A light-tailed conditionally heteroscedastic model with applications to river flows

PÉTER ELEK[1] AND LÁSZLÓ MÁRKUS[2]

Eötvös Loránd University

Department of Probability Theory and Statistics,

Pázmány Péter sétány 1/c, H-1117 Budapest, Hungary

[1]Corresponding author, email: elekpeti@cs.elte.hu

[2]Email: markus@cs.elte.hu

Abstract

A conditionally heteroscedastic model, different from the more commonly used ARMA-GARCH processes, is established and analysed here. The time dependent variance of innovations passing through an ARMA-filter is conditioned on the lagged values of the generated process, rather than on the lagged innovations, and is defined to be asymptotically proportional to those past values. Designed this way, the model incorporates certain feedback from the modelled process, the innovation is no longer of GARCH-type, and all moments of the modelled process are finite provided the same is true for the generating noise. The paper gives the condition of stationarity, and proves consistency and asymptotic normality of the Gaussian quasi-maximum-likelihood estimator of the variance parameters, even though the estimated parameters of the linear filter contain an error.

An analysis of six diurnal water discharge series observed along Rivers Danube and Tisza in Hungary demonstrates the usefulness of such a model. The effect of lagged river discharge turns out to be highly significant on the variance of innovations, and nonparametric estimation approves its approximate linearity. Simulations from the new model preserve well the probability distribution, the high quantiles, the tail behaviour and the high-level clustering of the original series, further justifying model choice.

Keywords: GARCH-type models, nonlinear time series, quasi maximum likelihood estimation, extreme value theory, river flow modelling

1 Introduction

The conditionally heteroscedastic (ARCH-type) model, this paper presents, differs from the commonly used ARMA-GARCH processes (Hamilton, 1994) in the generating innovations of the ARMA-filter. Here, the innovation is not just a simple ARCH/GARCH process, because the variance is not conditioned on the lagged innovations themselves. Instead, conditioning is made on the past value of the observed process. Though models based on this idea are known (Borkovec and Klüppelberg, 2001; Masry and Tjostheim, 1995), they are much less elaborated in the literature. This type of conditioning is designed to capture a feedback from the modelled process, making it appropriate e.g. for river discharge description. 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 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 value, corresponding in our application to the above mentioned feedback effect.

The analysis of a 100 year long daily water discharge measurements of River Tisza reported in Elek and Márkus (2004) shows that a linear, long range dependent model with independent non-Gaussian innovations does not approximate well the probability density and high quantiles of the discharge series, even if the distribution of the innovations changes seasonally. These results differ from the findings of Montanari et al. (1997), where a linear 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. 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.

Elek and Márkus (2004) attribute the nonlinear behaviour to the driving force of the linear system by modelling the uncorrelated innovations as a modified ARCH-type process. The conditional variance in that model was - somewhat unusually - a *bounded* function of the past value of the innovations. Conditioning the variance on the lagged *innovations* may not always fit the character of the phenomenon described, and consequently the obtained fit can be improved by better conditioning. As it was noted above, in certain situations it is more natural to condition on the value of the process itself. This gives the motivation to develop the idea further along this line and consider the following model.

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

$$\epsilon_t = \sigma(X_{t-1}) Z_t, \quad (2)$$

where

$$\sigma(x) = (\alpha_0 + \alpha_1(x-m)_+)^{1/2}. \tag{3}$$

Here, c_t is a deterministic periodic function (with period of e.g. one year), representing the seasonal component of the mean. y_+ denotes $\max(y,0)$. We impose the following assumptions on the parameters and on the noise sequence.

Assumption 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 2. *The characteristic polynomials $x^p - a_1x^{p-1} - \cdots - a_p$ and $x^q - b_1x^{q-1} - \cdots - b_q$ have all roots within the unit circle.*

Assumption 3. $\alpha_0 > 0$ *and* $\alpha_1 \geq 0$.

Assumption 3 implies that the variance function is a positive constant for lagged values smaller than a threshold m and increases linearly for lagged values higher than m. When m is close to the minimum of the observed series, the linear relationship applies for most times and in fact α_0 is needed only to ensure the strict positivity of the conditional variance.

Putting (1) into words the process X_t is generated from the ϵ_t innovation sequence like an ARMA-process with seasonally changing mean. Assumption 2 ensures that the corresponding ARMA-model with independent innovations and constant c_t is stationary and invertible. The ϵ_t innovations, however, are uncorrelated but not independent as their temporally changing variance creates interdependence. Their variance depends on the lagged value of the observed process X_t itself, realising the feedback mentioned in the beginning of this section. Moreover, the form of dependence – characterised by the function $\sigma^2(x)$ – is not quadratic but linear over a threshold.

This choice of linear specification of the conditional variance is justified in our application by a nonparametric estimation procedure. (See Section 4 for details.) The asymmetric $\sigma(x)$ function provides – even in the case of symmetrically distributed Z_t noise – a process with highly skewed marginal distribution, in accordance with our application. (When a linear ARMA-model is used instead, innovations with asymmetric distribution are necessary in order to produce skewness.)

The paper is organised as follows. Since our model does not fit directly into any previously considered setup, Section 2 proves the stationarity of the model, using results of Meyn and Tweedie (1993) on the stability of Markov chains. It proves, moreover, that the stationary process created from the linear specification of conditional variance has finite moments of all orders, provided the same holds true for the Z_t noise. (This is not the case for the usual, quadratic specification.) Section 3 discusses model estimation. For computational reasons, the linear filter is estimated by least squares while the variance equation is estimated by Gaussian quasi maximum likelihood. Even though the fitted ϵ_t innovations contain an error, we prove that the QML estimator of the variance equation is consistent and asymptotically normally distributed. Section 4 reveals the focal points of the application, explains the model choice, guides through the identification process and provides the details of the estimation results for six water discharge data sets. Section 5 justifies the model choice of section 4 and addresses goodness of fit by means of simulation from the fitted model. It shortly discusses issues of the simulation and then evaluates the model fit in a non-standard way, in terms of probability density, high quantiles, tail behaviour and extremal clustering – by comparing these for the empirical and the simulated series. In the application the model outperforms in all respect both the fractionally integrated linear model driven by an independent value noise with seasonal distribution, and the nonlinear model of Elek and Márkus (2004).

2 Existence and moments of the stationary distribution

First we address the problem of existence of a stationary solution to the proposed model (1)-(3). In the absence of moving average coefficients (i.e. $q = 0$), our model reduces to an autoregressive model with ARCH-type innovations, where the conditional variance is increasing slower than a quadratic function: $\lim_{|x|\to\infty} \sigma^2(x)/x^2 = 0$. In this case geometric ergodicity and hence existence of a unique stationary distribution was proven 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 ARCH(1)-case with linear variance equation). They proved that under the finiteness of all moments of Z_t, all moments of the stationary distribution of X_t are finite, too. (These results go against the classical ARCH(1) model of quadratic heteroscedasticity where the domain of stationarity depends on the parameters of the $\sigma^2(x)$ function. Moreover, not all moments of the stationary distribution are finite. For details see Borkovec and Klüppelberg, 2001.)

We now prove the stationarity of our conditionally heteroscedastic model. To avoid problems with seasonality, we assume that $c_t = c$ is a constant.

Theorem 1. *Assume that $c_t = c$ and Assumptions 1-3 hold. Then, the X_t process defined by* (1)-(3) *is geometrically ergodic and has a unique stationary distribution. If $E(|Z_t|^r)$ is finite for an $r \geq 2$ real then the stationary distribution of X_t has finite r-th moment.*

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,

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

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

$$\mathbf{Y_t} = \mathbf{A}\mathbf{Y_{t-1}} + \mathbf{E}_t, \qquad (4)$$

where $\mathbf{E}_t = (\epsilon_t, 0, \ldots, 0, \epsilon_t)^T$ 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 1. *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 $\mathbf{y} \in \mathbf{R}^{p+q}$*

$$||\mathbf{QBy}||_r \leq \mu||\mathbf{Qy}||_r \tag{5}$$

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 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 $1 > \beta > 0$ such that

$$E(V(\mathbf{Y_1})|\mathbf{Y_0} = \mathbf{y}) \leq (1-\beta)V(\mathbf{y}) + bI_C(\mathbf{y}) \tag{6}$$

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 $|\epsilon_1|$ there exists an $s > 0$ such that $||\mathbf{QPE}_1||_r = s|\epsilon_1| = s\sigma\left(Y_0^1\right)|Z_1|$. The identity $\mathbf{PA} = \mathbf{BP}$, the triangle inequality for the r-norm and inequality (5) 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

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

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, (6) is satisfied with $1 - \beta = \mu_2$ and with a suitably chosen b. This means that the stationary distribution exists and has finite V-moments and consequently has finite r-th moments. This concludes the proof. □

Remark 1. *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:*

$$\frac{1}{n}\sum_{t=1}^{n} g(X_t) \to E_\pi(g(X_t)) \text{ a.s.} \tag{7}$$

According to Theorem 1, if Z_t is light tailed in the sense that it has finite moments of all orders then X_t will be light tailed, too. This fact, however, does not determine the exact tail behaviour even in the special cases when Z_t is, for instance, normally or double exponentially distributed. Robert (2000) examines a related model

$$X_t = (\alpha_0 + \alpha_1 |X_{t-1}|^\kappa)^{1/2} Z_t$$

and proves that for $1 < \kappa < 2$ and for a normally distributed Z_t noise, $P(X_t > x) \sim K x^{b_1} \exp\left(-x^{b_2}\right)$ with constants K, b_1 and b_2. However, in the case of $\kappa = 1$ (which roughly corresponds to our model without ARMA-terms), no result is known about the form of the tail, hence the tail behaviour of our more general model is unknown, too.

It is well known in extreme value theory that, under general conditions, the tail of a random variable Y can be approximated by a generalized Pareto distribution (GPD) with distribution function $F_{\xi,\beta}$, i.e. there exist a shape parameter ξ and a scale parameter $\beta > 0$ such that for all $y > 0$

$$\lim_{u \to x_Y} P\left(Y < y + u | Y > u\right) = F_{\xi,\beta}\left(y\right)$$

where x_Y is the upper end point of the support of the distribution of Y (see Embrechts et al., 1997 or Coles, 2001). The finiteness of all moments and the infinite support of a distribution imply that the shape parameter (if the distribution belongs to the domain of attraction of a GPD at all) is zero, or may be negative in some pathological cases. Hence, e.g. a normal or double exponential noise can only generate a stationary distribution with $\xi = 0$ in our model, by virtue of Theorem 1. In contrast, the usual AR-(G)ARCH models with quadratic variance structure belong to the $\xi > 0$ case (see Borkovec and Klüppelberg, 2001).

3 Model estimation

Let us fix the values of p, q and m (details of their identification are described in Section 4) and denote the set of the linear filter and variance equation parameters respectively by $\boldsymbol{\theta}^0 = (a_1, \ldots, a_p, b_1, \ldots, b_q)$ and by $\boldsymbol{\alpha}^0 = (\alpha_0, \alpha_1)$. (As we are concerned with asymptotic properties, and these are not affected by replacing the expectation by the empirical mean, in the following we assume that the mean parameter c is known.) In order to achieve a fast algorithm for large sample sizes, a two stage estimation procedure is applied instead of full maximum likelihood. As it will turn out, we do not lose efficiency by this method during the estimation of $\boldsymbol{\alpha}^0$.

First, based on the ARMA representation of X_t the least squares estimator of $\boldsymbol{\theta}^0$ is obtained, i.e. 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 $\hat{\epsilon}_t$ for the fitted innovations at $\hat{\boldsymbol{\theta}}_n$, i.e. $\hat{\epsilon}_t = e_t(\hat{\boldsymbol{\theta}}_n)$.

To get consistency of the least squares estimator, we only need the uncorrelatedness – and not the independence – of the ϵ_t innovations. It was proven in Francq and Zakoian (1998) that $\hat{\boldsymbol{\theta}}_n$ is consistent under our assumptions, i.e. $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^0$ a.s. For asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ we need the additional

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

By Theorem 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.

The estimation of the parameter vector $\boldsymbol{\alpha}^0 = (\alpha_0, \alpha_1)$ in the variance equation can be carried out by Gaussian quasi maximum likelihood. (The term "quasi" refers to the fact that Z_t is not assumed to be Gaussian in the consistency and asymptotic normality results.) Denoting the sample size by n, this means the maximization of the following term:

$$\hat{L}_n(\boldsymbol{\alpha}) = \frac{1}{n} \sum_{t=1}^{n} l(\hat{\epsilon}_t, X_{t-1}, \boldsymbol{\alpha}), \tag{8}$$

where

$$l(y, x, \boldsymbol{\alpha}) = -\log \sqrt{2\pi} - \frac{1}{2} \log(\sigma^2(x, \boldsymbol{\alpha})) - \frac{1}{2} \frac{y^2}{\sigma^2(x, \boldsymbol{\alpha})} \tag{9}$$

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 (3). (With this notation we emphasize that σ depends on the parameter vector $\boldsymbol{\alpha}$, too.)

To prove consistency and asymptotic normality of the quasi maximum likelihood estimator we need to impose some restrictions on the parameter space. We assume that the parameter space is compact, the true parameter value lies in its interior and the set of possible constant terms in the variance equation is separated from zero:

Assumption 5. $\boldsymbol{\alpha}^0 \in \mathrm{int}(\mathbf{K})$, *where* $\mathbf{K}$ *is a compact subset of* $\mathbf{R}^+ \times \mathbf{R}^{0+}$.

Then the following holds.

Theorem 2. *Under Assumptions 1, 2 and 5 the quasi maximum likelihood estimator is consistent, i.e.*

$$\hat{\boldsymbol{\alpha}}_n \to \boldsymbol{\alpha}^0 \quad a.s. \tag{10}$$

If in addition Assumption 4 holds, the resulting estimator is asymptotically normally distributed, i.e.

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0) \to_d N\left(\mathbf{0}, \mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\mathbf{V}\left(\boldsymbol{\alpha}^0\right)\mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\right) \tag{11}$$

where

$$\mathbf{V}(\boldsymbol{\alpha}) = E_\pi\left(\frac{\partial l(\epsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\frac{\partial l(\epsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^T\right) \tag{12}$$

$$\mathbf{H}(\boldsymbol{\alpha}) = E_\pi\left(-\frac{\partial^2 l(\epsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial^2 \boldsymbol{\alpha}}\right). \tag{13}$$

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 $\mathbf{V}(\hat{\boldsymbol{\alpha}}_n)$, with expectations 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 $\hat{\epsilon}_t = \epsilon_t$ holds for all t). Then the maximum likelihood estimator maximizes

$$L_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{t=1}^{n} l(\epsilon_t, X_{t-1}, \boldsymbol{\alpha}). \tag{14}$$

According to the ergodic theorem

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

and it can be proven that $L(\boldsymbol{\alpha})$ obtains its maximum at the true parameter value $\boldsymbol{\alpha}^0$. Afterwards the proof of consistency goes directly along the lines of Kristensen and Rahbek (2005). (Conditions denoted by C.1.-C.4. in that paper are needed for consistency. However, because of the compactness of the parameter space C.1. is not required and C.4. (ii) can be weakened to the trivial $E_\pi[|\log \sigma_t^2|] < \infty$ condition. Moreover, C.3. (ii) can be replaced with the assumption that (ϵ_t, X_{t-1}) is embedded into a multidimensional Markov-chain. The other conditions follow directly from our Assumptions, too.)

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

$$\mathbf{0} = \mathbf{S}_n(\hat{\boldsymbol{\alpha}}_n) = \mathbf{S}_n(\boldsymbol{\alpha}^0) - \mathbf{H}_n(\boldsymbol{\alpha}_n^*)(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0) \tag{15}$$

where

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

It can be proven that $\sqrt{n}\mathbf{S}_n(\boldsymbol{\alpha}^0) \to N\left(\mathbf{0}, \mathbf{V}\left(\boldsymbol{\alpha}^0\right)\right)$ in distribution and $\mathbf{H}_n(\boldsymbol{\alpha}_n^*) \to \mathbf{H}(\boldsymbol{\alpha}^0)$ a.s. so (11) follows. (For details see Kristensen and Rahbek, 2005.)

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

Lemma 2. *If $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}^0$,*

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{\epsilon}_t^2 - \epsilon_t^2) \to 0 \quad \textit{a.s.} \tag{16}$$

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\epsilon}_t^2 - \epsilon_t^2) \to 0 \quad \textit{a.s.} \tag{17}$$

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, we use a standard Taylor-expansion:

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

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. □

Getting back to the proof of Theorem 2 we can compare the log-likelihood $\hat{L}_n(\boldsymbol{\alpha})$ calculated from the $\hat{\epsilon}_t$ series and the log-likelihood $L_n(\boldsymbol{\alpha})$ calculated from the true ϵ_t series:

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

By (16) 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:

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

Equations (9) and (17) 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 (15) remain valid and so asymptotic normality of the parameters in the variance equation holds in this case, too. □

4 Identification and estimation results for the water discharge data sets

Rivers Danube and Tisza both have a long history of damaging inundations in Hungary and the record-high floods in years 2000, 2001 and 2002 drew again attention to river flow management. This necessitates a renewed study of the registered measurements of river discharge. As, understandably, the primary concern is the occurrence of high discharge values, the estimation of high quantiles of the discharge series has become a crucial aim of the analysis. The main motivation behind the above presented time series model was the need for a description of the diurnal discharge series that reflects well its dependence structure including the mentioned feedback, its probability distribution including the skewness and high quantiles and, additionally, its extremal clustering behaviour. As water discharge measurements of river Danube and Tisza tend to be light-tailed, the proven finiteness of moments further strengthens the arguments for modelling them by time series of the above type.

In the following we present the analysis of six, 50 to 100 year long daily water discharge series measured at various monitoring stations along the two main rivers in Hungary. The monitoring sites in question are Komárom, Nagymaros, Budapest at River Danube and Tivadar, Vásárosnamény, Záhony at River Tisza. As the results for all the six stations are very much similar, some details of the identification and estimation procedure are only displayed for the site Nagymaros.

All series exhibit substantial seasonality both in mean and variance and they are highly autocorrelated. As a starting point, the seasonal components (c_t) of the series were estimated using a local polynomial fitting (LOESS) procedure. (Details can be found in Elek and Márkus, 2004.) Figure 1 displays the probability density of the water discharge series at Nagymaros and also the estimated seasonal component c_t at that site.

Put Figure 1 about here

Afterwards, ARMA(p,q) models were fitted to the deseasonalised $X_t - c_t$ series with various p and q orders. We did not use here likelihood-based model selection criteria like the AIC or BIC because the combination of the MA and the ARCH parts of the model make the form of the exact likelihood overly complicated. An ARMA(3,1) model seemed to be appropriate at each site in the sense that there remained little autocorrelation in the $\hat{\epsilon}_t = e_t(\hat{\boldsymbol{\theta}}_n)$ innovation sequences. However, the p-values derived from the standard Box-Pierce and Ljung-Box tests could not be used to formally evaluate the goodness of fit of these models since the ϵ_t innovation process was only assumed to be uncorrelated but not independent (see Francq et al., 2005). Instead of calculating the exact asymptotic distribution of the Box-Pierce statistics of the fitted innovations under the null hypothesis that the true model is our ARMA-ARCH one, we used a posterior evaluation strategy to justify the adequateness of the choice $p = 3$, $q = 1$. After estimating the (α_0, α_1) ARCH-parameters as well (see below), we simulated synthetic ARMA-ARCH series with the estimated parameters, fitted ARMA(3,1) models to them and obtained empirical critical values for the goodness of fit based on the Box-Pierce statistics of the resulting innovations. The fit was accepted at all sites at the 99% level for not too large lags in the autocorrelations. For instance, at site Nagymaros the simulated 99% critical value of the Box-Pierce statistics of the first $r = 8$ autocorrelations of the innovations was 22.1, while the observed value was 21.6. (The 99% quantile of a χ^2 distribution with $r - p - q = 4$ degrees of freedom is 13.3, hence model selection based on the standard p-values would have rejected the ARMA(3,1) model and thus would have resulted in overfitting.) It is worth noting that a more restricted model (i.e. ARMA(2,1) or AR(3)) was not appropriate at any site.

As the Box-Pierce test shows there remain some autocorrelations at high lags but this problem cannot be overcome by increasing the model order. This is possibly related to the otherwise detectable long memory behaviour of the series (cf. Elek and Márkus, 2004). Introducing a fractional differentiation filter on top of the ARMA one would completely eliminate the remaining autocorrelations. At the same time, however, it would overcomplicate the model while leaving the obtained density, quantile and extremal clustering characteristics practically the same as in the simpler ARMA case, and would make the theoretical properties (e.g. stationarity) unclear.

Figure 2 shows the probability density and the autocorrelation function of the $\hat{\epsilon}_t$ fitted innovations as well as the autocorrelation function of their squares and of their absolute values at Nagymaros. These plots indicate that the innovations are highly non-Gaussian, highly peaked and somewhat skewed. Moreover, although they are uncorrelated, their squares and their absolute values are autocorrelated sequences, pointing to an ARCH-type effect.

Put Figure 2 about here

We used 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) were grouped into 50 groups according to their magnitude and the fitted innovations (the $\hat{\epsilon}_t$-s) were classified based on the group of X_{t-1}. Thereafter, the variance of the fitted innovations and the mean of the discharges in each group were calculated and plotted against each other.

Figure 3 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). The empirical relationship is close to linear for all stations, and this fact makes the choice of the simplest model for $\sigma^2(x)$, the linear, sensible. (To ensure regularity, $\sigma^2(x) > 0$ should hold for all x, pointing to the parametrization in (3).) It is also clear from the plots that the break point m should lie very close to (or perhaps even lower than) the observed minimum of the water discharge series. Note that there would be no apparent pattern in the figures if an ARMA model with independent innovations were appropriate.

Figure 3 also shows the fitted lines $y = \alpha_0 + \alpha_1 (x - m)$ with maximum likelihood 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 $\sigma^2(x)$.

Put Figure 3 about here

Another indication of our choice comes from the (upper) tails of the observed water discharge series. As it has already been pointed out at the end of section 2, the quadratic ARCH-specification would imply a $\xi > 0$ shape parameter for a generalized Pareto distribution fitted to the tails, while the linear variance structure is consistent with the $\xi = 0$ case. Table 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. (We used the runs method with run length of 15 days to identify separate clusters of exceedances. See Embrechts et al., 1997 for details of the method.) It turns out that the point estimates are close to zero and are generally negative, being only the estimate at Záhony significantly different from zero. This gives a strong case against $\xi > 0$ and hence rejects the applicability of traditional GARCH processes. (For more details on extreme value modelling 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 is well known that if a Gumbel distribution fits well to the annual maxima, then the exceedances over high thresholds can be approximated by a GPD with $\xi = 0$, i.e. by an exponential distribution.) 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 and thus heavy tails are usually advised for smaller catchments.)

Table 1: Shape parameter estimates (with standard errors in parentheses) of the GPD-s fitted to exceedances over the specified thresholds measured in m^3/s at different monitoring stations

Monitoring station	Threshold	ξ	Threshold	ξ
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)

Because of non-differentiability, the estimation of parameter m would be a nonregular problem. Thus, after a pilot study, we have fixed m to lie at the minimum of the observed water discharge series at each site. (Different choices had no substantial effect on the properties of the series simulated from the models, as long as m was not chosen too large. However, the estimate of α_0 obviously varied with different values of m.) With such a specification the linear relationship for the conditional variance applies for the whole observed series, therefore different models where m is lower than the minimum discharge value cannot be distinguished from each other. In practice, however, this is not a crucial problem as they produce only slightly different simulation outputs.

With these in mind, Table 2 shows the quasi maximum likelihood estimates of the α_0 and α_1 parameters (along with the threshold value m) for the six stations. Asymptotic standard errors from Theorem 2 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.

Based on the estimated parameters, the fitted noise sequence ($\hat{Z}_t = \hat{\epsilon}_t/\sigma(X_{t-1})$) can be calculated easily. Figure 4 shows its probability density, its autocorrelation function as well as the

Table 2: Parameter estimates with standard errors in parentheses (the 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)

autocorrelation function of its square and of its absolute value at Nagymaros. Similarly to the innovations of the ARMA-model, 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).

Put Figure 4 about here

5 Simulations

As mentioned earlier, the most important quantities from an application point of view concern the extremal behaviour of river flows. Questions such as "how often will the river exceed a certain high value" or "what is the average length of a flood above a given high level" are of interest. To answer them, high quantiles, tail behaviour and extremal clustering properties of the series should be estimated. Based on limit theorems about sample maxima and exceedances over a high threshold, classical extreme value theory makes inference about such quantities using only a very small fraction of the discharge series sampled at a daily frequency. However, the rich information on the dynamics of the series contained in the sample is lost with such an approach and, as Table 1 demonstrates, the necessary parameters (i.e. the parameters of the GPD) can be estimated only with a relatively large standard error. If the dynamics can be captured adequately, it opens a way to simulate much longer artificial discharge series and draw reliable inference on the extremes of the true discharges from the simulated ones. Our model intends to capture the dynamics but can only be accepted if simulations from it reproduce the main features, especially the extremes, of the empirical river flows so the goodness of fit has to be analysed from this slightly unusual point of view as well. Beyond classical criteria like independence of the residuals, the series simulated from the model ought to preserve the probability distribution, the high quantiles, the tail behaviour and the high-level clustering of the original discharges.

The simulation procedure is as follows. First – as it is common practice in the hydrological literature (see e.g. Montanari et al., 1997) – seasonal resampling is used to generate an 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 (18000-36000 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. 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 2.

Figure 5 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 calculated from a 1000-year long synthetic series at each site.) Figure 6 compares various quantiles (50%, 75%, 90%, 95%, 98%, 99%, 99.5%, 99.9%, 99.95%) and the sample maximum of the empirical discharge series to the corresponding quantities of 50 synthetic series of the same length. The quantiles fit generally well at sites of River Danube (upper row), while some quantiles are underestimated at sites of River Tisza. However, the overall fit of the probability density is adequate at the latter sites, too.

Put Figure 5 about here

Put Figure 6 about here

Let us compare specifically the extremal behaviours, i.e. the tails and the extremal clustering tendencies of the simulated and original series. Figure 7 shows the histogram of the estimated shape, scale parameters and the expected values of the GPD-s fitted to the upper tails of 500 simulated series at Nagymaros, the vertical lines indicating the results for the observed series. (See also Table 1. 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 equal to 100 years.) The median and mean of the simulated shape parameters (0.04) are a bit higher than zero. However, the observed shape parameter (-0.10) is still well within the acceptable range since around 10% of the simulated shape parameters lie below this level. 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 GPD-s fitted to the simulations approximate well the same quantity of the observed series. (This quantity estimates the average height of exceedance above the chosen threshold and is also an important hydrological parameter.)

Put Figure 7 about here

Finally turning to the clustering tendency of high-level exceedances, let u_n denote a sequence of thresholds such that $n(1 - F(u_n)) \to \tau > 0$ where $F(x)$ is the marginal distribution function of a stationary series X_t. Let N_n be the point process of exceedances:

$$N_n = \sum_{t=1}^{n} \nu_{t/n} I(X_t > u_n)$$

i.e. N_n puts a point at t/n if X_t exceeds u_n. It is well known that under appropriate distributional mixing conditions N_n converges weakly to a compound Poisson-process with intensity $\Theta\tau$ where Θ is the so-called extremal index. As another definition, Θ is the reciprocial of the expectation of the i.i.d. cluster size distribution in the compound Poisson-process. Heuristically, high-level exceedances in a stationary time series occur in clusters and $1/\Theta$ is the average cluster size in the limit, hence its estimation is crucial in various applied problems. (For details on the extremal index, see Embrechts et al. (1997) and the references therein.)

A threshold-dependent estimator of the extremal index can be obtained by the so-called blocks method. 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 which contain at least one exceedance of u. Furthermore, let N denote the total number of exceedances of u. Then the block estimator of Θ is given by

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

If $u = u_n$ and r tends to infinity at an appropriate rate, $\hat{\Theta}_n \to \Theta$ a.s. (As an approximation we get $\hat{\Theta}_n \approx K/N$, directly indicating the link to the definition as the reciprocal of the average cluster size.) Figure 8 displays the estimates of Θ for the observed and 50 simulated series at site Nagymaros. Block lengths r are chosen as 45, 90, 180 and 365 days and the u-values are determined as various quantiles (95%, 97.5%, 99%, 99.5% and 99.9%) of the observed water discharge distribution. The estimates do not vary much with r but depend very strongly on the threshold chosen. For instance, for the observed series at Nagymaros, the average cluster size of exceedances above the 95% quantile is estimated to lie between 5.8 and 8.3 days (depending on r) whereas above the 99.9% quantile it is between 2.7 and 2.9 days. In general the agreement between the empirical and simulated estimates is good, although at some levels the clustering is stronger (i.e. Θ is smaller) in the simulated series than in the observed one. 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 well captured by the model, too.

Put Figure 8 about here

To summarize, the above described and theoretically examined light-tailed conditionally heteroscedastic model captures the most important properties of empirical river flow series, making it a promising candidate for analysing water discharge measurements.

Acknowledgement

This research was partially funded by Hungarian National Research and Development Project, No. 3/067/2001 and by Hungarian National Research Fund OTKA, grant No. T047086.

References

BORKOVEC, M. and KLÜPPELBERG, K. (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. The Annals of Applied Probability 11, 1220-1241.

BOZSÓ, D., RAKONCZAI, P. and ZEMPLÉNI, A. (2005) Floods of River Tisza and of its tributaries. Statisztikai Szemle 83, 919-936 (in Hungarian).

BÜHLMANN, P. and MCNEIL, A. (2002) An algorithm for nonparametric GARCH modelling. Journal of Comput. Statist. and Data Analysis 40, 665-683.

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. and 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.

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.

FRANCQ, C. and ZAKOIAN, J. M. (1998) Estimating linear representations of nonlinear processes. Journal of Statistical Planning and Inference 68, 145-165.

FRANCQ, C., ROY, R. and ZAKOIAN, J. M. (2005) Diagnostic checking in ARMA models with uncorrelated errors. Journal of the American Statistical Association 100, 532-544.

GUEGAN, D. and DIEBOLT, J. (1994) Probabilistic properties of the β-ARCH-model. Statistica Sinica 4, 71-87.

HAMILTON, J. D. (1994) Time series analysis. Princeton University Press, Princeton, N. J.

KRISTENSEN, D. and RAHBEK, A. (2005) Asymptotics of the QMLE for a class of ARCH(q) models. Econometric Theory 21, 946-961.

MASRY, E. and TJOSTHEIM, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory 11, 258-289.

MEYN, S. P. and TWEEDIE, R. L. (1993) Markov Chains and Stochastic Stability. Springer, London.

MONTANARI, A., ROSSO, R. and TAQQU, M. S. (1997) Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. Water Resources Research 33, 1035-1044.

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.

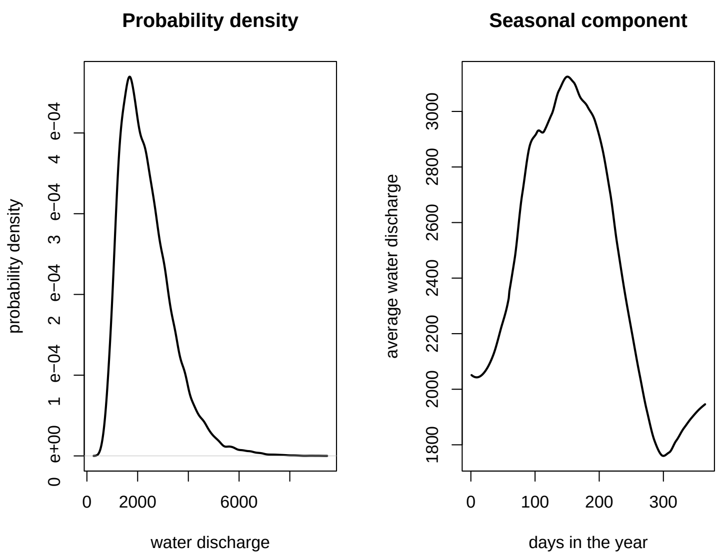

Figure 1: Probability density and the estimated seasonal component at Nagymaros

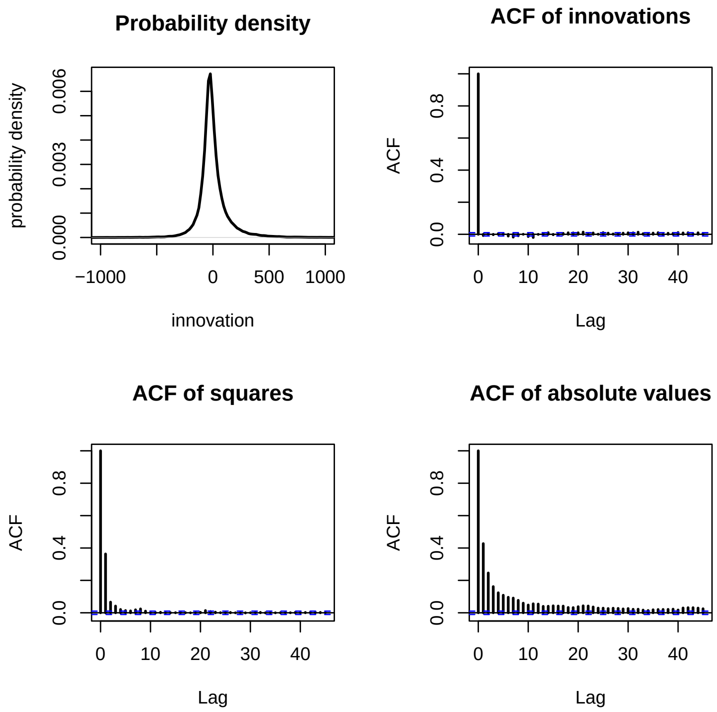

Figure 2: Probability density of the innovations, autocorrelation function of the innovations, of their squares and of their absolute values at Nagymaros

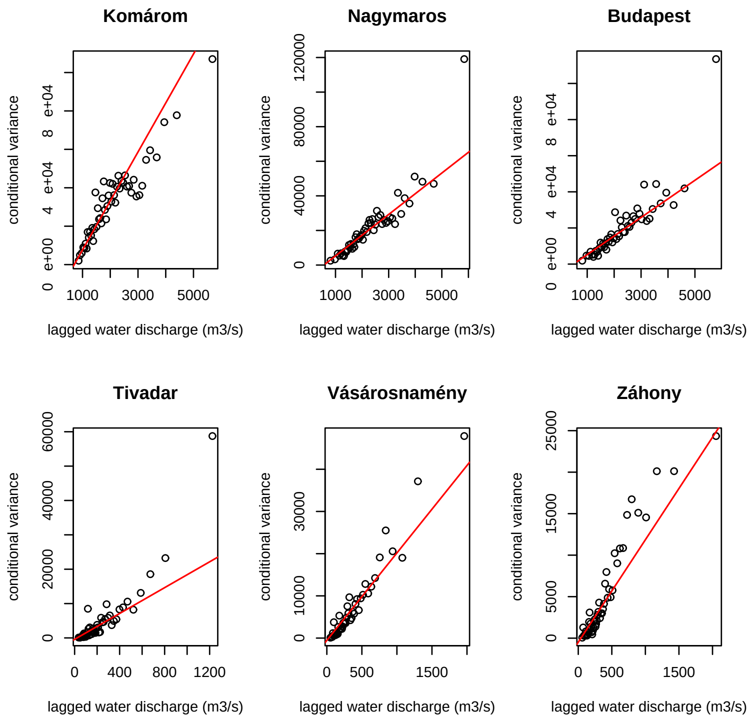

Figure 3: Conditional variance as a function of lagged water discharge at six stations, the lines showing the fitted relationships with maximum likelihood estimates α_0 and α_1

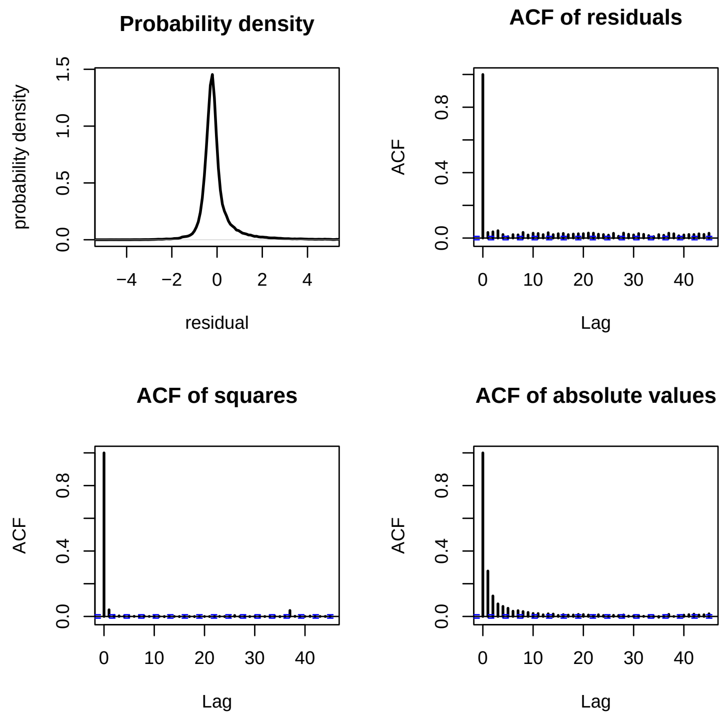

Figure 4: Probability density of the noise, autocorrelation function of the noise, of its square and of its absolute value at Nagymaros

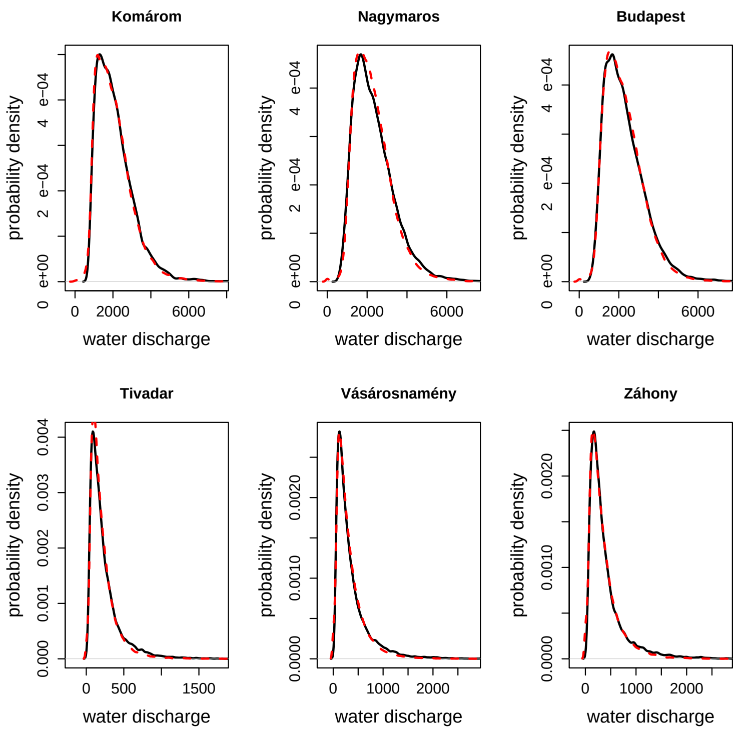

Figure 5: Probability density of the empirical (continuous) and simulated (dashed) series at six monitoring stations. The almost perfect fit makes the lines almost indistinguishable.

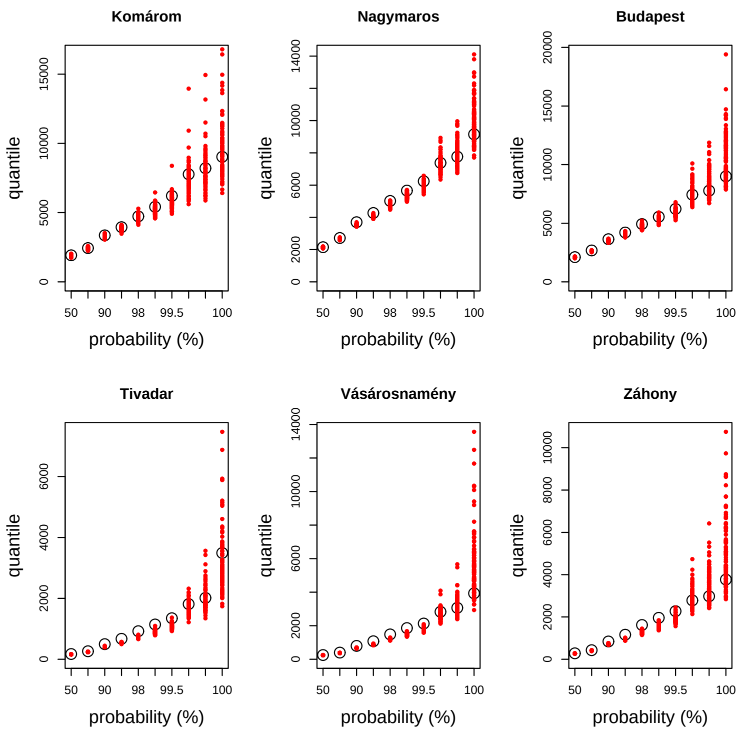

Figure 6: Quantiles (50%, 75%, 90%, 95%, 98%, 99%, 99.5%, 99.9%, 99.95%) and sample maxima of the empirical (large dots) and of 50 simulated (small dots) series at six monitoring stations

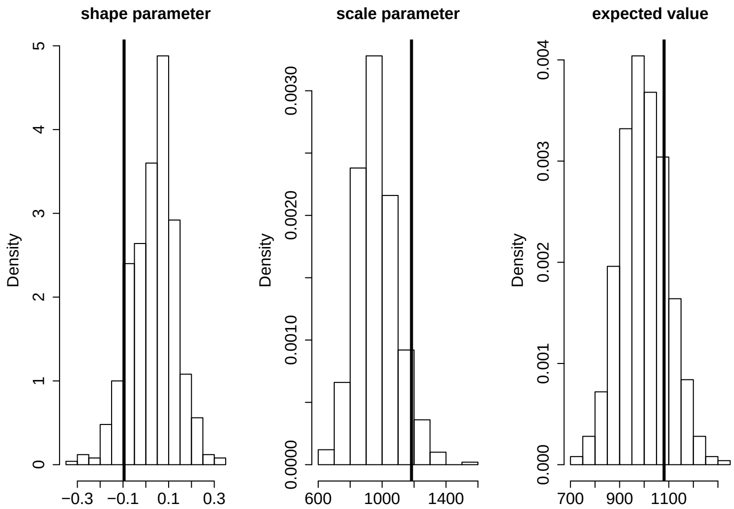

Figure 7: Histograms of shape and scale parameters and expected values of the GPD-s fitted to the upper tails of 500 simulated series at Nagymaros. Vertical lines indicate the observed values, the threshold is chosen as 4300 m^3/s.

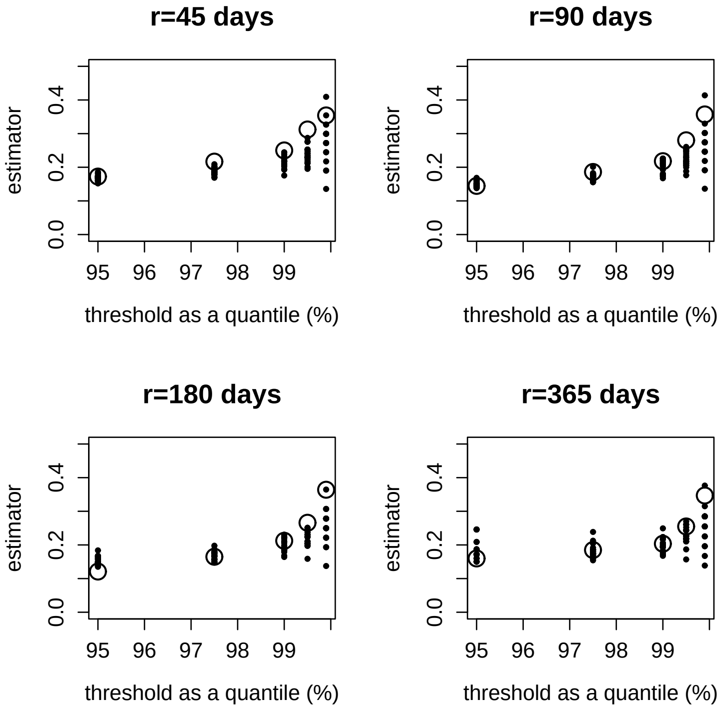

Figure 8: $\hat{\Theta}_n(u, r)$ (with various u and r) for the original (large dots) and 50 simulated (small dots) series at site Nagymaros