m)} - Z_m$ and hence also $M^{(m)}$ have asymptotically Exp (κ) upper tail.*

Remark 5.16. *It is clear from the proof that the $0 \leq a_0$ assumption can be substituted with a weaker one which ensures that even when $-1 < a_0 < 0$, the lower tail of $\varepsilon_{1,t}$ does not influence the upper tail of Y_m. In particular, if $a_0 < 0$ and there exists a $\kappa_- > 0$ such that*

$$(1 - p_1)\, L_{-\varepsilon_{1,t}}(\kappa_-) = 1,$$

then V_m has asymptotically Exp (κ_-) lower tail and $a_0 V_m$ has Exp $(-\kappa_-/a_0)$ upper tail. Hence the upper tail of Y_m is not affected and the theorem is valid if $\kappa_-/\kappa > -a_0$. Or, in the case of $L_{-\varepsilon_{1,t}}(s) < \infty$ for all $s > 0$ (e.g. in Examples 5.1 and 5.3), $|a_0| < 1$ is sufficient for the statement of the theorem to hold.

Corollary 5.17. *If $\varepsilon_{1,t}$ is normally distributed with mean μ and variance σ^2, and Assumptions 5.1 and 5.3 hold,*

$$P(X_t > u) \sim K \exp\left(-\frac{\left(\mu^2 - 2\sigma^2 \log(1 - p_1)\right)^{1/2} - \mu}{\sigma^2} u\right).$$

Corollary 5.18. *If $\varepsilon_{1,t}$ is distributed as skewed double exponential with parameters c, λ_L and λ_U, and Assumptions 5.1 and 5.3 holds,*

$$P(X_t > u) \sim K \exp(-\kappa u)$$

where κ is the positive root of (5.7).

Corollary 5.19. *If $\varepsilon_{1,t}$ is $\Gamma(\alpha, \lambda)$-distributed and Assumptions 5.1 and 5.3 hold, then*

$$P(X_t > u) \sim K \exp\left(-\lambda\left(1 - (1 - p_1)^{1/\alpha}\right) u\right).$$

5.3 Extremal clustering behaviour

In contrast to section 4.4 where it was conjectured that the extremal index of the ARMA-β-TARCH model is one and thus its clustering at extreme levels is trivial, in this section we prove that the extremal index of the MS-AR model is smaller than one and hence other aspects of its extremal clustering such as the limiting cluster size or limiting aggregate excess distribution should also be investigated.

Not surprisingly, the key to analysing extremal functionals in the MS-AR model is to examine their behaviour in a typical nonstationary regime exceeding a high threshold u. Thus let

$$C'(u) = \left(\sum_{t=\zeta_m+1}^{\xi_m} g\left(X_t - u\right) \right) | \left(M^{(m)} > u\right) = \sum_{t=\gamma_m(u)}^{\xi_m} g\left(X_t - u\right).$$

To obtain the distributional limit of $C'(u)$ as $u \to \infty$, let us first return to the random walk S_n governed by increments distributed as $\varepsilon_{1,t}$. Since Assumption 5.2 automatically implies that $E\varepsilon_{1,t}^2 < \infty$, we obtain from Lemma 5.6 that $E\varepsilon_{1,t} \geq 0$ ensures $P\left(\tau_u < \infty\right) = 1$ and B_u has a distributional limit B_∞ given by (5.9). On the other hand, if $E\varepsilon_{1,t} < 0$ then $L'_{\varepsilon_{1,t}}(0) < 0$ and so Assumption 5.2 ensures that there is a κ' such that $L_{\varepsilon_{1,t}}\left(\kappa'\right) = 1$ and $L'_{\varepsilon_{1,t}}\left(\kappa'\right) < \infty$. Hence, although $P\left(\tau_u < \infty\right) \to 0$ as $u \to \infty$ in the case with negative drift, it is still true by Asmussen (1982) that the distributional limit of $B_u | \left(\tau_u < \infty\right)$ exists (and it will also be denoted by B_∞).[2]

Now, define the S_n^* random walk as

$$S_0^* = B_\infty, \qquad S_n^* = S_{n-1}^* + \varepsilon_n \quad (n = 1, 2, \dots)$$

where ε_n is i.i.d., distributed as $\varepsilon_{1,t}$ and chosen independently of B_∞. Let T be a Geom (p_1)-distributed random variable, independent of $\{S_n^*\}$. Define

$$C^* = \sum_{k=0}^{T-1} g\left(S_k^*\right). \tag{5.17}$$

Then

Proposition 5.20. *Under Assumptions 5.1–5.3, $C'(u) \to_d C^*$ as $u \to \infty$.*

Proof. If $M^{(m)} > u$, let $\varepsilon^*(u) = X_{\gamma_m(u)} - u$. By conditioning on the value of the end of the last $I_t = 0$ regime (Z_m),

$$\bar{F}_{\varepsilon^*(u)}\left(y\right) = \int_{-\infty}^{u} \bar{F}_{B_{u-z}}(y) f_{Z_m|M^{(m)}>u}(z) dz + \bar{F}_{Z_m|M^{(m)}>u}(u+y). \tag{5.18}$$

By Remark 5.15, as $u \to \infty$,

$$\begin{aligned} f_{Z_m|\left(M^{(m)}>u\right)}\left(z\right) &= P\left(M^{(m)} > u | Z_m = z\right) f_{Z_m}\left(z\right) / P\left(M^{(m)} > u\right) \\ &\sim K \exp\left(-\kappa\left(u - z\right)\right) f_{Z_m}\left(z\right) / \exp\left(-\kappa u\right) = K \exp\left(\kappa z\right) f_{Z_m}\left(z\right). \end{aligned}$$

[2] Asmussen (1982) derives various limit theorems of functionals of $\varepsilon_1, \dots, \varepsilon_{\tau_u}$ subject to the condition $\tau_u < \infty$, showing that the behaviour of the functionals under this condition is similar to the case when $\{\varepsilon_n\}$ are independent and distributed as $F_{\kappa'}$.

It follows from the proof of Theorem 5.11 that $L_{Z_m}(s) < \infty$ for every

$$0 < s < \min(\kappa/a_0, s_0) > \kappa,$$

hence for every $\delta > 0$ there exists a z_0 such that $\lim_{u\to\infty} \bar{F}_{Z_m|\left(M^{(m)}>u\right)}(z_0) < \delta$. Thus the second term on the right hand side of (5.18) and also the integral on (z_0, ∞) is negligible. Therefore, since $\lim_{u\to\infty} \bar{F}_{B_{u-z}}(y) = \bar{F}_{B_\infty}(y)$ for every fixed z and $y > 0$, we obtain that $\lim_{u\to\infty} \bar{F}_{\varepsilon^*(u)}(y) = \bar{F}_{B_\infty}(y)$. Hence, as $u \to \infty$, $\{X_t : \gamma_m(u) \leq t \leq \xi_m\}$ behaves like $\{S^*_k : 0 \leq k \leq T-1\}$, and the statement of the proposition holds. □

Proposition 5.22 below states that the asymptotic distribution of an extremal functional in the case of our Markov-switching autoregressive model depends only on its behaviour in a typical $I_t = 1$ regime. It also gives a formula for calculating the extremal index of the process in terms of the solution of a Wiener-Hopf equation. The proof relies on Lemma 5.21 which states that there is no extremal clustering among values in different $I_t = 1$ regimes because they are asymptotically independent in the extreme value sense. That is, e.g. for the end points of two subsequent such regimes, $P(Y_m > u|Y_{m-1} > u) \to 0$ as $u \to \infty$.

Lemma 5.21. *Let $g(x) = 0$ for $x < 0$, and $g(x) = o(\exp(\kappa x))$ as $x \to \infty$. Then there exists a $K > 0$ such that for all j integers*

$$E\left(g\left(Y_m - u\right)|Y_{m-j} > u\right) \leq K \exp\left[-\kappa\left(1 - a_0^{|j|}\right) u\right]. \tag{5.19}$$

The same bound holds for $E\left(g\left(X_t - u\right)|X_{t-l} > u\right)$ provided that there are j stationary regimes in $(t-l, t)$, and also for $E\left(g\left(M^{(m)} - u\right)|M^{(m-j)} > u\right)$ where $M^{(m)}$ is the maximum of the m-th nonstationary regime.

It follows with the choice $g(x) = \chi_{\{x>0\}}$ that for $j \neq 0$, as $u \to \infty$,

$$P(Y_m > u|Y_{m-j} > u) \to 0.$$

Proof. Let us first assume that $j = 1$. We know from the proof of Theorem 5.11 that U_m+V_m and Y_{m-1} are independent and both have asymptotically Exp (κ) tail. If $Y_{m-1} > 0$, $Y_m \leq a_0 Y_{m-1} + U_m + V_m$. Since $g(x) = o(\exp(\kappa x))$, the bound (5.19) for $j = 1$ follows from Lemma 5.5 with the choice $Q_1 = Y_{m-1}$, $Q_2 = U_m + V_m$ and $a = a_0$.

To prove the bound for $E\left(g\left(X_t - u\right)|X_{t-l} > u\right)$, Remark 5.13 implies that we only have to deal with the case when X_t and X_{t-l} are both in nonstationary regimes. Let $m = D(t)$ be the end of the last stationary regime before t, $Q_1 = X_{t-l}$ and $Q_2 = a_0\left(Y_{m-1} - X_{t-l}\right) + U_m + \left(X_t - Z_m\right)$. Then, for $Y_{m-1} > 0$, $X_t \leq a_0 Q_1 + Q_2$. By the constant hazard property of the geometric distribution, $X_t - Z_m$ has asymptotically Exp (κ)-tail. $a_0\left(Y_{m-1} - X_{t-l}\right)$ is lighter tailed than $X_t - Z_m$, and independent of X_{t-l}, U_m and

$(X_t - Z_m)$. Hence, after using Lemma 5.4 to obtain the tail of Q_2, we can apply Lemma 5.5 with $a = a_0$ to get the required upper bound.

Finally, in the case of regime maxima,

$$E\left(g\left(M^{(m)} - u\right) | M^{(m-1)} > u\right) \leq E\left(g\left(M^{(m)} - u\right) | Y_{m-1} > u\right),$$

and for $Y_{m-1} > 0$, $M^{(m)} \leq a_0 Y_{m-1} + U_m + (M^{(m)} - Z_m)$. By Remark 5.15, $M^{(m)} - Z_m$ has Exp (κ)-tail, and it is independent of $a_0 Y_{m-1}$ and U_m, hence the statement holds by Lemma 5.5.

The $j \neq 1$ cases can be treated similarly. □

Proposition 5.22. *If $g(x) = 0$ for $x < 0$ and $g(x) = o(\exp(\kappa x))$ as $x \to \infty$, the conditions of Theorem 2.20 are satisfied with C^* defined by* (5.17). *That is, $C_n(u_n)$ defined there converges in distribution to a Poisson sum of independent copies of C^*.*

The extremal index θ of the process can be calculated as

$$\theta = \int_{-\infty}^{0} \kappa \exp(\kappa x) Q(x) dx \tag{5.20}$$

where $Q(x)$ is the solution of the Wiener-Hopf equation

$$Q(x) = p_1 + (1 - p_1) \int_0^\infty Q(y) f_{\varepsilon_{1,t}}(x - y) dy. \tag{5.21}$$

Proof. It follows from the existence of the test function constructed in the proof of Theorem 5.11 that Y_m is geometrically ergodic and hence strong mixing with an exponential rate. A routine calculation then yields the strong mixing of X_t with mixing coefficient $\alpha_l = K\rho^l$ for some $0 < \rho < 1$. Thus p_n in (2.12) can be chosen as $K \log n$ with an appropriate $K > 0$. Meanwhile, Theorem 5.11 implies that $u_n \sim \log n / \kappa$ for u_n defined in (2.8).

Moreover, Lemma 5.21 and Theorem 5.11 yield that for all $g(x) = o(\exp(\kappa x))$ and $m > 0$

$$E\left(g\left(X_t - u_n\right) | X_{t-l} > u_n\right) \leq K n^{a_0^j - 1},$$

provided that there are at least j stationary regimes in $(t - l, t)$. Hence, for each $0 < \eta < 1$,

$$\begin{aligned} &E\left(g\left(X_t - u_n\right) | X_{t-l} > u_n\right) \\ &\qquad \leq K n^{a_0^{\lfloor \eta l \rfloor} - 1} + K' P\left(\text{less than} \lfloor \eta l \rfloor \text{stationary regimes in} \left(t - l, t\right)\right). \end{aligned}$$

Here the last term can be bounded from above by $K' P(N_0 > l/2) + K' P(N_1 > l/2)$ where N_0 is the sum of the durations of $\lfloor \eta l \rfloor$ independent stationary regimes and N_1 is the sum of the lengths of $\lfloor \eta l \rfloor$ nonstationary ones. Since regime durations are independent and geometrically distributed, $N_i - (\lfloor \eta l \rfloor - 1)$ $(i = 0, 1)$ are negative binomially distributed with parameter $(\lfloor \eta l \rfloor, p_i)$. It can then be shown easily that if $\eta < p_i/2$, $P(N_i > l/2) \leq K_i \rho_i^l$ $(i = 0, 1)$ with $0 < \rho_i < 1$. Therefore,

$$E\left(g\left(X_t - u_n\right) | X_{t-l} > u_n\right) \leq K n^{a_0^{\eta l} - 1} + K_1 K' \rho_1^l + K_2 K' \rho_2^l, \tag{5.22}$$

hence

$$\sum_{k=p}^{p_n} E\left(g\left(X_k - u_n\right) | X_0 > u_n\right) \leq K p_n n^{a_0^{\eta p} - 1} + K_1^{'} \rho_1^p + K_2^{'} \rho_2^p,$$

which tends to 0 as $n \to \infty$ for all p. Thus, putting $g(x) = \chi_{(x>0)}$, the X_t process satisfies (2.13).

This argument also yields that the number of nonstationary regimes which exceed u_n in time interval $[1, p_n]$, provided that $M_{1,p_n} > u_n$, converges in probability to one as $n \to \infty$. Thus, using (5.22) again, (2.14) follows. Similarly, if $M_{1,p} > u$ for a fixed p, the distribution of $C_p(u)$ deviates from the distribution of $C'(u)$ only for two reasons. First, there may be at least two nonstationary regimes exceeding u in $[1, p]$ (the probability of this event vanishes as $u \to \infty$) and second, the single such regime (say, the m-th one) may be too long to belong entirely to $[1, p]$. If the latter event occurs, $\xi_m - \gamma_m(u) \geq p^{1/2}$ or $\xi_m \in [1, p^{1/2}]$ or $\xi_m \in [p, p + p^{1/2}]$. Since $(\xi_m - \gamma_m(u)) \,|\, \left(M^{(m)} > u\right)$ is geometrically distributed,

$$\limsup_{u \to \infty} \left| P\left(C_p\left(u\right) \leq y | M_{1,p} > u\right) - P\left(C'\left(u\right) \leq y\right) \right| \leq 2p^{1/2} / \left(p + p^{1/2}\right) + \left(1 - p_1\right)^{p^{1/2}}$$

for all p integers and $y \geq 0$. Letting $p \to \infty$ and using Proposition 5.20, (2.15) holds with C^* defined by (5.17). Hence all conditions of Theorem 2.20 have been checked, and the limit result on $C_n(u_n)$ holds.

Turning to the calculation of the extremal index, let us first recall that according to (2.16)-(2.17) θ can be calculated as $\theta = \lim_{p \to \infty} \theta_p$ where

$$\theta_p = \lim_{u \to \infty} P\left(M_{1,p} \leq u | X_0 > u\right).$$

At the same time, with similar arguments as above, one can show that the probability that at least two nonstationary regimes exceed u in $[0, p]$ given $X_0 = u - x$ converges to zero uniformly as $u \to \infty$ if x is in a compact subset of the negative half-line. Moreover,

$$P\left(X_1 > u | X_0 = u - x, I_1 = 0\right) = \bar{F}_{\varepsilon_{0,t}}\left(\left(1 - a\right) u + a x\right) \to 0$$

uniformly on compact subsets, too, therefore the conditional maximum of the $\{X_t\}$ process on $[1, p]$ can be approximated by the maximum of the random walk $\{S_n\}$ stopped at a random time:

$$\limsup_{u\to\infty} \left|P\left(M_{1,p} \le u | X_0 = u - x\right) - P\left(M^S_{1,\min\{T-1,p\}} \le x\right)\right| = 0$$

uniformly if $x \in [-K, 0]$ for any $K > 0$. Hence

$$\limsup_{u\to\infty} \left|P\left(M_{1,p} \le u | X_0 = u - x\right) - P\left(M^S_{1,T-1} \le x\right)\right| \le (1 - p_1)^{p+1}$$

and by the uniform convergence on compact subsets

$$\limsup_{u\to\infty} \left|P\left(M_{1,p} \le u | X_0 > u\right) - \int_{-\infty}^{0} P\left(M^S_{1,T-1} \le x\right) \kappa \exp\left(\kappa x\right) dx\right| \le (1 - p_1)^{p+1} .$$

Letting $p \to \infty$ we obtain

$$\theta = \int_{-\infty}^{0} P\left(M^S_{1,T-1} \le x\right) \kappa \exp\left(\kappa x\right) dx = \int_{-\infty}^{0} Q\left(x\right) \kappa \exp\left(\kappa x\right) dx$$

where $Q(x) = P\left(M^S_{1,T-1} \le x\right)$ can be calculated by the Wiener-Hopf equation

$$\begin{aligned} Q\left(x\right) &= p_1 + (1 - p_1) \int_0^\infty P\left(M^S_{2,T-1} \le x | S_1 = x - y, T \ge 2\right) f_{\varepsilon_{1,t}}\left(x - y\right) dy \\ &= p_1 + (1 - p_1) \int_0^\infty Q\left(y\right) f_{\varepsilon_{1,t}}\left(x - y\right) dy. \end{aligned}$$

□

Put shortly, as long as extremes are concerned, our Markov-switching autoregressive model behaves as a Markov chain, moreover, like a random walk with defective step distribution function $F^*(z)$ which puts p_1 mass to $-\infty$: $F^*\left(z\right) = p_1 + (1 - p_1) F_{\varepsilon_{1,t}}\left(z\right)$. It is not surprising in light of section 2.2.3 that the extremal index is given in terms of the solution of a Wiener-Hopf equation.

Extremal index in some special cases

It follows from Assumption 5.2 that $P\left(\varepsilon_{1,t} > 0\right) > 0$, thus $Q\left(x\right) < 1$ for all $x < 0$ which yields that $\theta < 1$. However, θ can be rarely obtained analytically because the Wiener-Hopf equation rarely has an explicit solution. A trivial exception is when $\varepsilon_{1,t} \ge 0$ a.s., then the extremal index is obviously p_1. This is the case for Example 5.3. Another distribution for which an explicit expression is available is the skewed Laplace one, introduced in Example 5.2. To illustrate this, the following proposition gives the result for $c = 0$.

Proposition 5.23. *If $\varepsilon_{1,t}$ has a skewed Laplace distribution with parameters $c = 0$, λ_L and λ_U, the extremal index of X_t is*

$$\theta = p_1 + (1 - p_1) \left(\frac{\kappa}{\kappa + \lambda_L} \right)^2$$

where κ is defined by (5.7). *In particular, for $\lambda_U = \lambda_L$,*

$$\theta = \frac{2p_1}{1 + p_1^{1/2}}. \tag{5.23}$$

Proof. The Wiener-Hopf equation can now be solved explicitly because the right tail of $\varepsilon_{1,t}$ is exponential. Let $S'_0 = 0$, $S'_n = S'_{n-1} + \varepsilon'_n$ be a random walk with $E\varepsilon'_n < 0$ and with exponential right tail, i.e. $f_{\varepsilon'_n}(x) = \alpha \exp(-\delta x)$ for $x > 0$ (and no assumption is made on the form of the distribution for $x \leq 0$.) Let $\kappa' > 0$ satisfy $L_{\varepsilon'_n}(\kappa') = 1$ and let M' be the maximum of the random walk. Then Asmussen (1987, Chapter IX., Thm. 1.2.) states that the Cramer-Lundberg approximation is exact, or more precisely

$$P(M' > u) = \left(1 - \frac{\kappa'}{\delta}\right) \exp(-\kappa' u).$$

for $u > 0$.

As in the heuristic proof of Proposition 5.7, let $\varepsilon'_n = -\infty$ with probability p_1 and be distributed as $\varepsilon_{1,t}$ with probability $1 - p_1$. Since $M' = M^S_{0,T-1} = \max\left(0, M^S_{1,T-1}\right)$ and $\kappa = \kappa'$ and $\delta = \lambda_U$ in our case we obtain

$$Q(x) = P\left(M^S_{1,T-1} < x\right) = 1 - \left(1 - \frac{\kappa}{\lambda_U}\right) \exp(-\kappa x)$$

for $x > 0$. Hence for $x \leq 0$ (5.21) yields

$$Q(x) = p_1 + (1 - p_1) \exp(\lambda_L x) \frac{\kappa}{\kappa + \lambda_L}.$$

(A formal proof just shows that the above two formulas satisfy the Wiener-Hopf equation.) Finally integration in (5.20) shows the statement.

□

There is no closed form for $Q(x)$ in the Gaussian case (Example 5.1). Instead, we approximate the extremal index by Monte Carlo simulations in this setting.[3] It is easy to show that θ is determined by p_1 and μ/σ, so Figure 5.1/a displays θ as a function of $r = \mu/\sigma$ for $p_1 = 1/8$, 1/4, 3/8 and 1/2. Obviously, $\theta > p_1$ and $\theta \to p_1$ as $r \to \infty$. However, according to Figure 5.1/b, θ is only moderately higher than p_1 even for $\mu = r = 0$. (The

[3] The simulation is straightforward by the fact that (5.20) yields $\theta = P\left(M^S_{1,T-1} \leq -Y | Y \sim \text{Exp}(\kappa)\right)$.

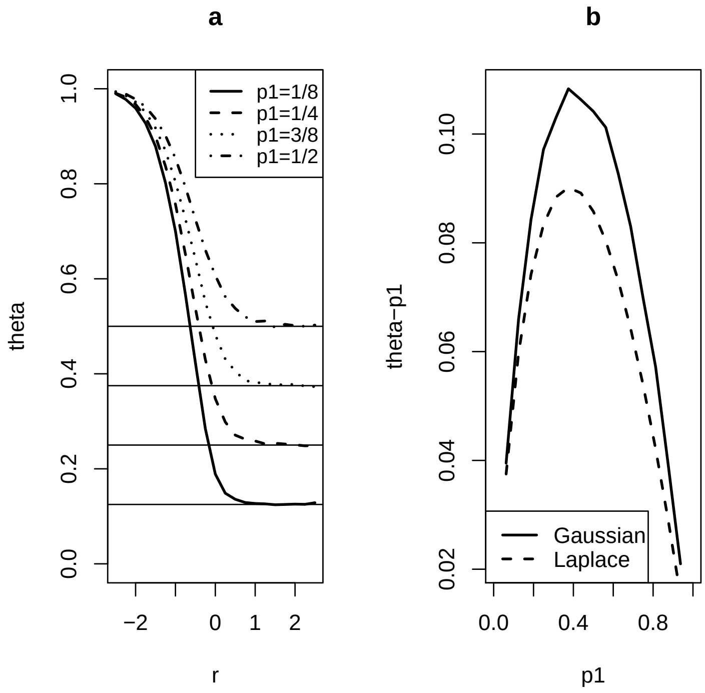


Figure 5.1: a) θ as a function of $r = \mu/\sigma$ for four different values of p_1 in the Gaussian model. b) $\theta - p_1$ as a function of p_1 for the Gaussian and the Laplace-distributed model if $E(\varepsilon_{1,t}) = 0$.

difference is highest when p_1 is close neither to 0 nor to 1.) Figure 5.1/b also displays $\theta - p_1$ as a function of p_1 for the symmetric Laplace-distributed case with $c = 0$ (see (5.23)). It clearly shows that if $E\left(\varepsilon_{1,t}\right) = 0$ and p_1 is the same, clustering is higher (i.e. θ is lower) for the process driven by a Laplace-noise than for the one generated by a Gaussian noise.

Extremal cluster functionals in special cases

More elaborate extremal characteristics such as the limiting cluster size distribution or the limiting aggregate excess distribution can be calculated explicitly in even fewer cases. For instance, when $\varepsilon_{1,t} > 0$ a.s. (as in Example 5.3), $S_k^* > 0$ for all k, hence the limiting cluster size distribution is geometric with parameter p_1. However, this is no longer the case in the Gaussian or double exponential setting.

To illustrate this, we simulated limiting cluster size distributions (denoted by N) for $p_1 = 0.25$, 0.5 and 0.75 when the generating noise is the Laplace distribution with $c = 0$. The calculated empirical hazard functions $P(N = k|N \geq k)$ – plotted in Figure 5.2 –

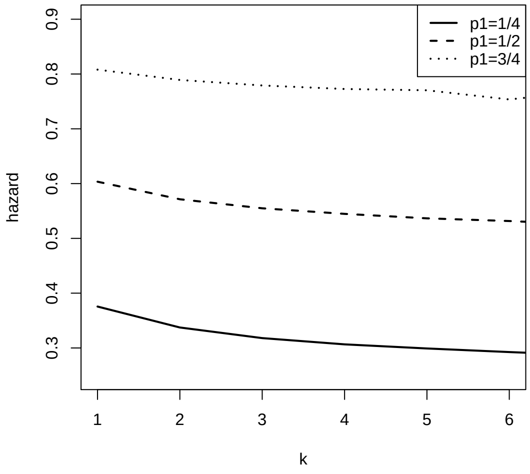


Figure 5.2: Hazard function of the limiting cluster size distributions of the model with Laplace-distributed noise with zero mean for three different values of p_1

show that hazards are decreasing with k, therefore the cluster size distributions belong to the DFR (decreasing failure rate) family for these parameters.

Turning to the limiting aggregate excess distribution, it cannot be obtained in a closed form even in a relatively simple case, i.e. when $\varepsilon_{1,t}$ is Gamma-distributed. Nevertheless, we prove below that in this case the tail of the limiting aggregate excess distribution can be approximated by Weibull-like distributions with exponent parameter 1/2. This finding will turn out to be important in the application presented in section 5.6.

Let

$$W_n(u_n) = \sum_{t=1}^{n} (X_t - u_n)^+$$

be the aggregate excess above u_n, then $W_n(u_n)$ converges in distribution to a Poisson sum of i.i.d. random variables distributed as W^*, which has the following property in the Gamma-distributed case. The proof relies on renewal theoretical arguments and on Laplace's method of sums.

Theorem 5.24. *Let Assumption 5.1 and 5.3 hold and the condition in Example 5.3 be satisfied. Then there exist $K_i > 0$ $(i = 1, 2)$ constants such that*

$$K_1 \exp\left(-2^{3/2}\left(\lambda_0^{-1} - \alpha\lambda_0\right)(\lambda y)^{1/2}\right) \le \bar{F}_{W^*}(y) \le K_2 \exp\left(-2\left(\lambda_0^{-1} - \alpha\lambda_0\right)(\lambda y)^{1/2}\right) \tag{5.24}$$

where λ_0 is the unique real number satisfying

$$\lambda_0^{-2} - 2\alpha \log \lambda_0 + \log(1 - p_1) - \alpha(1 + \log \alpha) = 0. \tag{5.25}$$

Proof. We may assume that $\lambda = 1$. By Propositions 5.20 and 5.22, and since $S_k^* > 0$ for all k in the Gamma-case,

$$W^* = \sum_{k=0}^{T-1} (S_k^*)^+ = TB_\infty + \sum_{k=1}^{T-1} (T-k)\varepsilon_k. \tag{5.26}$$

T is a Geom (p_1)-distributed random variable, ε_k $(k = 1, 2, \dots)$ are distributed as $\Gamma(\alpha, 1)$ and the density of B_∞ is given by (5.9). Since $\varepsilon_i \geq 0$ a.s., B_0 in (5.9) is also distributed as $\Gamma(\alpha, 1)$.

Let us first examine the $\alpha \geq 1$ case. Then the $\Gamma(\alpha, 1)$ distribution belongs to the class of NBUE-distributions ("new better than used in expectation"), thus for all $y \in \mathbf{R}_+$

$$\bar{F}_{B_\infty}(y) \leq \bar{F}_{\Gamma(\alpha,1)}(y). \tag{5.27}$$

Since $\bar{F}_{\Gamma(\alpha,1)}(v) \geq \exp(-v) E(\Gamma(\alpha, 1))$ for all $v \geq 0$, a lower bound can also be given straightforwardly for $\bar{F}_{B_\infty}(y)$:

$$\bar{F}_{B_\infty}(y) = \frac{\int_y^\infty \bar{F}_{\Gamma(\alpha,1)}(v)}{E(\Gamma(\alpha, 1))} \geq \bar{F}_{\mathrm{Exp}(1)}(y).$$

To get an upper bound for the tail of W^* we first observe that $W^* \leq R_1 = T\left(\sum_{k=0}^{T-1} \varepsilon_k\right)$ by (5.26) and (5.27) (here ε_0 is also chosen as $\Gamma(\alpha, 1)$), hence it is enough to examine the tail of R_1. Since $R_1|(T = k)$ is distributed as $\Gamma(k\alpha, 1/k)$, we obtain for $y \to \infty$ that

$$f_{R_1}(y) = \sum_{k=1}^{\infty} p_1 (1-p_1)^{k-1} \frac{1}{\Gamma(k\alpha)} \left(\frac{1}{k}\right)^{k\alpha} y^{k\alpha-1} \exp(-y/k) = \sum_{k=1}^{\infty} \exp(h(y, k)) \tag{5.28}$$

where

$$h(y, k) = \log p_1 + (k-1)\log(1-p_1) - \log\Gamma(k\alpha) + (k\alpha - 1)\log y - k\alpha\log k - y/k.$$

The asymptotic behaviour of this sum as $y \to \infty$ can be examined by Laplace's method for sums (see e.g. Bender and Orszag (1999)). After finding the location of the maximum $k_{\max}(y)$ of the function $k \to h(y, k)$, we use Taylor series expansion around $k_{\max}(y)$ to obtain the following result (the proof is given later):

Lemma 5.25. *There exists a $K > 0$ constant such that with λ_0 defined by* (5.25)

$$f_{R_1}(y) \sim Ky^{-1/2} \exp\left(-2\left(\lambda_0^{-1} - \alpha\lambda_0\right) y^{1/2}\right). \tag{5.29}$$

Integrating this directly gives the upper bound in (5.24). To give a lower bound for $P(W^* > y)$, let us introduce $R_2 = (T-1)\left(\sum_{k=1}^{T-1} \varepsilon_k\right)$. (The notation implies that R_2

takes 0 with probability p_1.) Clearly, for $y > 0$, $f_{R_2}(y) = (1-p_1)\, f_{R_1}(y)$, hence the approximation in Lemma 5.25 – though with a different constant – applies for $f_{R_2}(y)$. Moreover, the variables $W_1^{'} = \sum_{k=1}^{T-1}(k-1/2)\,\varepsilon_k$ and $W_1^{''} = \sum_{k=1}^{T-1}(T-k-1/2)\,\varepsilon_k$ are identically distributed, take only positive values and $R_2 = W_1^{'} + W_1^{''}$, hence $\bar{F}_{R_2}(2y) \le 2\bar{F}_{W_1^{''}}(y)$ for all y. Additionally, by (5.26), $W^* > W_1^{''}$ a.s., hence $\bar{F}_{R_2}(2y) \le 2\bar{F}_{W^*}(y)$, which yields the lower bound in (5.24). This concludes the proof when $\alpha \ge 1$.

When $\alpha < 1$, similar calculations give $\bar{F}_{\Gamma(\alpha,1)}(y) \le \bar{F}_{B_\infty}(y) \le \bar{F}_{\text{Exp}(1)}(y)$ for all $y > 0$ hence the lower bound for $P(W^* > y)$ can be obtained by observing that the variables $W_2^{'} = \sum_{k=0}^{T-1}(k+1/2)\varepsilon_k$ and $W_2^{''} = \sum_{k=0}^{T-1}(T-k-1/2)\,\varepsilon_k$ are identically distributed, their sum is R_1 and they are stochastically smaller than W^*. On the other hand, $R_3 = \sum_{k=T}^{T+\lfloor 1/\alpha \rfloor}\varepsilon_k$ is distributed as $\Gamma\left(\alpha\lfloor 1+1/\alpha\rfloor, 1\right)$, hence is stochastically larger than an Exp(1)-distributed variable. Therefore, W^* is stochastically smaller than $R_4 = T\left(\sum_{k=1}^{T+\lfloor 1/\alpha\rfloor} E_k\right)$ and this can be easily shown to have the same tail as R_1, though with a different constant. This concludes the proof for $\alpha < 1$, too. □

Proof of Lemma 5.25. Denoting the logarithmic derivative of the Gamma-function by $\Psi(.)$ and using $\Psi(x) = \log x + O(x^{-1})$, $\Psi'(x) = x^{-1} + O(x^{-2})$ and $\Psi''(x) = O(x^{-2})$ (see Abramowitz and Stegun (1965)) we obtain that

$$\begin{aligned}
h_k'(y,k) &= \log(1-p_1) - \alpha\Psi(k\alpha) + \alpha\log y - \alpha\log k - \alpha + yk^{-2} = \\
&= yk^{-2} + \alpha\log\left(yk^{-2}\right) + \log(1-p_1) - \alpha(1+\log\alpha) + O\left(k^{-1}\right) \\
h_k''(y,k) &= -\alpha^2\Psi'(k\alpha) - \alpha k^{-1} - 2yk^{-3} = -2\alpha k^{-1} - 2yk^{-3} + O\left(k^{-2}\right) \\
h_k'''(y,k) &= O\left(k^{-2}\right) + O\left(yk^{-4}\right).
\end{aligned}$$

Solving the equation $h_k'(y,k) = 0$ for k yields $k_{\max}(y) = \lambda_0 y^{1/2} + O(1)$ with λ_0 defined by (5.25). (It is easy to check that λ_0 satisfies $0 < \lambda_0 < \alpha^{-1/2}$.) If we use the notation $k_y = \lambda_0 y^{1/2}$ and apply Stirling's formula, we obtain from above and from the definition of λ_0 that

$$\begin{aligned}
h(y,k_y) &= \lambda_0 y^{1/2}\left(\log(1-p_1) - \alpha\log\alpha + \alpha - 2\alpha\log\lambda_0 - \lambda_0^{-2}\right) - \\
&\quad - 3/4\log y + K + O\left(y^{-1/2}\right) = \\
&= -2\lambda_0 y^{1/2}\left(\lambda_0^{-2} - \alpha\right) - 3/4\log y + K + O\left(y^{-1/2}\right) \\
h_k'(y,k_y) &= O\left(y^{-1/2}\right) \\
h_k''(y,k_y) &= -2\lambda_2 y^{-1/2} + O\left(y^{-1}\right) \\
h_k'''(y,k_y) &= O\left(y^{-1}\right)
\end{aligned}$$

where $\lambda_2 = \alpha\lambda_0^{-1} + \lambda_0^{-3}$.

To examine the sum in (5.28) we distinguish between different values of k. If $|k-k_y| < y^{3/10}$, a Taylor-series expansion around k_y gives

$$\begin{aligned} h\left(y,k\right)-h\left(y,k_y\right) &= \left(k-k_y\right)O\left(y^{-1/2}\right)-\left(k-k_y\right)^2\left(\lambda_2 y^{-1/2}+O\left(y^{-1}\right)\right)+\\ &+\left(k-k_y\right)^3 O\left(y^{-1}\right) = -\left(k-k_y\right)^2\lambda_2 y^{-1/2}+O\left(y^{-1/10}\right). \end{aligned} \tag{5.30}$$

Therefore, as $y\to\infty$,

$$\begin{aligned} \sum_{|k-k_y|<y^{3/10}}\exp\left(h\left(y,k\right)-h\left(y,k_y\right)\right) &\sim \sum_{|j|<y^{3/10}}\exp\left(-\frac{1}{2}\left(\frac{2^{1/2}\lambda_2^{1/2}}{y^{1/4}}j\right)^2\right)\sim\\ &\sim \frac{y^{1/4}}{\left(2\lambda_2\right)^{1/2}}\int_{-(2\lambda_2)^{1/2}y^{1/20}}^{(2\lambda_2)^{1/2}y^{1/20}}\exp\left(-t^2/2\right)dt\sim Ky^{1/4}. \end{aligned}$$

On the other hand, using the fact that $k\to h_k'(y,k)$ is a decreasing function for all y and $h_k'\left(y,k_y\right)=0$, we obtain from (5.30) that for all $|k-k_y|\geq y^{3/10}$

$$h\left(y,k\right)-h\left(y,k_y\right)\leq -y^{6/10}\lambda_2 y^{-1/2}+O\left(y^{-1/10}\right).$$

Moreover, if $k>y$ then $h(y,k)<-\log\Gamma(k\alpha)$. Hence, as $y\to\infty$,

$$\begin{aligned} &\sum_{\left\{k>0:\ |k-k_y|\geq y^{3/10}\right\}}\exp\left(h\left(y,k\right)-h\left(y,k_y\right)\right)\\ &\qquad\leq Ky\exp\left(-\lambda_2 y^{1/10}\right)+\exp\left(-h\left(y,k_y\right)\right)\sum_{k>y}1/\Gamma\left(k\right)=o(1). \end{aligned}$$

Combining the above estimates yields

$$\begin{aligned} f_{R_1}\left(y\right) &\sim \exp\left(h\left(y,k_y\right)\right)\sum_{k=1}^{\infty}\exp\left(h\left(y,k\right)-h\left(y,k_y\right)\right)\\ &\sim Ky^{1/4}\exp\left(h\left(y,k_y\right)\right)\sim Ky^{-1/2}\exp\left(-2\left(\lambda_0^{-1}-\alpha\lambda_0\right)y^{1/2}\right). \end{aligned}$$

□

5.4 Extremes of Markov-switching, conditionally Markov models

Section 2.2.3 has revealed that examining the extremal behaviour of random walks is not only interesting on its own right but sheds light on the extremal clustering properties of certain Markov chains as well. Therefore, it is a natural question to ask whether the Markov-switching autoregressive model of the previous sections (which behaves as a random walk in one regime) also arises as an extremal limit of certain more general models. Obvious candidates are the Markov-switching, conditionally Markov models, defined in the following.

Assumption 5.4. *Let I_t be a two-state discrete time Markov chain with transition probabilities given by* (5.3)-(5.4). *Let X_t be a stationary process whose conditional distribution, provided that I_t is known, only depends on the value of X_{t-1} (i.e. X_t is conditionally Markov in each regime). Formally, for $A_t \subset \mathbf{R}$ Borel-sets and $j_t \in \{0,1\}$,*

$$\begin{aligned} P\left(X_t \in A_t | I_t = j_t, X_{t-i} \in A_{t-i}, I_{t-i} = j_{t-i}, i = 1, 2, \ldots\right) & \\ = P\left(X_t \in A_t | X_{t-1} \in A_{t-1}, I_t = j_t\right). & \end{aligned}$$

Moreover, for each t, conditionally on $(I_1, I_2, \ldots, I_t)$, the set of variables $(X_1, X_2, \ldots, X_t)$ is independent of $(I_{t+1}, I_{t+2}, \ldots)$.

The simplest model of this class is obtained when the sequence $\{X_t\}$ is conditionally independent given $\{I_t\}$ and the distribution of X_t only depends on I_t. Extremes of this restricted model were examined in detail by Resnick (1971) and his results were later generalised to allow some form of conditional dependence (see e.g. Turkman and Oliveira (1992)). However, these generalisations still assume that the distribution of X_t only depends on I_t, i.e. that I_{t-i} $(i \geq 1)$ does not yield new information on X_t provided that I_t is known. This restriction does not necessarily hold for models satisfying Assumption 5.4, hence our following analysis is basically novel. For instance, Markov-switching AR(1) processes lie within the framework of Assumption 5.4 but do not satisfy the conditions of Turkman and Oliveira (1992): the distribution of X_t in such a process depends not only on I_t but on the location of t in the actual regime, thus e.g. on I_{t-1} as well.

Furthermore, we assume that the stationary distribution of the process X_t is asymptotically exponential in each regime, i.e. using the notations $F_j(x) = P\left(X_t < u | I_t = j\right)$ and $\bar{F}_j(u) = 1 - F_j(u)$ $(j = 0, 1)$ the following holds. (Note that in the sequel, unless indicated otherwise, all probability statements are made under the stationary distribution of X_t.)

Assumption 5.5. *X_t has an absolutely continuous distribution with respect to the Lebesgue-measure and there exist $K_0 > 0$ and $K_1 > 0$ such that*

$$\bar{F}_1(u) \sim K_1 e^{-\kappa u} \tag{5.31}$$

$$\bar{F}_0(u) \sim K_0 e^{-\kappa u / a} \tag{5.32}$$

where $0 < a \leq 1$ holds.

This assumption is more straightforward than it looks at a first glance. (5.31) can always be satisfied by a marginal transformation of the process, which leaves the extremal *dependence* structure essentially unchanged. As far as the second equation is concerned, we may assume without much loss of generality that F_0 also belongs to the domain of attraction of a GPD and that $\bar{F}_0(u) \leq \bar{F}_1(u)$ for sufficiently large u. Then, as the exponential distribution is the GPD with shape parameter $\xi_1 = 0$, the shape parameter of the GPD corresponding to F_0 can only be $\xi_0 \leq 0$. The $\xi_0 < 0$ case is not particularly interesting because the support of F_0 has a finite upper end point in this case. Hence only $\xi_0 = 0$ remains and one may obtain (5.32) as a reasonable approximation of the tail of F_0.[4]

Assumption 5.5 immediately implies that $P(I_t = 0|X_t > u) \to 0$ if $0 < a < 1$. With this in mind, the $I_t = 1$ regime will be called as the dominant regime, and the other as the dominated one, although strictly speaking this distinction is not valid for $a = 1$.

Turning to the dependence structure, it is completely determined for given marginals by the bivariate distributions $((X_{t-1}, X_t)|I_t = j)$ $(j = 0, 1)$ because of the conditional Markovity assumption. Indeed, with a slightly vague use of notation, the joint distribution can be written as

$$P(X_n, \ldots, X_1) = \sum_{I_n, I_{n-1}, \ldots, I_1} P(X_n, X_{n-1}, \ldots, X_1|I_n, I_{n-1}, \ldots, I_1)\, P(I_n, I_{n-1}, \ldots, I_1)$$

with

$$\begin{aligned} & P(X_n, X_{n-1}, \ldots, X_1|I_n, \ldots, I_1) \\ & \quad = P(X_n|X_{n-1}, \ldots, X_1, I_n, \ldots, I_1)\, P(X_{n-1}, X_{n-2}, \ldots, X_1|I_n, I_{n-1}, \ldots, I_1) \\ & \quad = P(X_n|X_{n-1}, I_n)\, P(X_{n-1}, X_{n-2}, \ldots, X_1|I_n, \ldots, I_1) \\ & \quad = P(X_n|X_{n-1}, I_n)\, P(X_{n-1}, X_{n-2}, \ldots, X_1|I_{n-1}, \ldots, I_1) \end{aligned}$$

where we used Assumption 5.4. Furthermore,

$$P(X_n|X_{n-1}, I_n) = \frac{P(X_{n-1}, X_n|I_n)}{\sum_{I_n} P(X_{n-1}|I_{n-1})\, P(I_{n-1}|I_n)},$$

hence recursively

$$P(X_n, X_{n-1}, \ldots, X_1|I_n, I_{n-1}, \ldots, I_1) = \prod_{t=1}^{n} \frac{P(X_{t-1}, X_t|I_t)}{\sum_{I_t} P(X_{t-1}|I_{t-1})\, P(I_{t-1}|I_t)}.$$

Concerning the bivariate dependence structure, it is natural to assume that the conditional joint distributions $((X_{t-1}, X_t)|I_t = j)$ for $j = 0, 1$ both belong to the domain of

[4] It is clear from section 2.2.1 that $\xi = 0$ is compatible with many different distributions. However, by Theorem 2.6 the exponential distribution is the best choice because in this case exceedances themselves follow the GPD.

attraction of bivariate extreme value laws with spectral measures H_0 and H_1, respectively, as this assumption is satisfied by all practically relevant bivariate distributions. A short introduction into the topic of bivariate extreme value theory was given in section 2.2.3.

Equation (2.22) suggests that if both marginals of a bivariate extreme value law are unit exponential, the joint distribution behaves (under some regularity conditions) as a random walk above high thresholds. Although the unit exponentiality assumption is not valid for either conditional joint distributions in the Markov-switching, conditionally Markov model, the idea behind (2.22) can be applied after appropriate transformations as follows.

For later reference, using the notations $a_1 = 1$ and $a_0 = a$, define for $j = 0, 1$

$$F_j^u(z) = P\left(X_t < a_j u + z | X_{t-1} = u, I_t = j\right).$$

We will examine whether $F_j^u(z)$ has a limit (in the spirit of (2.22)) as $u \to \infty$.

It is a well known fact that I_t – being a two-state Markov chain – is a Markov chain in reversed time as well, and its transition matrix is equal to that of the original chain:

$$P\left(I_{t-1} = i | I_t = j\right) = P\left(I_t = i | I_{t-1} = j\right)$$

for $i, j \in \{0, 1\}$. As

$$P\left(X_{t-1} > u | I_t = i\right) = \sum_{j=0}^{1} P\left(X_{t-1} > u | I_{t-1} = j\right) P\left(I_{t-1} = j | I_t = i\right)$$

for $i = 0, 1$, one can obtain

$$P\left(X_{t-1} > u | I_t = 1\right) \sim \left(1 - p_1\right) K_1 e^{-\kappa u} + p_1 K_0 e^{-\kappa u/a} \tag{5.33}$$

$$P\left(X_{t-1} > u | I_t = 0\right) \sim p_0 K_1 e^{-\kappa u} + \left(1 - p_0\right) K_0 e^{-\kappa u/a}. \tag{5.34}$$

Thus (5.33) implies

$$P\left(X_{t-1} + c_1 > u | I_t = 1\right) \sim K_1 e^{-\kappa u} \sim P\left(X_t > u | I_t = 1\right), \tag{5.35}$$

where $c_1 = \log\left(1 - p_1\right)/\kappa$ if $0 < a < 1$ and $c_1 = \log\left(1 - p_1 + p_1 K_0 / K_1\right)/\kappa$ if $a = 1$. Hence, if $I_t = 1$, both marginals of $\left(\kappa\left(X_{t-1} + c_1\right), \kappa X_t\right)$ are asymptotically unit exponential. Taking into account the bivariate extreme value assumption, it follows from (2.22) that – under further regularity conditions – the limit

$$\lim_{u\to\infty} P\left(\kappa X_t < u + z | \kappa\left(X_{t-1} + c_1\right) = u, I_t = 1\right)$$

and hence also the limit

$$F_1^*(z) := \lim_{u\to\infty} F_1^u(z) \tag{5.36}$$

exist for all z. Similarly, (5.34) yields that

$$P\left(aX_{t-1}+c_0>u|I_t=0\right)\sim K_0e^{-\kappa u/a}\sim P\left(X_t>u|I_t=0\right) \tag{5.37}$$

with $c_0 = a/\kappa\log\left(K_0/\left(p_0K_1\right)\right)$ if $0<a<1$ and $c_0=\log\left(1-p_0+p_0K_1/K_0\right)/\kappa$ if $a=1$. Hence, if $I_t=0$, both marginals of $\left(\kappa/a\left(aX_{t-1}+c_0\right),\kappa X_t/a\right)$ are asymptotically unit exponential. Thus

$$\lim_{u\to\infty}P\left(\kappa X_t/a<u+z|\kappa/a\left(aX_{t-1}+c_0\right)=u,I_t=0\right)$$

and also

$$F_0^*\left(z\right):=\lim_{u\to\infty}F_0^u\left(z\right) \tag{5.38}$$

exist under appropriate regularity conditions. (We will also use the notations $\bar{F}_j^*\left(z\right)=1-F_j^*\left(z\right)$ for $j=0,1$.) It is natural to assume the following, slightly stronger versions of (5.36) and (5.38) for X_t :

Assumption 5.6. *The joint distributions* $\left(X_{t-1},X_t\right)|\left(I_t=j\right)$ $(j=0,1)$ *are absolutely continuous with respect to the Lebesgue-measure. There exist (possibly improper) distribution functions* $F_j^*(z)$ *such that* $F_j^u\left(z\right)\to F_j^*\left(z\right)$ *as* $u\to\infty$ *uniformly on all compact intervals* $(j=0,1)$. *Moreover, if* $F_j^*\left(-\infty\right)=\lim_{z\to-\infty}F_j^*(z)>0$ *for a* j, *then*

$$\lim_{M\to\infty}\limsup_{u\to\infty}\sup_{y\ge M}P\left(X_t>a_iu|X_{t-1}=u-y,I_t=i\right)=0 \tag{5.39}$$

is satisfied for $i=0,1$ *where* $a_1=1$ *and* $a_0=a$.

Note that $F_j^*(\infty)=\lim_{z\to\infty}F_j^*\left(z\right)=1$ always holds because of the exponential decay of the marginal distributions. Indeed, if $F_j^*\left(z\right)$ exists for a z, then Assumption 5.5 and equations (5.35) and (5.37) imply

$$\begin{aligned}K_je^{-\kappa(u+z/a_j)}&\sim P\left(X_t>a_ju+z|I_t=j\right)\\&\ge\int_u^\infty P\left(X_t>a_jv+z|X_{t-1}=v,I_t=j\right)f_{(X_{t-1}|I_t=j)}\left(v\right)dv\\&\sim\bar{F}_j^*\left(z\right)P\left(X_{t-1}>u|I_t=j\right)\sim\bar{F}_j^*\left(z\right)K_je^{-\kappa(u+c_j/a_j)}\end{aligned}$$

hence $\bar{F}_j^*\left(z\right)\le e^{-\kappa(z-c_j)/a_j}$.

On the other hand, $F_j^*(-\infty)>0$ may well happen. For instance, if X_t is independent conditionally on I_t, then $F_j^*(z)=1$ for all z. Condition (5.39) is needed in order to rule out models which "jump" from a moderate or an extremely low level to a very high one in a single step (e.g. "tail-switching" models such as ARCH-type processes).

Let us now introduce a (not necessarily stationary) Markov-switching autoregressive process, with the I_t process in the background:

$$Y_t = Y_{t-1} + \varepsilon_{1,t} \qquad \text{if} \quad I_t = 1, \tag{5.40}$$

$$Y_t = aY_{t-1} + \varepsilon_{0,t} \qquad \text{if} \quad I_t = 0 \tag{5.41}$$

where $0 < a \leq 1$ and $\varepsilon_{j,t}$ $(j = 0, 1)$ are both i.i.d. (possibly non-finite) random variables with distribution functions $F_j^*(z)$, and they are independent of each other as well. (They take $-\infty$ with probability $F_j^*(-\infty)$.) Note that Y_t automatically satisfies Assumptions 5.4 (apart from the stationarity condition if $a = 1$) and 5.6, while Theorem 5.11 ensures that Assumption 5.5 also holds under further technical conditions on the distribution of $\varepsilon_{j,t}$.

As the following Proposition states, X_t and Y_t behave similarly in the region of extremes. For ease of notation, for any symbol $\mathbf{w} \in \{\mathbf{X}, \mathbf{Y}, \mathbf{x}, \mathbf{y}, \mathbf{r}, \mathbf{s}\}$ let $\mathbf{w}_{k,l} = (w_k, \ldots, w_l)$. (If $k = 1$ and $l = p$, the subscripts will be occasionally omitted.) We will use $\mathbf{w}_{k,l} < \mathbf{v}_{k,l}$ if $w_i < v_i$ $(k \leq i \leq l)$. Also, for a fixed set of $j_i \in \{0, 1\}$ $(i = k, \ldots, l)$ let $\mathbf{a}_{k,l} = (a_{j_k}, \ldots, a_{j_l})$, $a_{k,l}^C = \prod_{i=k}^l a_{j_i}$ and $\mathbf{a}_{k,l}^C = \left(a_{k,k}^C, \ldots, a_{k,l}^C\right)$. Finally, for given $\{j_i\}$, $\mathbf{x}_{k,l}$ and $\mathbf{y}_{k,l}$, define the following notations for the events

$$\begin{aligned} A_{k,l} &= \{I_i = j_i \, (i = k, \ldots, l)\} \\ B_{k,l}^{u,\mathbf{x}_{k,l},\mathbf{y}_{k,l}} &= \left\{\mathbf{a}_{k,l}^C u + \mathbf{x}_{k,l} \leq \mathbf{X}_{k,l} < \mathbf{a}_{k,l}^C u + \mathbf{y}_{k,l}\right\} \\ C_{k,l}^{u,\mathbf{x}_{k,l},\mathbf{y}_{k,l}} &= \left\{\mathbf{a}_{k,l}^C u + \mathbf{x}_{k,l} \leq \mathbf{Y}_{k,l} < \mathbf{a}_{k,l}^C u + \mathbf{y}_{k,l}\right\}. \end{aligned}$$

Proposition 5.26. *Let us assume Assumptions 5.4-5.6 and let $a_1 = 1$ and $a_0 = a$. Then, for all p, $j_i \in \{0, 1\}$ and y_i $(i = 1, \ldots, p)$,*

$$\lim_{u\to\infty} \left|P\left(\mathbf{X}_{1,p} < \mathbf{a}_{1,p}^C u + \mathbf{y}_{1,p} | X_0 = u, A_{1,p}\right) - P\left(\mathbf{Y}_{1,p} < \mathbf{a}_{1,p}^C u + \mathbf{y}_{1,p} | Y_0 = u, A_{1,p}\right)\right| = 0.$$

Proof. Let $j_i \in \{0, 1\}$ and $\mathbf{r} < \mathbf{s}$. We first prove by induction that

$$\lim_{u\to\infty} \sup_{\mathbf{r}<\mathbf{x}<\mathbf{y}<\mathbf{s}} \left|P\left(B_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,p}\right) - P\left(C_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,p}\right)\right| = 0. \tag{5.42}$$

Indeed, for $p = 1$,

$$\begin{aligned} &P\left(B_{1,1}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,1}\right) - P\left(C_{1,1}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,1}\right) \\ &\qquad = \left(F_{j_1}^u(y_1) - F_{j_1}^*(y_1)\right) - \left(F_{j_1}^u(x_1) - F_{j_1}^*(x_1)\right), \end{aligned}$$

which tends to 0 uniformly on all compact intervals according to Assumption 5.6. Then

$$\begin{aligned}
&\left|P\left(B_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,p}\right) - P\left(C_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,p}\right)\right| \\
&\quad = \left|\int_{x_1}^{y_1} P\left(B_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid X_1 = a_{j_1}u + v, A_{2,p}\right) dF_{j_1}^{u}\left(v\right)\right. \\
&\qquad \left. - \int_{x_1}^{y_1} P\left(C_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid Y_1 = a_{j_1}u + v, A_{2,p}\right) dF_{j_1}^{*}\left(v\right)\right| \\
&\quad \leq \int_{x_1}^{y_1} \left|P\left(B_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid X_1 = a_{j_1}u + v, A_{2,p}\right)\right. \\
&\qquad \left. - P\left(C_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid Y_1 = a_{j_1}u + v, A_{2,p}\right)\right| dF_{j_1}^{*}\left(v\right) \\
&\qquad + \left|F_{j_1}^{u}\left(y_1\right) - F_{j_1}^{*}\left(y_1\right)\right| + \left|F_{j_1}^{u}\left(x_1\right) - F_{j_1}^{*}\left(x_1\right)\right| \\
&\quad \leq \sup_{x_1 < v < y_1} \left|P\left(B_{2,p}^{u',\mathbf{x}_{2,p} - \mathbf{a}_{2,p}^{C}v,\mathbf{y}_{2,p} - \mathbf{a}_{2,p}^{C}v} \mid X_1 = u', A_{2,p}\right)\right. \\
&\qquad \left. - P\left(C_{2,p}^{u',\mathbf{x}_{2,p} - \mathbf{a}_{2,p}^{C}v,\mathbf{y}_{2,p} - \mathbf{a}_{2,p}^{C}v} \mid Y_1 = u', A_{2,p}\right)\right| \\
&\qquad + \left|F_{j_1}^{u}\left(y_1\right) - F_{j_1}^{*}\left(y_1\right)\right| + \left|F_{j_1}^{u}\left(x_1\right) - F_{j_1}^{*}\left(x_1\right)\right|
\end{aligned}$$

where $u' = a_{j_1}u + v$. Here the supremum of the first term in the last inequality over $\mathbf{r} < \mathbf{x} < \mathbf{y} < \mathbf{s}$ goes to zero as $u' \to \infty$ by the induction argument, while the suprema of the second and third terms tend to zero by Assumption 5.6. Hence (5.42) is proven.

So far we examined the probabilities of events bounded from both sides. But what happens when $x_i \to -\infty$? It is easy to see that if $\{Z\} = \{X\}$ or $\{Z\} = \{Y\}$,

$$\begin{aligned}
F_{j_i}^{*}\left(-\infty\right) \leq \overset{*}{\lim_{u\to\infty}} P\left(Z_i < a_{1,i}^{C}u - iM \mid Z_{i-1} \geq a_{1,i-1}^{C}u - \left(i-1\right)M, I_i = j_i\right) \\
\leq F_{j_i}^{*}\left(-\left(i\left(1 - a_{j_i}\right) + a_{j_i}\right)M\right)
\end{aligned}$$

because $\left(a_{1,i}^{C}u - iM\right) - a_{j_i}\left(a_{1,i-1}^{C}u - \left(i-1\right)M\right) = -\left(i\left(1 - a_{j_i}\right) + a_{j_i}\right)M$, where we used the notation $\lim^{*}$ for either $\liminf$ or $\limsup$. Hence

$$\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} P\left(Z_i < a_{1,i}^{C}u - iM \mid Z_{i-1} \geq a_{1,i-1}^{C}u - \left(i-1\right)M, I_i = j_i\right) = F_{j_i}^{*}\left(-\infty\right). \tag{5.43}$$

Trivially, the above statement also holds when the condition is $Z_{i-1} = a_{1,i-1}^{C}u - \left(i-1\right)M$, $I_i = j_i$.

(5.43) means that if $F_j^{*}\left(-\infty\right) = 0$ $\left(j = 0, 1\right)$ the X_t or Y_t process, starting from a relatively high region, will not reach a much lower region compared to its usual path in a single step with probability close to one. On the other hand, if $F_j^{*}\left(-\infty\right) > 0$, we obtain from (5.39) that for all $l \geq i + 1$, using the notation $d_{i,k} = i - \left(k - i\right)/\left(p - i\right)$,

$$\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} P\left(Z_l > a_{1,l}^{C}u - a^{l-i}d_{i,l}M \mid Z_{l-1} < a_{1,l-1}^{C}u - a^{l-i-1}d_{i,l-1}M, A_{1,p}\right) = 0$$

since

$$\left(a_{1,l}^{C}u - a^{l-i}d_{i,l}M\right) - a_{j_l}\left(a_{1,l-1}^{C}u - a^{l-1-i}d_{i,l-1}M\right) \geq a^{l-i}M/\left(p-i\right).$$

Thus once the process has reached a low level compared to its usual sample path, it will not get back to a much higher region with probability close to one:

$$\lim_{M\to\infty}\overset{*}{\lim_{u\to\infty}} P\left(\exists l : i+1 \leq l \leq p, Z_l > a_{1,l}^{C}u - a^{l-i}d_{i,l}M \mid Z_i < a_{1,i}^{C}u - iM, A_{1,p}\right) = 0. \tag{5.44}$$

(5.43) and (5.44) together imply

$$\begin{aligned}\lim_{M\to\infty}\overset{*}{\lim_{u\to\infty}} \big| P\left(B_{1,p}^{u,-\infty,\mathbf{y}} \text{and} \quad \exists i : X_i < a_{1,i}^{C}u - iM \mid X_0 = u, A_{1,p}\right)\\ - P\left(C_{1,p}^{u,-\infty,\mathbf{y}} \text{and} \quad \exists i : Y_i < a_{1,i}^{C}u - iM \mid Y_0 = u, A_{1,p}\right)\big| = 0.\end{aligned}$$

Finally, the combination of this equation and (5.42) with the choice $x_i = -iM$ $(i = 1,\ldots,p)$ yields the statement in the Proposition. □

Proposition 5.26 suggests approximating X_t above sufficiently high thresholds with Y_t, a Markov-switching autoregressive process. One of the regimes in Y_t is a random walk and the other may be a random walk (if $a = 1$) or a stationary autoregression (if $0 < a < 1$).

For instance, if X_t is precisely a two-state MS-AR(1) process with autoregressive parameters 1 and $0 < a < 1$, respectively, then this extremal approximation is certainly exact. As another extreme example, if X_t is conditionally independent given I_t, then $\varepsilon_{1,t} = \varepsilon_{0,t} = -\infty$ with probability 1 in the limiting representation.

Adapting Appendix 1 of Smith et al. (1997), one could explicitly calculate $F_j^*(z)$ if the joint distribution of $((X_{t-1}, X_t)|I_t = j)$ were in the domain of attraction of specific bivariate extreme value laws such as the logistic, bilogistic, negative bilogistic or asymmetric ones. For instance, $F_j^*(z)$ are proper distribution functions in all these cases apart from the negative bilogistic one.

The given extremal approximation for Markov-switching, conditionally Markov processes generalises the idea presented in section 2.2.3 where simple Markov chains with asymptotically exponential marginal distributions are modelled as a random walk above high thresholds. In fact, if $0 < a < 1$ the conditionally Markov processes are asymptotically still Markov chains because the dominant regime determines their behaviour at "very high" thresholds. Since in this case

$$\begin{aligned}P(X_t > u + x|X_{t-1} = u, I_t = 0) &\approx P(aX_{t-1} + \varepsilon_{0,t} > u + x|X_{t-1} = u)\\ &= P\left(\varepsilon_{0,t} > \left(1-a\right)u + x\right) \to 0\end{aligned} \tag{5.45}$$

for all x as $u \to \infty,$ the asymptotic step distribution function of the "limiting" Markov chain takes $-\infty$ with probability $p_1 + (1 - p_1)F_1^*(-\infty)$. (Here p_1 is the probability of switching to the dominated regime and the other term is the probability of jumping in the limit to $-\infty$ while staying in the dominant regime.) By taking into account the dominated regime, the two-state model gives more insight into the subasymptotic behaviour of a Markov-switching conditionally Markov process than the simple Markov chain representation.

We should note however that, similarly to the original Markov chain methodology for extremes (Smith (1992), Smith et al. (1997)) and to its generalisations (e.g. Bortot and Tawn (1998) or Sisson and Coles (2003)), our motivation comes from statistical modelling. In statistical practice the regularity conditions behind Assumption 5.6 are not considered restrictive thus Proposition 5.26 gives an appropriate basis for modelling extremes of Markov-switching conditionally Markov processes with exponential tail.

5.5 Model estimation and simulation

Assume that a process exactly follows a Markov-switching autoregressive structure defined in (5.1)-(5.2) (or, more generally, a regime switching AR(1) model with negative binomial regime durations, see section 5.6) and $\varepsilon_{j,t}$ $(j = 0, 1)$ have parametric distributions. Then, theoretically, model parameters can be estimated by maximum likelihood – although in practice, due to computational reasons, Markov Chain Monte Carlo (MCMC) algorithms are needed even for not too large sample sizes. In Vasas et al. (2007) we present and compare two MCMC estimation schemes for the more general regime switching AR(1) model: the I_t-s serve as auxiliary parameters in the first approach, while the change points of the regimes do the same in the second, reversible jump framework. (Reversible jump MCMC methods were introduced by Green (1995) and allow for the change of the dimension of the parameter space.)

In this dissertation we do not go into details of MCMC estimation methods. Instead, we develop a framework that is more suitable for extreme value analysis. Suppose that X_t is a Markov-switching, conditionally Markov process (which is a much weaker assumption than the MS-AR one) with exponential marginals and suppose in the spirit of threshold methods that we observe data only above a high threshold $u.$ The aim is to estimate the extremal dependence structure of the process based on these high-level observations, which we model according to Proposition 5.26 by a Y_t Markov-switching autoregressive process. Compared to fitting an MS-AR model to the whole series, this approach has the advantage that it uses only the high-level exceedances for the estimation of extremes (thus it is unaffected by potentially different dynamics at normal levels) and at the same time has a theoretical foundation. Furthermore, estimation is much less computationally intensive compared to the MCMC scheme.

So, assume that u is high enough for the approximation of X_t by Y_t to be valid and that the following condition holds.

Assumption 5.7. $0 < a < 1$, $F_1^*(0) = 0$ *and the distributions of* $\varepsilon_{j,t}$ $(j = 0, 1)$ *are absolutely continuous with respect to the Lebesgue-measure with density function* $h_j(z)$.

This condition is not crucial for the estimation procedure but makes the interpretation easier. Since $\varepsilon_{1,t} \geq 0$ a.s. in this case the $I_t = 1$ regime can be called asymptotically the "ascending" regime and the other regime – because of $0 < a < 1$ and (5.45) – the "descending" one.

If all data were observed and the whole process followed a Markov-switching autoregression (5.40)-(5.41), the likelihood function would just be the product of the individual conditional likelihood terms $f_t = f(Y_t|Y_{t-1}, \ldots, Y_1)$:

$$L(Y_1, Y_2, \ldots, Y_n) = f(Y_1) \prod_{t=2}^{n} f_t \tag{5.46}$$

and f_t could be calculated easily by the following recursion. Let

$$r_t = P(I_t = 1|Y_t, Y_{t-1}, \ldots, Y_1)$$

denote the conditional probability of belonging to the dominant regime at time t given all observations up to time t, then

$$\begin{aligned}
r_{1,t} &= P(I_t = 1|Y_{t-1}, \ldots, Y_1) = (1 - p_1)r_{t-1} + p_0(1 - r_{t-1}) \\
f_{1,t} &= f(Y_t, I_t = 1|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = r_{1,t}h_1(Y_t - Y_{t-1}) \\
f_{0,t} &= f(Y_t, I_t = 0|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = (1 - r_{1,t})h_0(Y_t - aY_{t-1}) \\
f_t &= f(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = f_{0,t} + f_{1,t} \\
r_t &= f_{1,t}/f_t
\end{aligned}$$

and the starting values r_1 and $f(Y_1)$ do not influence the estimates in large samples. The resulting maximum likelihood estimator is consistent, see Francq and Roussignol (1998).

As in the Markov chain case in section 2.2.3 due to censoring we only observe (Z_t, δ_t) where $Z_t = \max(Y_t, u)$ and $\delta_t = \chi_{\{Y_t > u\}}$. The aim is to derive an approximation of the likelihood (5.46) based only on (Z_t, δ_t) . Then, if $Y_{t-1} > u$ and $Y_t > u$ both Y_{t-1} and Y_t are known and hence the same recursion as above can be applied. However, approximations are needed in the other three cases.

When $Y_{t-1} \leq u$ and $Y_t > u$ two approximations are used to determine f_t. First, since $P\left(I_t = 1 | Y_t > u, Y_{t-1} \leq u\right) \to 1$ as $u \to \infty$, it is assumed in this case that $r_t \approx 1$. Second, we cannot observe $Y_t - Y_{t-1}$, hence the distribution of $Q = (Y_t - u | Y_t > u, Y_{t-1} \leq u)$ has to be approximated. It is easy to show that if $0 < a < 1$ holds (Assumption 5.7) then

$$P\left(I_t = 1, I_{t-1} = 1, \ldots, I_{t-m} = 1 | Y_t > u, Y_{t-1} \leq u\right) \to 1 \tag{5.47}$$

for any fixed m as $u \to \infty$ (i.e. a long $I_t = 1$ regime is needed for the process to cross a high level u). Thus at the time of reaching u the process behaves similarly to S_n defined in (5.8) where ε_n has density function $h_1(x)$. Therefore, Q is approximately distributed as the limiting overshoot of S_n, which was denoted by B_∞ in Lemma 5.6. Since the nonnegativity of $\varepsilon_{1,t}$ implies $B_0 = \varepsilon_1$, (5.9) yields that the probability density of Q is approximately

$$f_Q(y) = \bar{F}_{\varepsilon_{1,t}}\left(y\right) / E\varepsilon_{1,t}. \tag{5.48}$$

(If $\varepsilon_{1,t}$ is e.g. exponentially distributed, Q is also exponential because of the constant hazard property of the exponential distribution.) Taking together the two approximations, we obtain $f_t \approx f_Q(Y_t - u)$ when $Y_{t-1} \leq u$ and $Y_t > u$.

When $Y_{t-1} > u$ and $Y_t \leq u$ we get $r_t = 0$ because $\varepsilon_{1,t} \geq 0$ a.s. Thus $P\left(Y_t \leq u | Y_{t-1}\right) = P\left(\varepsilon_{0,t} \leq u - aY_{t-1}\right)$, hence $f_t \approx \int_{-\infty}^{u-aY_{t-1}} h_0(y)dy$.

Finally, when $Y_{t-1} \leq u$ and $Y_t \leq u$ we simply take $f_t \approx P\left(Y_t \leq u\right)$, which is a reasonable approximation for most of the sample. By Assumption 5.5, $P\left(Y_t \leq u\right)$ depends on K_1 and K_0 which do not enter the approximate likelihood at other places, hence these terms do not influence the maximum likelihood estimation of the structural parameters.

Having obtained the approximate likelihood (which is a function of (Z_t, δ_t) only) the maximum likelihood estimates of the parameters can be calculated. It follows from the approximations that when u is not high enough a bias may appear even at large samples, but it certainly tends to zero as $u \to \infty$. Moreover, the smaller the parameters a and p_0 are, the smaller the estimation bias is.

To illustrate the performance of the approximate likelihood estimator, consider an X_t Markov-switching autoregressive process where the underlying Markov chain is governed by $p_1 = 0.6$ and $p_0 = 0.025$ transition probabilities and the two regimes are characterised by $a = 0.8$, $\varepsilon_{1,t} \sim \text{Exp}\left(\lambda\right)$ and $\varepsilon_{0,t} \sim \text{N}\left(0, \sigma^2\right)$ with $\lambda = 1$ and $\sigma = 0.5$. (Apart from a scaling factor, these parameters roughly correspond to the estimates obtained for the river discharge data set in section 5.6.) Let us examine the parameter estimates resulting from the approximate likelihood as a function of the threshold u. The threshold ranges from the 95% to the 99.9% quantile of the marginal distribution of the process. Figure 5.3 shows that the parameters of the dominated regime, p_0, a and σ are essentially unbiased even at

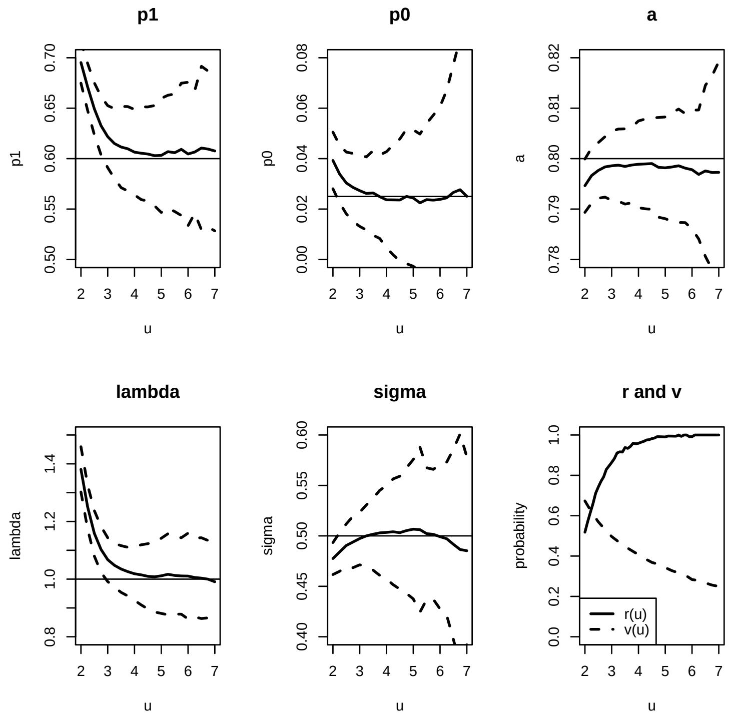


Figure 5.3: Parameter estimates (and approximate 95% confidence bands) as functions of the threshold u for a Markov-switching AR(1) process with original length 100000. The thresholds range from the 95% to the 99.9% quantile of the marginal distribution. The horizontal lines show the true parameter values. The lower right panel displays the probabilities $r(u)$ and $v(u)$.

reasonably small thresholds while the p_1 (transition probability) and λ (scale) parameters of the dominant regime are both overestimated for moderate u values – though with a vanishing margin as $u \to \infty$. (Their bias essentially disappears at around $u = 4$, the 99.3% quantile of the marginal distribution.)

The lower right panel of the figure displays two probabilities. The first is

$$r(u) = P\left(I_t = 1 | X_t > u, X_{t-1} \leq u\right),$$

which is an approximation of r_t at the time of reaching u, hence it is not surprising that the distance of $r(u)$ from 1 strongly determines the bias of p_1 and λ. The other probability shown in the figure is $v(u) = P(I_t = 0 | X_t > u)$, which certainly tends to zero as $u \to \infty$, but this convergence is very slow. There is a wide range of thresholds where the bias of the variables are negligible but $v(u)$ is still far from zero. These are the thresholds of particular interest: the parameter estimation gives reliable results but the dominated regime plays a substantial role in determining the behaviour of the process at such levels.

The choice of an appropriate u constitutes a bias-variance problem often encountered in extreme value analysis: an increase of the threshold reduces the estimation bias but raises the variance by lowering the effective sample size. As an illustration, Figure 5.3 also shows the calculated 95% confidence intervals of the parameters estimated on the basis of various threshold exceedances of the Markov-switching autoregressive process with original length 100000. Similarly to e.g. the usual diagnostic check of fitting a GPD to i.i.d. exceedances (see page 19), a simple way to find a suitable threshold is to select one above which the parameter estimates look roughly constant. An alternative solution would be to directly estimate $r(u)$ from the sample but this seems to be complicated as only the large values are observed.

The final aim of analysing the extremal dependence structure of the Markov-switching, conditionally Markov model is to describe the behaviour of its limiting extremal cluster functionals denoted by C^* in section 2.2.2 (Theorem 2.20). After estimating p_1, p_0, a and the parameters of $h_1(x)$ and $h_0(x)$, an extremal cluster can be simulated straightforwardly from the autoregressive approximation as follows. By (5.47) we can assume that $I_t = 1$ at the start of a cluster and thus the first value above a high threshold u is distributed as Q in (5.48). Then we simulate the Markov chain I_t and the process Y_t according to (5.40)-(5.41) until the process decreases sufficiently below u. Finally, we calculate the desired extremal cluster functional.

5.6 Application to water discharge data

Figure 3.1/a already suggested why Markov-switching models – together with shot noise processes and neural networks – are among the most widely used tools to study hydrological phenomena. That figure shows the pulsatile nature of the water discharge series: short but steep ascending periods are followed by longer, gradually descending ones. Hydrological evidence suggests that the two periods are governed by completely different physical phenomena (Jain and Srinivasulu, 2006), pointing to a Markov-switching (or generally, a regime switching) process.

Among the various switching models presented in the literature two recent articles deserve most attention from the point of view of this dissertation. Lu and Berliner (1999) develop a three-state Markov-switching AR(1) model where the innovations are normally distributed in each state and the lagged precipitation is included as a regressor. They estimate the model by MCMC methods and find that it reproduces the usual features of river flow series reasonably well.

However, in their model the inclusion of precipitation – which is not available to us – is necessary to reproduce the pulsatile nature of the series. In the absence of an exogenous trigger such as precipitation the easiest way to model the pulsatility of river flow data in the regime switching autoregressive framework is to allow a.s. positive increments in one of the regimes. Szilágyi et al. (2006) propose a two-state Markov-switching model where in the first, "random walk-type" regime the increments are Weibull-distributed and are ranked in increasing order within the regime (i.e. always the larger increments occur at the end of the regime). In the second regime the process decays exponentially, without a random term. Since the first regime contains the increasing periods and the second the decreasing ones identification of the regimes is very easy, however, statistically appropriate estimation is impossible on real data because the decay is never deterministic in reality. Nevertheless, the model (estimated on ad hoc grounds) performs well in practical hydrological simulations.

In Vasas et al. (2007) we introduce a generalisation of the above model[5], which is also a slight generalisation of the MS-AR model discussed in this chapter. We define a regime switching AR(1) process (5.1)-(5.2), where regime durations are independent and the duration of the random walk regime is negative binomially distributed[6] with parameter (b_1, p_1) , while the duration of the stationary regime is Geom (p_0)-distributed. Since the MS-AR model has independent and geometrically distributed regime durations, it is obtained as a special case by choosing $b_1 = 1$. The noise in the random walk regime is assumed to follow a $\Gamma(\alpha, \lambda)$ distribution, while the noise in the stationary regime is Gaussian with zero mean and σ^2 variance.

In Vasas et al. (2007) we obtain MCMC parameter estimates for the water discharge series at Tivadar and show that synthetic water discharge series simulated from the fitted model reproduce the empirical features (such as the probability density) of the observed series reasonably well. As in section 4.6 with the ARCH-type model, one could also make inference on the extremes based on these simulations. (Moreover, the theoretical tail behaviour and extremal clustering could also be investigated with methods similar to those used in this chapter for the MS-AR model.) However, when the extremal behaviour is of interest, the elegant threshold method developed in section 5.5 is a more suitable estimation technique because it relies only on high-level exceedances, moreover, it assumes only the

[5] apart from the ranking condition of the increments in the rising regime

[6] The negative binomial distribution $\{q_{k,1}\}$ $(k \geq 1)$ with parameter (b_1, p_1) is defined as:

$$q_{k,1} = \frac{\Gamma(k + b_1 - 1)}{\Gamma(k)\Gamma(b_1)} p_1^{b_1} (1 - p_1)^{k-1}, \tag{5.49}$$

i.e. regime durations can be positive integers and their distribution is IFR (increasing failure rate) if $b_1 > 1$, DFR (decreasing failure rate) if $b_1 < 1$ and geometric if $b_1 = 1$.

conditional Markov structure (and not the conditional AR structure) of the whole series. If the switching AR model is estimated on the whole series instead of on the exceedances, the measures of extremal clustering may be biased due to model misspecification. Misspecification may occur e.g. because the expectation of the increments is threshold-dependent – a hydrological phenomenon captured by ranking the increments in Szilágyi et al. (2006). Therefore we focus on the threshold method in the sequel.

As a price for using the threshold method one should stick to the Markov-switching regime structure (i.e. geometrically distributed durations) instead of the more general negative binomial framework. However, as far as extremes are concerned, this is not a very strong restriction for the following reason. If we have a conditionally Markov model with negative binomial regime durations where Assumption 5.7 holds, it is still true that a long $I_t = 1$ regime is needed to reach a high threshold (equation (5.47)). Therefore, since an N negative binomially distributed random variable with parameter (b_1, p_1) has the property that $(N - k) \,|\, (N > k) \to_d \text{Geom}\,(p_1)$ as $k \to \infty$, the duration of the $I_t = 1$ regime can be approximated as geometric above high thresholds, providing some theoretical justification of the Markov-switching limiting representation Y_t even when the ascending regime durations are negative binomially distributed.

Application of the threshold method

The aim is to analyse the distributions of flood peaks, flood durations and flood volumes in the water discharge data set measured at Tivadar. (In hydrological practice, flood peaks, durations and volumes together describe the severity of a flood event.) More precisely, we ultimately seek to determine a value x such that all flood volumes (or flood durations or flood peaks) in the next n years (e.g. $n = 50$) will lie below x with a certain pre-specified probability q. (The value of q is close to one.) If we assume – in the spirit of Theorem 2.20 – that the number of floods in the coming n years (denoted by K) is Poisson-distributed with parameter μ and the floods are independent of each other, x can be given easily in terms of the quantile function of the corresponding cluster functional distribution during a particular flood (C') since

$$q = P\left(\max\left(C'_1, C'_2, \dots, C'_K\right) \le x\right) = \exp\left(-\mu\left(1 - P\left(C' \le x\right)\right)\right). \tag{5.50}$$

The threshold defining a flood is chosen as $u_0 = 1250 \text{ m}^3/\text{s}$, which is the 99.3% quantile of the marginal distribution of the river flow series and roughly corresponds to the water height of the first level of preparedness in the flood alert system. As an operational definition – which slightly differs from the usual declustering procedures – we regard two floods distinct if there is at least one day when the water discharge goes below a $u < u_0$,

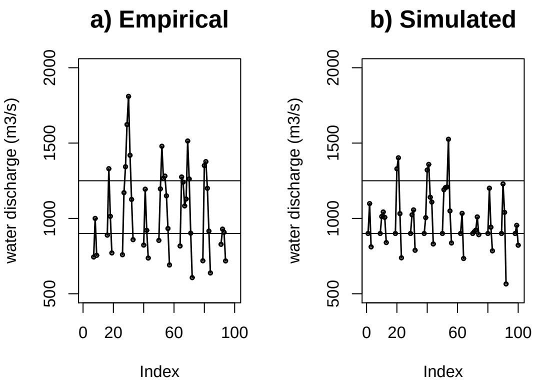


Figure 5.4: A few typical observed and simulated flood scenarios

say $u = 1050\ \mathrm{m}^3/\mathrm{s}$ auxiliary threshold between them. The hydrological reason behind this scheme lies in allowing the natural reservoirs to give away excess water during the inter-flood periods. Note however, that the precise definition of flood does not influence much the results given below.

Figure 5.4/a provides a rough picture about the shape of a few observed flood scenarios above the threshold while the upper row in Figure 5.5 displays the histogram of flood peaks, flood durations and flood volumes.

A variety of studies has examined the distribution of flood peaks and modelled them in line with extreme value theory by GPD. (In fact, it follows from section 3.2 that the GPDs fitted to flood peaks of medium and larger rivers tend to be close to the exponential distribution.) A few papers have also investigated flood duration and flood volume distributions, but they have usually chosen the parametric family used in the analysis on ad hoc grounds. (Nonparametric modelling is rarely feasible because of the small sample size: e.g. in the case at Tivadar there are only 48 flood events.) Anderson and Dancy (1992) proposed a Weibull distribution for aggregate excesses, while Grimaldi and Serinaldi (2006) applied Gamma distribution for them and lognormal for the durations. Nevertheless, when the use of a particular parametric family is to be justified, one has to provide a dependence structure asymptotically yielding that family for the distribution of the extremal cluster functional. Threshold methods are tailor-made for this purpose.

In view of section 2.2.3 the first idea for threshold-based modelling of the water discharge data is to assume a Markov chain structure, estimate the transition density based on the censored observations and use the random walk approximation of Smith et al. (1997) for simulation. However, a simple analysis reveals that the river flow data cannot be treated as Markov even above high thresholds. Denoting the time series by $X_t,$ Figure 5.6 displays

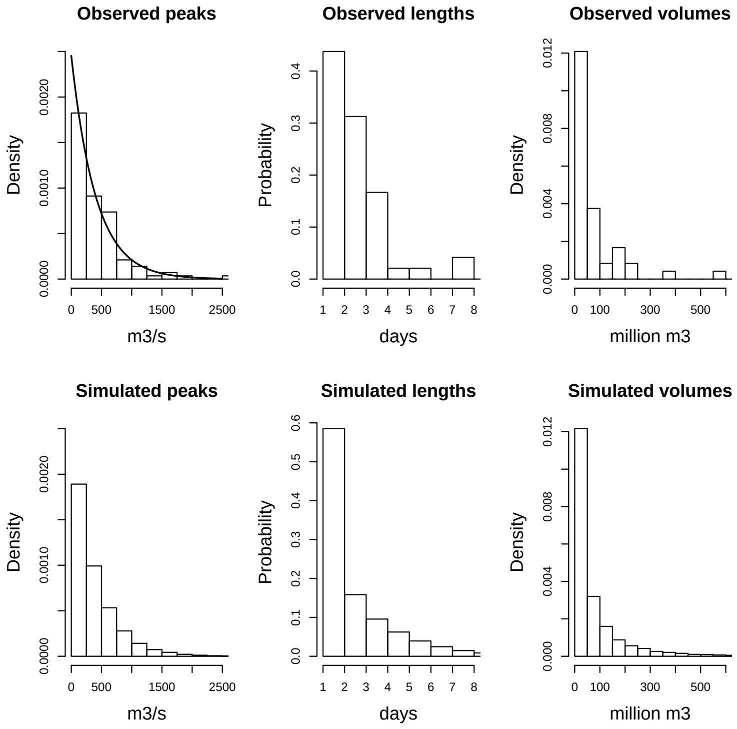


Figure 5.5: Histograms of observed and 50000 simulated peaks, durations and volumes above $u_0 = 1250\ \mathrm{m}^3/\mathrm{s}$, with auxiliary threshold $u = 1050\ \mathrm{m}^3/\mathrm{s}$. Observed and simulated peaks are close to exponential.

the plots of $X_t - X_{t-1}$ against X_{t-1} if $X_{t-1} - X_{t-2}$ is larger or smaller than zero, respectively. Although the figures only show the cases when X_{t-1} is larger than the 98% quantile of the marginal distribution, the two plots are not similar, indicating that the series is not first order Markov even above this high threshold.

Being more general, our threshold method assumes that the data come from a Markov-switching, conditionally Markov model. The approximation in section 5.4 then suggests that the process behaves asymptotically as a random walk in one regime and as a stationary autoregression (or perhaps as another random walk) in the other one. In our case, because of the markedly ascending ("pulsatile") nature of the process in the first regime we may assume that $\varepsilon_{1,t} \geq 0$ a.s., and because of the "descending" nature in the other regime we may assume $0 < a < 1$. (The justification also comes from the previously mentioned hydrological paper Szilágyi et al. (2006).) So in this case the dominant regime can be called the "ascending", and the dominated regime – at extreme levels – the "descending" one.

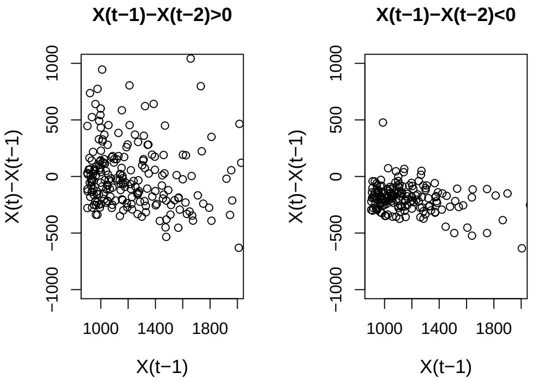


Figure 5.6: Plot of the increments against the previous day's discharge values above the 98% quantile, conditioned on the sign of the previous day's increment

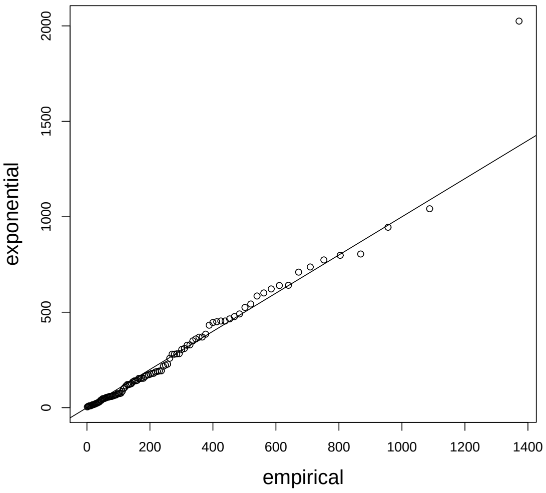


Figure 5.7: Exponential QQ-plot of the positive increments above the threshold $u = 1050\,\mathrm{m}^3/\mathrm{s}$

It also follows that a first approximation to the distribution of $\varepsilon_{1,t}$ can be given by examining the positive increments measured above a high threshold u, i.e. $(X_t - X_{t-1})|(X_t - X_{t-1} > 0, X_{t-1} > u)$. Figure 5.7 displays the exponential QQ-plot of these *increments* above $u = 1050\,\mathrm{m}^3/\mathrm{s}$, indicating that their distribution is close to exponential (apart from an outlier). Thus we assume that $\varepsilon_{1,t}$ is Exp (λ)-distributed, and – as a standard condition – $\varepsilon_{0,t}$ is normally distributed with zero mean and σ^2 variance. Hence Assumption 5.7, needed for the estimation procedure of section 5.5, is satisfied in this case.

Following this procedure, we fit the extremal model with parameters p_0, p_1, a, λ and σ using thresholds ranging from $u = 500$ to 1800 m^3/s (or from the 90% to the 99.9% quantile of the marginal distribution). Figure 5.8 shows the parameter estimates as functions of u. We did not display p_0 because it lies below 0.05 irrespective of the threshold and thus

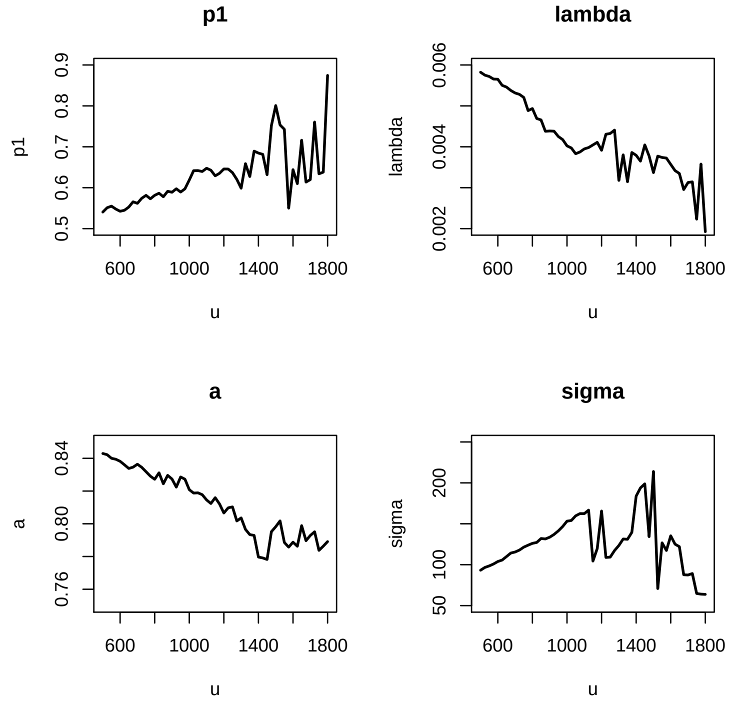


Figure 5.8: Parameter estimates for the water discharge series as functions of the threshold

it only slightly affects extremal clustering. (Because of the low value of p_0 nearly all observed dominated regimes above a high threshold are censored.) The estimate of p_1 seems to be constant from around 1050 m^3/s (the 98,7% quantile), while λ and σ do not have a strong trend from that threshold until around 1500 m^3/s (above which the sample size is less than 50 hence the estimates become very imprecise). The only parameter steadily decreasing is a but its whole range is narrow enough not to alter substantially the results.

Therefore, we have chosen the threshold $u = 1050\ \mathrm{m}^3/\mathrm{s}$ and show the resulting maximum likelihood parameter estimates along with their asymptotic standard errors in Table 5.1. Above this threshold, the average duration of an ascending regime is $1/p_1 = 1.56$ days and the average increment is $1/\lambda = 261.1\ \mathrm{m}^3/\mathrm{s}$. The persistence is quite high even in the stationary regime with $a = 0.819$. The value of p_0 is estimated very imprecisely because (nearly) all stationary regimes above the threshold are censored.

Table 5.1: Parameter estimates with standard errors

Parameter	p_1	p_0	$\lambda^{-1}(\mathrm{m}^3/\mathrm{s})$	a_0	$\sigma(\mathrm{m}^3/\mathrm{s})$
ML-estimate	0.642	0.0289	261.1	0.819	159.7
Standard error	0.045	0.0168	22.3	0.009	12.1

To give an impression of the model, Figure 5.4 also displays a few simulated flood scenarios. (Simulations start at u but the flood definition does not change: only floods above the first level of preparedness, $u_0 = 1250\ m^3/s$ are taken into account when calculating the extremal cluster functionals.) The shapes of the simulated and observed floods are similar. The average simulated flood duration above u_0 is 2.71 days, so the process in the dominated ("descending") regime remains above the threshold for more than one day on average after the peak. Since under the asymptotic structure the process would fall immediately below the threshold after reaching the peak, the subasymptotic component has an important effect on clustering tendencies at this level.

The lower row in Figure 5.5 shows the histogram of 50000 simulated flood peaks, flood durations and flood volumes above the threshold u_0. Although the probability of the one day long floods is significantly higher in the simulated flood duration distribution than in the observed one and thus a formal χ^2-test slightly rejects the fit of the two distributions, the observed and simulated averages (2.71 days) are equal. The goodness of fit is appropriate in the case of the peaks and volumes, which is illustrated in Figure 5.9 by the QQ-plots of the observed quantities with respect to their simulated counterparts. Since the model-based flood peaks can be given approximately as Geom (p_1) random sums of i.i.d. Exp (λ)-distributed variables, they are approximately exponentially distributed with mean $1/(\lambda p_1) = 406.7\ \mathrm{m}^3/\mathrm{s},$ in accordance with the observed data.

The flood volumes are clearly heavier tailed than the exponential distribution. In fact, since the threshold model is just an MS-AR(1) process above high thresholds, it follows from Theorem 5.24 that the limiting aggregate excess distribution in this model has approximately Weibull-like tail. Hence, the fit of the model to water discharge data also gives some theoretical justification to the method of Anderson and Dancy (1992), who proposed to approximate the aggregate excesses of hydrological data sets by Weibull distributions.

Based on the simulated flood duration and flood volume distributions, return values can be obtained for these quantities. We use e.g. $\mu = 48$ and $q = 0.95$ in (5.50) for the 50 years, 95% return value, and obtain a point estimate of $x = 1370$ million m^3 for flood volume and $x = 14$ days for flood duration. This means that, for instance, the chance of a flood volume greater than 1370 million m^3 in the coming 50 years is approximately 5%. Note that the highest observed flood volume in the last fifty years was about 570 million $\mathrm{m}^3/\mathrm{s},$ which corresponds to the 50 years, 44% return value according to the simulation-based flood volume distribution. The longest flood in the last fifty years took eight days, which corresponds to the model-based 50 years, 30% return value in the flood duration distribution.

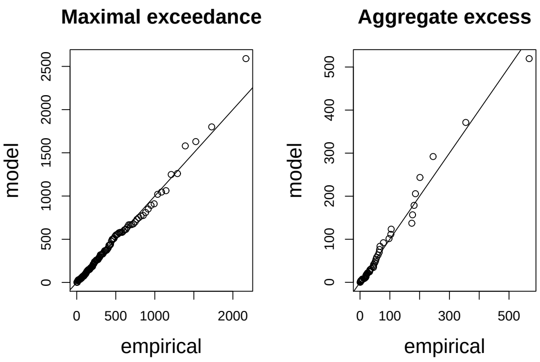


Figure 5.9: QQ-plots of observed flood peaks and volumes with respect to their simulated counterparts

5.7 Summary

Within the field of the analysis of Markov-switching structures, this chapter gave novel results in two areas. First, we examined the tail behaviour and extremal clustering of a certain Markov-switching autoregressive model which was not examined previously in the literature. In particular, we proved under general conditions that the tail of the model is asymptotically exponential, and obtained its extremal index in terms of a Wiener-Hopf integral equation, which can be solved explicitly under further restrictions. We also showed that the tail of the limiting aggregate excess distribution can be approximated by Weibull-like distributions if the increments in the first regime are Gamma-distributed.

Second, we proved that extremes of certain Markov-switching, conditionally Markov models can be approximated by the Markov-switching AR(1) model examined previously, and developed a threshold-based method to fit the MS-AR(1) model. Finally, we used this approach for inference on extremal cluster functionals of real water discharge data.

Chapter 6

Conclusions

In this dissertation we analysed two classes of models, an ARMA-ARCH-type process and a Markov-switching autoregressive process, and applied both of them to real hydrological data. Although both models proved useful to some degree in practical applications, it should be asked at the end of the day which model is more valuable. In our opinion, the MS-AR(1) model outperforms the ARMA-ARCH-type process both in its "physical" interpretability and in its ability to describe the extremes of hydrological data sets (asymptotically exponential vs. Weibull-like tail; smaller than one extremal index vs. clustering only at subasymptotic levels). This does not mean, however, that the conditional heteroscedasticity idea behind the ARMA-ARCH-type model should not be used in further analyses. On the contrary, its successful application in hydrology suggests that it may prove practically useful beyond its traditional fields of finance and economics.

To make a connection between the two models we show that the ARMA-ARCH-type model may be viewed as a statistical approximation to the MS-AR(1) process. It follows from Karlsen (1990) or Zhang and Stine (2001) that an MS-AR(1) model with k distinct regimes has a weak ARMA (p, q)-representation where $p \leq k$ and $q \leq k-1$, hence the model examined in chapter 5 is actually a weak ARMA(2,1) process.[1] Furthermore, the innovation of this weak ARMA process (denoted in the sequel by w_t) has an interesting property. By the uniqueness of Wold's decomposition w_t is the error of the linear projection of X_t on the space spanned by $\{X_s : s \leq t-1\}$, but according to Yang (2000, Theorem 4) it can also be given in our MS-AR setting as $w_t = X_t - E\left(X_t | \mathcal{G}_{t-1}\right)$, where

[1] Zhang and Stine (2001) define the model as $X_t = b_{I_t} + a_{I_t}\left(X_{t-1} - b_{I_{t-1}}\right) + \sigma_{I_t} V_t$, where V_t is a zero mean uncorrelated process, independent of the I_t Markov chain, and b_j, a_j are real numbers and $\sigma_j > 0$. With similar notations our model can be written as $X_t = b_{I_t} + a_{I_t} X_{t-1} + \sigma_{I_t} V_t$, where $b_j = E\left(\varepsilon_{j,t}\right)$ and $\sigma_j = D\left(\varepsilon_{j,t}\right)$ for $j = 0, 1$, hence it is not exactly of the same form as in Zhang and Stine (2001). However, the proofs remain valid with slight modifications, thus the weak ARMA representation still holds.

$\mathcal{G}_t$ is the σ-field generated by $\{X_s, I_s : s \leq t\}$. Obviously, $E\left(w_t | \mathcal{G}_{t-1}\right) = 0$, and the $E\left(w_t^2 | \mathcal{G}_{t-1}\right)$ conditional variance is a function of only X_{t-1} and I_{t-1}. Using the notations $p_{ji} = P\left(I_t = i | I_{t-1} = j\right), b_j = E\left(\varepsilon_{j,t}\right), \sigma_j^2 = D^2\left(\varepsilon_{j,t}\right)\ (i, j = 0, 1)$ and $a_1 = 1$, $a_0 = a$ we obtain in our model that

$$\begin{aligned} d_j^2\left(x\right) : &= D^2\left(w_t | X_{t-1} = x, I_{t-1} = j\right) = D^2\left(X_t | X_{t-1} = x, I_{t-1} = j\right) \\ &= \sum_{i=0}^{1} p_{ji}\left(\sigma_i^2 + \left(a_i x + b_i\right)^2\right) - \left(\sum_{i=0}^{1} p_{ji}\left(a_i x + b_i\right)\right)^2 . \end{aligned}$$

Since $d_j^2(x)$ is a quadratic function of x for both js the weak ARMA representation of the MS-AR(1) process exhibits conditional heteroscedasticity. The variance conditioned on X_{t-1} is given by

$$d^2\left(x\right) := D^2\left(w_t | X_{t-1} = x\right) = \sum_{j=0}^{1} d_j^2\left(x\right) P\left(I_{t-1} = j | X_{t-1} = x\right).$$

As $\lim_{x \to \infty} P\left(I_{t-1} = 1 | X_{t-1} = x\right) = 1$, $d^2(x)$ is asymptotically quadratic in x, thus the weak ARMA-ARCH representation of the process does not belong to the weak type of the framework of chapter 4. (There the conditional variance should be proportionate to $x^{2\beta}$ with $0 < \beta < 1$.) Nevertheless, due to the behaviour of $P\left(I_{t-1} = 1 | X_{t-1} = x\right)$ it may be possible for certain parameter choices of the MS-AR(1) process that $d^2(x)$ is approximately linear for a wide range of x values and thus the statistical fit of the ARMA-ARCH model with linear variance (as in section 4.6) may be appropriate. This is illustrated in Figure 6.1 where $d_0^2(x)$, $d_1^2(x)$ and $d^2(x)$ are plotted for an MS-AR(1) model with parameters estimated by MCMC for the river flow data at Tivadar, and approximate linearity is observed for values between about 200 and 1500.

Hence it is not surprising that if the true model is the MS-AR(1) process a statistical model fitting procedure may lead first to a weak ARMA(2,1) model (as at a few stations in section 3.1.1) and then to an ARMA-ARCH representation with linear conditional variance. It should be stressed, however, that this ARMA-ARCH representation is not a proper data generating process because even if the conditional variance were specified correctly w_t could not be written in the form $w_t = \sigma\left(X_{t-1}\right) Z_t$ with i.i.d. Z_t noise.

Beyond the theoretical results on the extremes, one of the main messages of the dissertation is that the combination of extreme value tools with time series methods may allow more insight into the extremal behaviour of processes than extreme value procedures alone would do. As illustrated especially nicely with the application of the threshold-based Markov-switching model in section 5.6, extremal clusters can be estimated more precisely if some – but not very specific – information on the time dependence of the data is utilised.

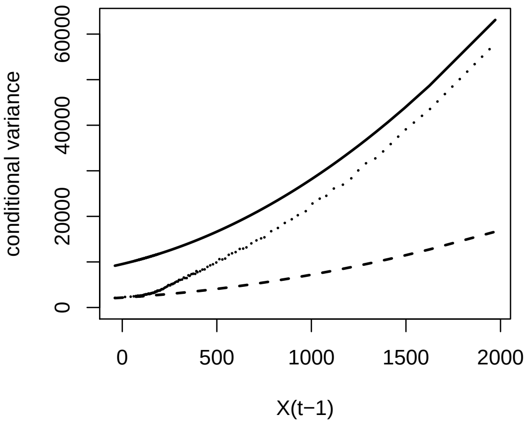


Figure 6.1: The functions $d_1^2(x)$ (continuous), $d_0^2(x)$ (dashed) and $d^2(x)$ (dotted line) for the MS-AR(1) model estimated for the water discharge data set

This combined approach is not constrained to hydrological applicatons. The same philosophy – or even a version of the Markov switching model presented here – could be helpful in examining extremal phenomena of other series e.g. in biology (endocrinology), macroeconomics or energy market analysis.

Another related question arises from a multivariate perspective. Similarly to the limiting cluster size or limiting aggregate excess distributions in the univariate time series case, there does not exist a unique parametric family to describe multivariate extreme value laws. Therefore, one might ask whether the extremal analysis of appropriate models (e.g. multivariate Markov-switching processes) could suggest a reasonable parametrisation for the joint occurence of e.g. extreme water discharges at different monitoring stations. With such an approach the understanding of multivariate extreme events could be improved.

Bibliography

Abramowitz, M. and Stegun, I. A., 1965. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55.

Anderson, C. W., 1990. Discussion of paper by A. C. Davison and R. L. Smith. J. R. Statist. Soc. Ser. B 52, 425-426.

Anderson, C. W. and Dancy, G. P., 1992. The severity of extreme events. Research Report 409/92, University of Sheffield.

Arató, M., Bozsó, D., Elek, P. and Zempléni, A., 2008. Forecasting and simulating mortality tables. Math. Comput. Modelling, accepted for publication.

Asmussen, S., 1982. Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the GI/G/1 queue. Adv. in Appl. Probab. 14(1), 143-170.

Asmussen, S., 1987. Applied Probability and Queues. Wiley, Chichester.

Balkema, A. A. and de Haan, L., 1974. Residual lifetime at great age. Ann. Probab. 2, 792-804.

Bender, C. M. and Orszag, S. A., 1999. Advanced Mathematical Methods for Scientists and Engineers. Springer, Berlin.

Beran, J., 1992. A goodness of fit test for time series with long-range dependence. J. Roy. Statist. Soc. Ser. B 54, 749-760.

Beran, J., 1994. Statistics for long-memory processes. Chapman and Hall, New York.

Bierens, H. J., 2004. Introduction to the Mathematical and Statistical Foundations of Econometrics. Cambridge University Press.

Bíró, A., Elek, P. and Vincze, J., 2008. Model-based sensitivity analysis of the Hungarian economy to shocks and uncertainties. Acta Oeconomica 58, 367-401.

Bollerslev, T., 1986. Generalised autoregressive conditional heteroscedasticity. J. Econometrics 31, 307-327.

Borkovec, M., 2000. Extremal behaviour of the autoregressive process with ARCH(1) errors. Stochastic Process. Appl. 85, 189-207.

Borkovec, M. and Klüppelberg, K., 2001. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. Ann. Appl. Probab. 11, 1220-1241.

Bortot, P. and Coles, S., 2000. A sufficiency property arising from the characterisation of extremes of Markov chains. Bernoulli 6, 183-190.

Bortot, P. and Tawn, J., 1998. Models for the extremes of Markov chains. Biometrika 85, 851-867.

Box, G. and Jenkins, G., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Bozsó, D., Rakonczai, P. and Zempléni, A., 2005. Floods of River Tisza and of its tributaries. Statisztikai Szemle 83, 919-936 (in Hungarian).

Brandt, A., 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. Adv. in Appl. Probab. 18, 211-220.

Breiman, L., 1965. On some limit theorems similar to the arc-sin law. Theory Probab. Appl. 10, 323-331.

Brochu, M., 1978. Computerized rivers. Science Dimension 10, 18-20.

Brockwell, P. J. and Davis, R. A., 1991. Time Series: Theory and Methods. Springer, New York.

Bühlmann, P. and McNeil, A., 2002. An algorithm for nonparametric GARCH modelling. Comput. Statist. Data Anal. 40, 665-683.

Cleveland, B. R., Cleveland, W. S., McRae, J. E. and Terpenning, I., 1990. STL: A seasonal trend- decomposition procedure based on loess. J. Off. Statist. 6, 3-73.

Coles, S. and Tawn, J., 1991. Modelling extreme multivariate events. J. R. Statist. Soc. Ser. B 53, 377-392.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.

Cunnane, C., 1989. Statistical distributions for flood frequency analysis. Operational Hydrology Report No.33, Secretariat of the World Meteorological Organization, WMO No.718, Geneva, Switzerland.

Elek, P., 2002. Statistical analysis of long range dependent processes. Unpublished MSc thesis (in Hungarian), Eötvös Loránd University, Budapest.

Elek, P., 2003. Extreme value theory and volatility models in the measurement of financial market risk. Unpublished MA thesis (in Hungarian), Corvinus University of Budapest.

Elek, P. and Márkus, L., 2004. A long range dependent model with nonlinear innovations for simulating daily river flows. Natur. Hazards Earth Systems Sci. 4, 277-283.

Elek, P. and Márkus, L., 2008. A light-tailed conditionally heteroscedastic time series model with an application to river flows. J. Time Ser. Anal. 29, 14-36.

Elek, P. and Márkus, L., 2009. Tail behaviour and extremes of β-TARCH processes. Manuscript.

Elek, P. and Zempléni, A., 2008. Tail behaviour and extremes of two-state Markov-switching autoregressive models. Comput. Math. Appl. 55, 2839-2855.

Elek, P. and Zempléni, A., 2009. Modelling extremes of time-dependent data by Markov-switching structures. J. Statist. Plann. Inference 139, 1953-1967.

Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Engle, R. F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica 50, 987-1007.

Fawcett, L. and Walshaw, D., 2006. Markov chain models for extreme wind speeds. Environmetrics 17, 795-809.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications. Wiley, New York.

Ferro, C. A. T. and Segers, J., 2003. Inference for clusters of extreme values. J. R. Statist. Soc. Ser. B 65, 545-556.

Fisher, R. A. and Tippet, L. H. C., 1928. Limiting forms of the frequence distribution of the largest or smallest member of a sample. Proc. Cambridge Phil. Soc. 24, 180-190.

Francq, C. and Roussignol, M., 1998. Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. Statistics 32, 151-173.

Francq, C. and Zakoian, J. M., 1998. Estimating linear representations of nonlinear processes. J. Statist. Plann. Inference 68, 145-165.

Francq, C., Roy, R. and Zakoian, J. M., 2005. Diagnostic checking in ARMA models with uncorrelated errors. J. Amer. Statist. Assoc. 100, 532-544.

Giraitis, L. and Surgailis, D., 1990. A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. Probab. Theory Related Fields 86, 87-104.

Giraitis, L. and Taqqu, M.S., 1999. Whittle estimator for finite-variance non-Gaussian time series with long memory. Ann. Statist. 27, 178-203, 1999.

Glosten, L. R., Jagannathan, R. and Runkle, D., 1993. On the relation between the expected value and the volatility of the normal excess return on stocks. J. Finance 48, 1779-1801.

Goldie, C. M., 1991. Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. 1, 126-166.

Green, P. J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711-732.

Greenwood, P., 1971. Wiener-Hopf decomposition of random walks and Lévy-processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 34, 193-198.

Grimaldi, S. and Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. Adv. in Water Resour. 29, 1155-1167.

Guegan, D. and Diebolt, J., 1994. Probabilistic properties of the β-ARCH-model. Statist. Sinica 4, 71-87.

Haan, L. de, 1990. Fighting the arch-enemy with mathematics. Statist. Neerlandica 44, 45-68.

Hamilton, J. D., 1990. Analysis of time series subject to changes in regime. J. Econometrics 45, 39-70.

Hamilton, J. D., 1994. Time series analysis. Princeton University Press, Princeton, N. J.

Hamilton, J. D. and Susmel, R., 1994. Autoregressive conditional heteroscedasticity and changes in regimes. J. Econometrics 64, 307-333.

Hsing, T., Hüsler, J. and Leadbetter, M. R., 1988. On the exceedance point process of a stationary sequence. Probab. Theory Related Fields 78, 97-112.

Hurst, H. E., 1951. Long-term storage capacity of reservoirs. Trans. of the Amer. Soc. of Civil Engineers 770-808.

Hwang, S. Y. and Basawa, I. W., 2003. Estimation for nonlinear autoregressive models generated by beta-ARCH processes. Sankhya 65(4), 744-762.

Jain, A. and Srinivasulu, S., 2006. Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. J. Hydrology 317, 291-306.

Karlsen, H. A., 1990. Doubly stochastic vector AR(1) processes. PhD. Dissertation, Univ. of Bergen, Norway.

Klüppelberg, C. and Lindner, A., 2005. Extreme value theory for moving average processes with light-tailed innovations. Bernoulli 11(3), 381-410.

Kristensen, D. and Rahbek, A., 2005. Asymptotics of the QMLE for a class of ARCH(q) models. Econometric Theory 21, 946-961.

Lawrance, A. J. and Kottegada, N. T., 1977. Stochastic modelling of river flow time series. J. Roy. Statist. Soc. Ser. A 140, 1-31.

Lu, Z. Q. and Berliner, L. M., 1999. Markov switching time series models with application to a daily runoff series. Water Resour. Res. 35(2), 523-534.

Masry, E. and Tjostheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory 11, 258-289.

McNeil, A. and Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Empirical Finance 7, 271-300, 2000.

Meyn, S. P. and Tweedie, R. L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Mikosch, T., 2006. Copulas: tales and facts. Extremes 9, 3-20.

Montanari, A., Rosso, R. and Taqqu, M. S., 1997. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. Water Resour. Res. 33, 1035-1044.

Montanari, A. Rosso, R. and Taqqu, M. S., 2000. A seasonal fractionally differenced ARIMA model: an application to the Nile River monthly flows at Aswan. Water Resour. Res. 36, 1249-1259.

Ooms, M. and Franses, P. H., 2001. A seasonal periodic long memory model for monthly river flows. Environmental Modelling and Software 16, 559-569.

Perfekt, R., 1994. Extremal behaviour of stationary Markov chains with applications. Ann. Appl. Probab. 4, 529-548.

Resnick, S. I., 1971. Asymptotic location and recurrence properties of maxima of a sequence of random variables defined on a Markov chain. Z. Wahrscheinlichkeitstheorie verw. Geb. 18, 197-217.

Resnick, S. I., 1987. Extreme values, point processes and regular variation. Springer, New York.

Robert, C., 2000. Extremes of alpha-ARCH models. In: Measuring Risk in Complex Stochastic Systems, Ed.: Franke, J., Hardle, W. and Stahl, G., 219-251. Springer, Berlin.

Saporta, B., 2005. Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients. Stochastic Process. Appl. 115, 1954-1978.

Segers, J., 2003. Functionals of clusters of extremes. Adv. in Appl. Probab. 35, 1028-1045.

Sisson, S. and Coles, S., 2003. Modelling dependence uncertainty in the extremes of Markov chains. Extremes 6, 283-300.t

Smith, R. L., 1992. The extremal index for a Markov chain. J. Appl. Probab. 29, 37-45.

Smith, R. L., Tawn, J. A. and Coles, S. G., 1997. Markov chain models for threshold exceedances. Biometrika 84, 249–268.

Spitzer, F., 1960. A tauberian theorem and its probability interpretation. Trans. Amer. Math. Soc. 94, 150-169.

Szilágyi, J., Bálint, G. and Csík, A., 2006. Hybrid, Markov chain-based model for daily streamflow generation at multiple catchment sites. J. Hydrologic Eng. 11, 245-256.

Tong, H., 1990. Non-linear Time Series: A dynamical system approach. Oxford University Press, New York.

Turkman, K. F. and Oliveira, M. F., 1992. Limit laws for the maxima of chain-dependent sequences with positive extremal index. J. Appl. Probab. 29, 222-227.

Vasas, K., Elek, P. and Márkus, L., 2007. A two-state regime switching autoregressive model with an application to river flows. J. Statist. Plann. Inference 137, 3113-3126.

Yang, M., 2000. Some properties of vector autoregressive processes with Markov-switching coefficients. Econometric Theory 16, 23-43.

Yao, J. F. and Attali, J. G., 2000. On stability of nonlinear AR-processes with Markov switching, Adv. in Appl. Probab. 32, 394-407.

Zhang, J. and Stine, R. A., 2001. Autocovariance structure of Markov regime switching models and model selection. J. Time Ser. Anal. 22, 107-124.

Summary

In this thesis I investigate probabilistic properties, extremal behaviour, parameter estimation and hydrological applications of two classes of stationary nonlinear time series models that share the common feature that all of their moments are finite.

After the Introduction, Chapter 2 summarizes the concepts of time series analysis and extreme value theory needed for the results of the thesis. Discussing in detail the motivation behind developing the models, I present the empirical properties of daily river discharge data of Danube and Tisza in Chapter 3. Besides being highly autocorrelated, these series belong to the max-domain of attraction of the Gumbel law and exhibit clustering at high levels. I also demonstrate that linear ARMA or fractional ARIMA processes are not suitable for analysing the data sets thus the use of nonlinear models is necessary.

Chapter 4 introduces an unusual ARMA-ARCH-type model where the conditional variance of the ARMA innovations grows as $Kx^{2\beta}$ where $0 < \beta < 1$. I prove that if all roots of the characteristic polynomials of the corresponding AR- and MA-terms lie outside the closed unit disk the model has a stationary solution, moreover, all of its moments are finite provided that the same is true for the generating noise. The model is shown to possess approximately a Weibull-like tail. It is estimated by a mixture of least squares and maximum likelihood methods and I prove the consistency and asymptotic normality of the estimator. Finally, the model is fitted to the water discharge data and is shown to be superior to the linear specifications.

Chapter 5 examines Markov-switching autoregressive (MS-AR) models with two states (regimes): a random walk one and a stationary one. I prove that if the generating noise of the random walk regime is light-tailed then the model asymptotically has an exponential tail, its extremal clustering is nontrivial and its limiting aggregate excess distribution approximately has a Weibull-like tail. I also show why the extremal behaviour of this model is a good statistical approximation of the extremes of more general Markov-switching conditionally Markov models. Then I develop a threshold-based procedure to fit the MS-AR model to high-level exceedances of an observed process and analyse the extremal cluster functionals of the water discharge data using this method. Finally, the concluding chapter examines the relationship between the two model families of the dissertation and outlines directions for further research.

Összefoglaló

A disszertációban két olyan nemlineáris idősormodell-család valószínűségi jellemzőit, extremális viselkedését, paramétereinek becslését és hidrológiai alkalmazásait vizsgálom, amelyek stacionárius eloszlásának minden momentuma véges.

A bevezetés után a 2. fejezetben foglalom össze az idősorelemzés és az extrémérték-elmélet néhány fontos eredményét. Mivel a modellvizsgálatokat eredetileg hidrológiai alkalmazások motiválták, a 3. fejezetben mutatom be a Duna és Tisza folyók vízhozamadatsorainak empirikus jellemzőit. Az idősorok erősen autokorreláltak, eloszlásuk a Gumbel-eloszlás maximum vonzási tartományába tartozik, és magas értékeik klaszterekben fordulnak elő. Azt is bemutatom, hogy lineáris ARMA- és frakcionális ARIMA-folyamatokkal nem lehet megfelelően modellezni az adatsorokat, ezért szükséges a nemlineáris modellezés.

A 4. fejezetben egy olyan, a szokásostól eltérő ARMA-ARCH-típusú modellt vizsgálok, amelyben az innovációk feltételes varianciája $Kx^{2\beta}$ nagyságrendű, ahol $0 < \beta < 1$. Bebizonyítom, hogy ha az AR- és MA-tagok karakterisztikus egyenletének minden gyöke a zárt egységkörön kívül van, akkor a modellnek létezik stacionárius megoldása, sőt a megoldás minden momentuma véges, ha ugyanez teljesül a generáló zajra. A stacionárius eloszlás széle közelítően Weibull-típusú módon cseng le. A modellt a maximum likelihood és a legkisebb négyzetek módszerének kombinációjával becsülöm, és bebizonyítom a becslőfüggvény konzisztenciáját és aszimptotikus normalitását. Végül vízhozamidősorokra illesztem a modellt és illusztrálom annak kedvező szimulációs tulajdonságait a lineáris specifikációhoz képest.

Az 5. fejezetben egy olyan kétállapotú Markov-rezsimváltó autoregresszív (MS-AR) modellt elemzek, amelyben az egyik rezsim véletlen bolyongás, a másik pedig stacionárius autoregresszió. Bebizonyítom, hogy ha az első rezsimben a zaj vékony szélű, akkor a modell stacionárius eloszlása aszimptotikusan exponenciális lecsengésű, a folyamat magas értékei klaszterekben fordulnak elő, és az aggregált túllépések határeloszlása hozzávetőlegesen Weibull-típusú. Megmutatom azt is, hogy bizonyos általánosabb Markov-rezsimváltó, feltételesen Markov-típusú folyamatok extremális viselkedése is jól közelíthető MS-AR-modellel. Ezután egy küszöbalapú becslési eljárást dolgozok ki az MS-AR-modell becslésére abban az esetben, ha csak magas értéknél cenzorált megfigyelések állnak rendelkezésre, és elemzem a vízhozam-idősorok extremális klaszter funkcionáljait az így illesztett modell alapján. Végül az utolsó, következtetéseket tartalmazó fejezet rámutat a disszertációban vizsgált két modellcsalád közötti kapcsolatra, és további lehetséges kutatási témákat is megfogalmaz.