Modeling rare events through a pRARMAX process

Marta Ferreira[1] Luísa Canto e Castro[2]

[1]University of Minho, Portugal

CMAT

[2]Faculty of Sciences, University of Lisbon, Portugal

DEIO-CEAUL

6th Conference on Extreme Value Analysis

Fort Collins, Colorado - EUA June 22-26, 2009

Motivation

- Most often, the processes under study when we are making inference on maxima, show short term dependence which can conveniently be modeled by a markovian structure;
- Data presenting sudden large peaks (e.g. telephone signals, stock market prices), are potential candidate for modeling as an ARMA process with heavy-tailed noises;
- ARMAX processes ($X_i = c X_{i-1} \vee Z_i$, $_{c \in (0,1)}$) present a nice treatment in what concerns extremal properties (Alpuim, 1989);
- MARMA processes have been already considered a good alternative for modeling this kind of data - similar paths to the heavy-tailed ARMA and present a nice treatment in what concerns extremal properties (Davis and Resnick, 1989).

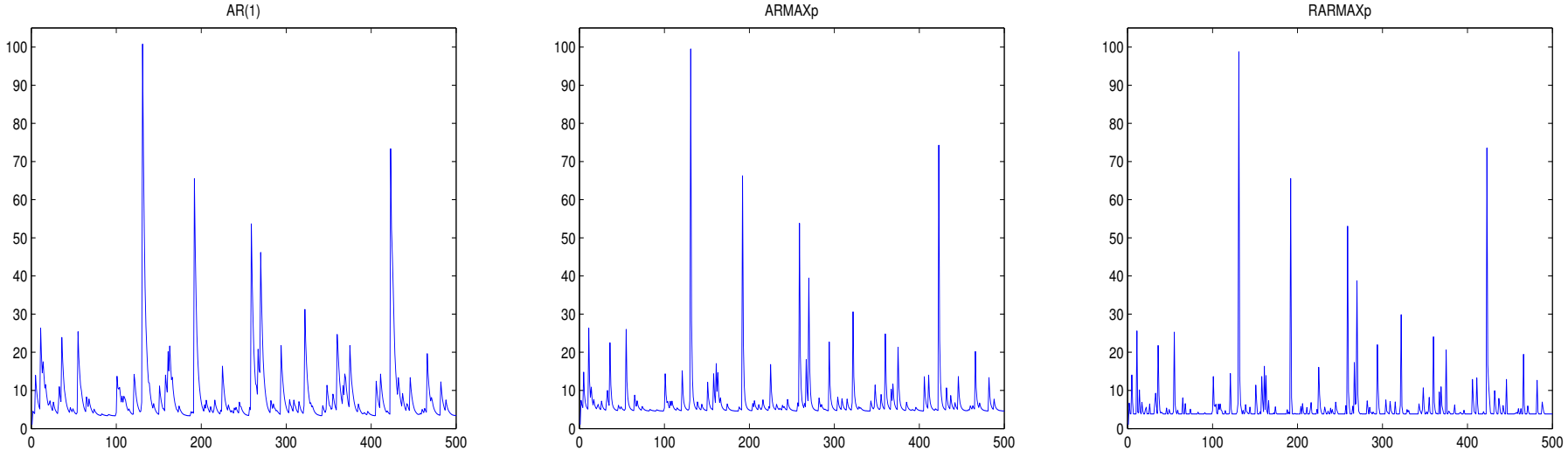
- The power max-autoregressive process, pARMAX, $X_i = X_{i-1}^c \vee Z_i$, $c \in (0,1)$, (Ferreira and Canto e Castro 2008), also presents similar paths to the heavy-tailed ARMA and easily derived extremal properties ;

- Motivation for model pARMAX:
  - the power parameter, **c**, is directly related to the coefficient of asymptotic tail dependence (η) of Ledford and Tawn (1996, 1997), calculated for random pairs $(X_i, X_{i+m}) \longrightarrow$ **c** can be estimated using η estimators.

$\twoheadrightarrow$ p**RARMAX** - a generalization of pARMAX by introducing a random component

- the same extremal properties and also similar sample paths;
- more applicable for modeling real phenomena: we present a methodology based on minimizing the Bayes risk in classification theory and analyze this procedure through a simulation study

500 realizations of an AR(1), a pARMAX and of a pRARMAX, respectively



Model pRARMAX

Consider:

- $\{Z_i\}_{i\in\mathbb{Z}}$ and $\{U_i\}_{i\in\mathbb{Z}}$ i.i.d. copies of Z and U, respectively;
- U_j independent of $\{Z_i\}_{i\in\mathbb{Z}}$, for all integer j;
- Z and U both with nonnegative support and non degenerate d.f. F_Z and F_U (resp.);

A stationary process $\{X_i\}_{i\in\mathbb{Z}}$ is pRARMAX process if it satisfies,

$$X_i = U_i\, X_{i-1}^c \;\vee\; Z_i\,, \qquad 0<c<1,\; i=0,\pm1,\pm2,\ldots. \tag{1}$$

(If $U=1$ we have pARMAX).

Assuming $-\infty \leq E\log U < 0$ and $E\log(Z \vee 1) < \infty$,

$$X_n = \bigvee_{j=0}^{\infty} \prod_{i=0}^{j-1} U_{n-i}^{c^i} Z_{n-j}^{c^j} \tag{2}$$

is a.s. finite and its law is the unique such that (1) holds. ($\prod_{i=0}^{j-1} U_{n-i}^{c^i} = 1$ for $j = 0$)

Any d.f. K that satisfies the stationarity equation,

$$K(x) = F_Z(x) \int_{\mathcal{U}} K((x/u)^{1/c}) dF_U(u)\,, \tag{3}$$

is a stationary marginal d.f. of the process.

Example: *If Z has d.f.,*

$$F_Z(x) = \frac{1 - x^{-1/\gamma}}{1 - \frac{c\gamma}{1+c\gamma} x^{-1/(c\gamma)}} \mathbf{1}_{\{x \geq 1\}},$$

and $U \frown \mathrm{U}(0,1)$, then by (3), $K(x) = \left(1 - x^{-1/\gamma}\right)\mathbf{1}_{[1,\infty)}(x)$, is a non degenerate marginal distribution of $\{X_i\}_{i\in\mathbb{Z}}$. $\square$

Assuming that $\{X_i\}_{i\in\mathbb{Z}}$ is a stationary pRARMAX process with marginal d.f. K, then:

- it is in the same domain of attraction of innovations Z with the same tail index γ;
- it is *regenerative* and *aperiodic*, hence, β-*mixing*;
- satisfies condition $\mathrm{D}''(u_n)$ (Leadbetter and Nandagopalan 1989) for $(u_n)_{n\geq 1}$, such that, $1-K(u_n)=O(1/n)$;
- if Z is in the Fréchet domain of attraction, then $\theta=1$ and $\eta=\max(1/2,c)$ (based on random pairs (X_i,X_{i+1}))

$$\Downarrow$$

exceedances tend to occur singly as the threshold increases, similar to the i.i.d. case

advantage (for inferential purposes) if θ is replaced by a pre-asymptotic form on

$$P\left(\bigvee_{i=1}^{n} X_i \leq u_n\right) \sim \left(K(u_n)\right)^{n\theta} \quad , n\to\infty$$

based on Bortot and Tawn (1998), we consider the pre-asymptotic extremal index,

$$\theta(u) \sim 1-(t(u))^{1-1/\eta}\mathcal{L}(t(u)), \text{ as } u\to\infty \text{ with } t(u)=(1-K(u))^{-1}$$

Given an observed time series, how can we decide if pRARMAX is a suitable model?

If so, then $X_i = \max(U_i X_{i-1}^c, Z_i)$:

- X_i either comes from the first or the second component of the maximum.

- G_0: the set of X_i's that come from the second component (Z_i)
- G_1: the set of X_i's coming from the first component ($U_i X_{i-1}^c$)

We will say that the model fits, if for the observations in $\mathtt{G}_0$ the assumptions considered for Z are not rejected, and for the observations in $\mathtt{G}_1$, when divided by X_{i-1}^c, the hypotheses assumed for U are not rejected as well.

- if $X_i \geq X_{i-1}^c$ then $X_i \in \mathrm{G}_0$
- if $X_i < X_{i-1}^c$ (?) we need to establish a criterion for decision

- F_0 is the d.f. of X_i in G_0: $F_0(x) = P(X_i \leq x, X_i < X_{i-1}^c, U_i X_{i-1}^c \leq Z_i)$
- F_1 is the d.f. of X_i in G_1: $F_1(x) = P(X_i \leq x, X_i < X_{i-1}^c, U_i X_{i-1}^c > Z_i)$

Similar to the procedure in *classification theory*, for each λ ($0 \leq \lambda \leq 1$),

$$\mathcal{B}_\lambda = \left\{ t : \frac{\pi_0 f_0(t)}{\pi_0 f_0(t) + \pi_1 f_1(t)} \leq \lambda \right\}, \qquad \pi_0 = P(X_i \in G_0) = 1 - \pi_1,$$

is the region that minimizes the *Bayes error* (Storey 2003; Rohatgi 1976) and we classify X_i in G_1 if $X_i \in \mathcal{B}_\lambda$.

Considering a pRARMAX process of the given Example:

$$f_1(x) = \frac{x^{-1/(c\gamma)-1} F_Z(x)}{1+c\gamma} \qquad \text{and} \qquad f_0(x) = \frac{x^{-1/(c\gamma)} f_Z(x)}{1+c\gamma}$$

For each fixed λ we obtain:

$$\mathcal{B}_\lambda = \left\{ t : \frac{\pi_0 t^{-1/(c\gamma)} f_Z(t)}{\pi_0 t^{-1/(c\gamma)} f_Z(t) + (1-\pi_0) t^{-1/(c\gamma)-1} F_Z(t)} \leq \lambda \right\} = \{ t : t \geq t_\lambda \},$$

- The captured observations U are no longer standard Uniform but Uniform conditional on X_i of the form $U_i X_{i-1}^c$ and under the criterion, $X_i > t_\lambda$. In this case they have distribution, $Beta\left(\frac{1}{c\gamma}+1, 1\right)$.

1. Test if, $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, is in the Fréchet($\gamma_X$) domain of attraction, for some $\gamma_X > 0$, i.e., test the extreme value condition, $X \in \mathcal{D}(G_{\gamma_X})_{\gamma_X \geq 0}$ (Dietrich *et al.* 2002) and estimate γ_X (e.g., Hill estimator);

2. Estimate parameter $\mathbf{c}$ of pRARMAX through the estimation of η (which is the tail index, γ_T, of the transformed r.v.'s $T_i^{(n)} = \min((n+1)/(n+1-R_i), (n+1)/(n+1-R_{i+1}))$, $i = 1, \ldots, n$, where R_i denotes the rank of X_i among $(X_1, \ldots, X_n)$;

3. Capture innovations $\mathbf{Z}$: if $X_i > X_{i-1}^{\widehat{c}}$, where $\widehat{c} = \widehat{\gamma_T}$ was obtained in 2., then $X_i = Z_i$, and test if $\mathbf{Z}$ is in the Fréchet domain of attraction;

4. Capture the random coefficients U: if $X_i < X_{i-1}^{\widehat{c}}$ and $X_i \in \mathcal{B}_\lambda$ ($X_i > t_\lambda$), then $U_i = X_i / X_{i-1}^{\widehat{c}}$;

5. Test if the sample of r.v.'s U captured above has distribution $Beta(1/(\widehat{\gamma}_X \widehat{c}) + 1, 1)$ (e.g., Kolmogorov-Smirnov test).

λ must be small in order to penalize more considerer X_i of the form $U_i X_{i-1}^c$, when in fact it is of the form Z_i.

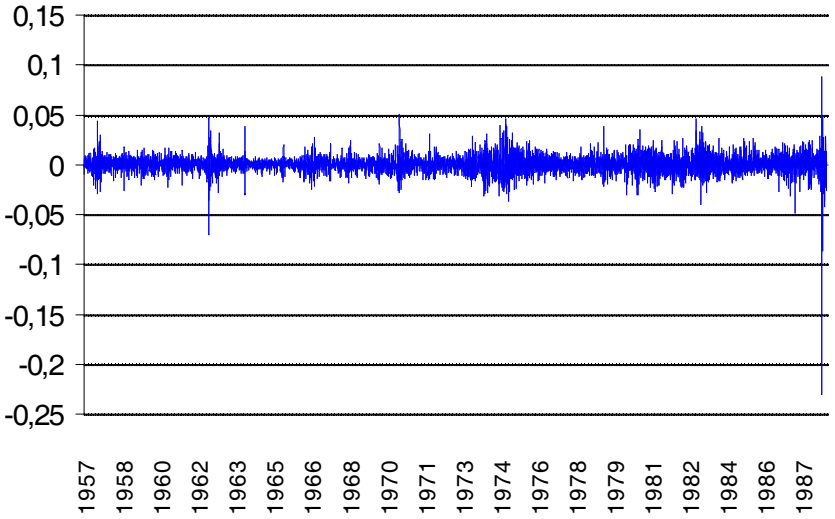
- as the first ones are less frequent, they must be captured with less uncertainty (although they should be in enough amount) in order to achieve reliable results in the goodness-of-fit test on step 5.

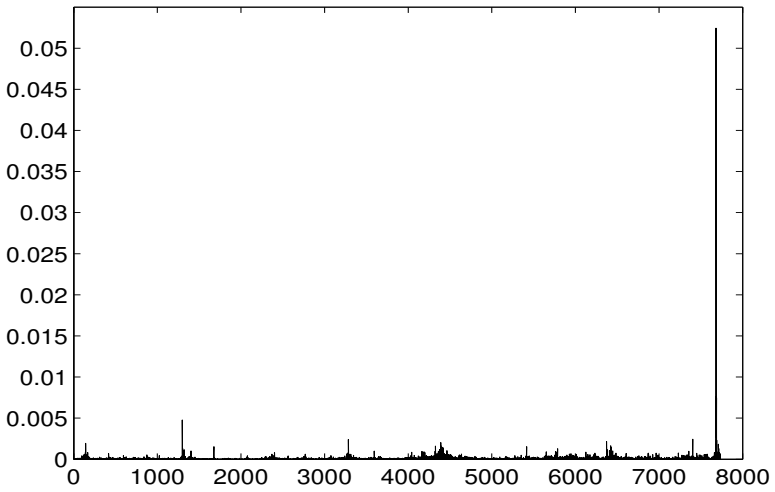
Suggestion for λ based on a simulation study

	$\gamma = 0.4$			$\gamma = 0.8$			$\gamma = 1.5$		
n	1000	3000	5000	1000	3000	5000	1000	3000	5000
$c = 0.5$	—	—	—	—	—	—	0.25	0.15	0.1
$c = 0.6$	—	—	—	—	0.25	0.2	0.15	0.1	0.05
$c = 0.7$	—	—	—	—	0.15	0.15	0.1	0.05	0.05
$c = 0.8$	—	0.35	0.25	0.2	0.1	0.1	0.05	0.05	0.05
$c = 0.9$	—	0.25	0.15	0.1	0.1	0.1	0.05	0.05	0.05

An application to financial data: *S&P500* stock market index

Consider the square of the log-returns $R_i = \log P_{i+1}/P_i$, $1 \leq i \leq n-1$ ("volatility" can be measured through $|R_i|$ or R_i^2); P_i is the closing index of the ith trading day in years 1957-1987, $n = 7733$; large peak: Monday stock market crash 19th October 1987.





Obtain a data series that can be modeled by a pRARMAX with standard Pareto marginals $\dashrightarrow$ robust regression: our analysis focuses on the transformed data, $X_i = aR_i^2 + b$, with estimates $a = 13618.3$ and $b = 1.1$.

Step 1: Test the extreme value condition, $X \in \mathcal{D}(G_{\gamma_X})_{\gamma_X \geq 0}$ and estimate γ_X

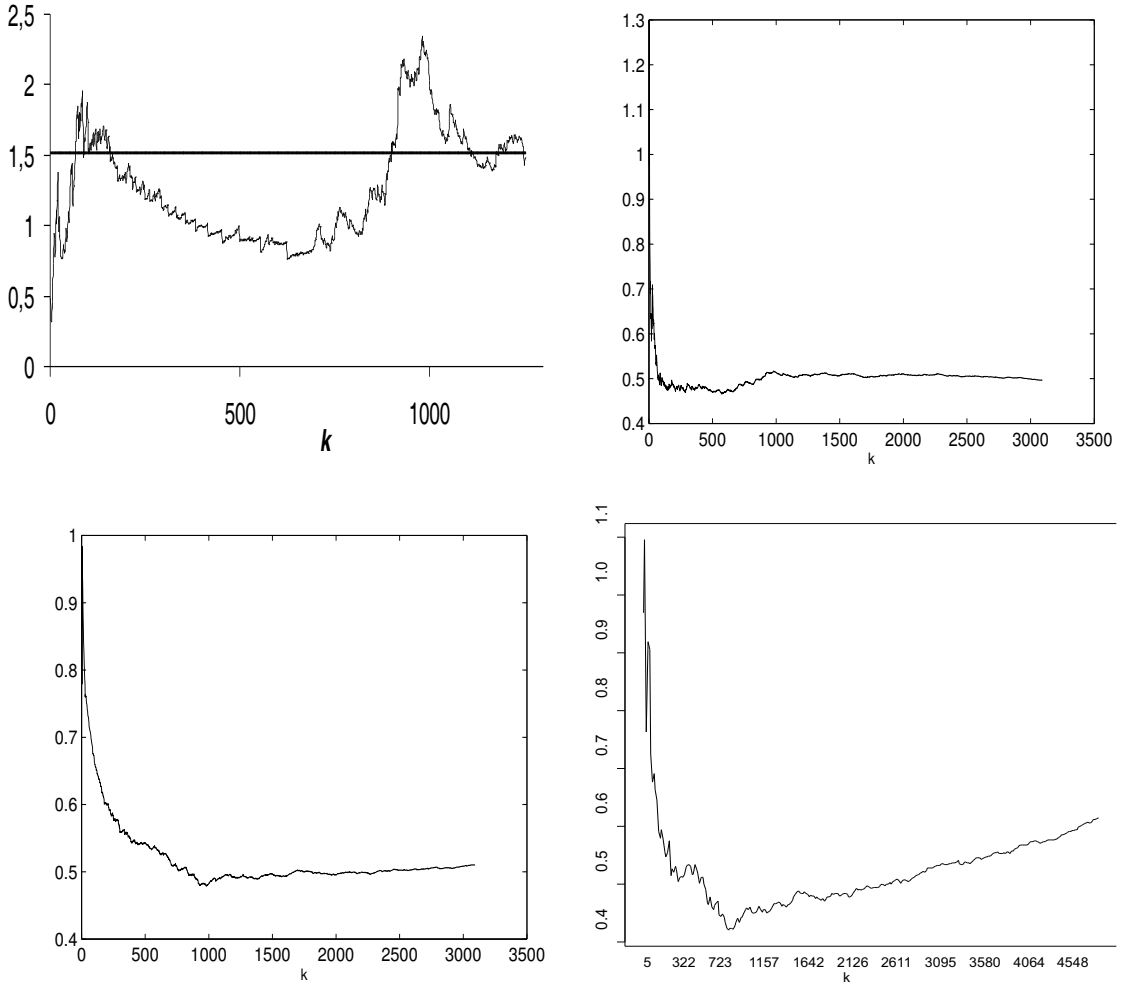


Figure: Top-left: sample path of the test statistic and horizontal line is the critical value above which reject $X \in \mathcal{D}(G_\gamma)_{\gamma \geq 0}$; sample paths of Hill, moment and maximum likelihood estimators, resp.

- extreme value condition not rejected for $165 \lesssim k \lesssim 900$; $\widehat{\gamma}_X \approx 0.5$

Evaluate the effect on the tail of the large peak: data considered until the day before

$\dashrightarrow \gamma_X \approx 0.4$

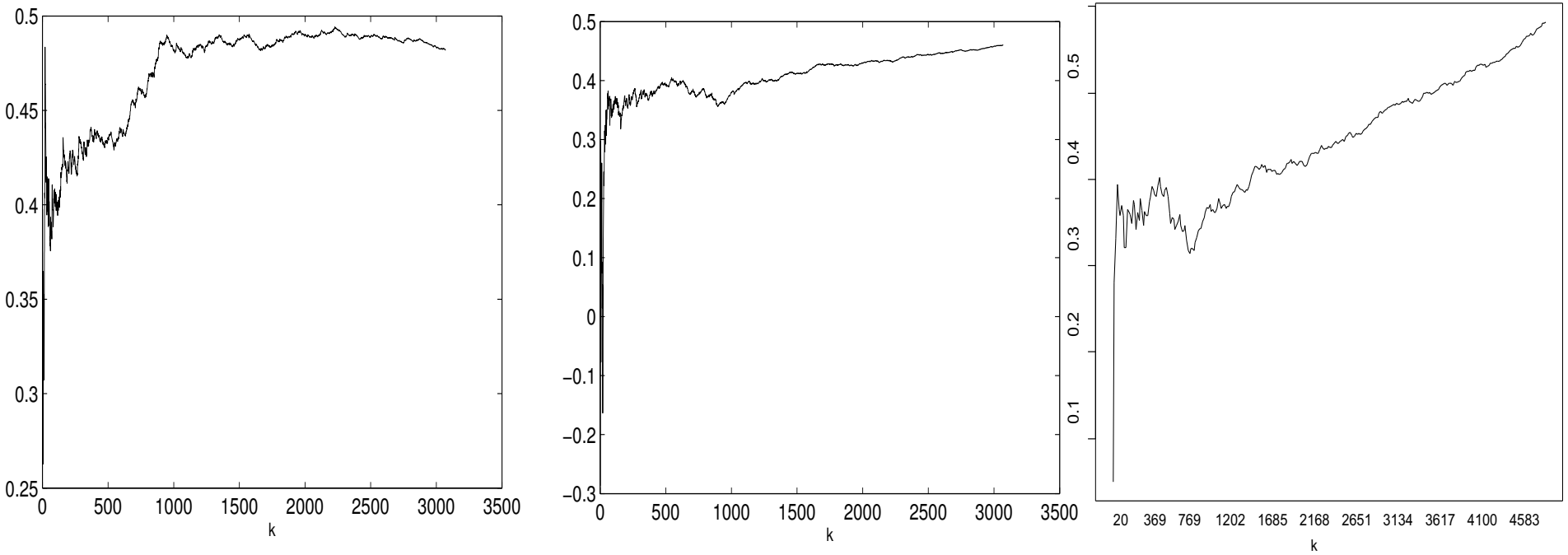


Figure: Sample paths of Hill (left), moment (center) and maximum likelihood (right).

Step 2: estimate parameter c through η (tail index of the transformed $\mathbf{T}^{(\mathbf{n})}$)

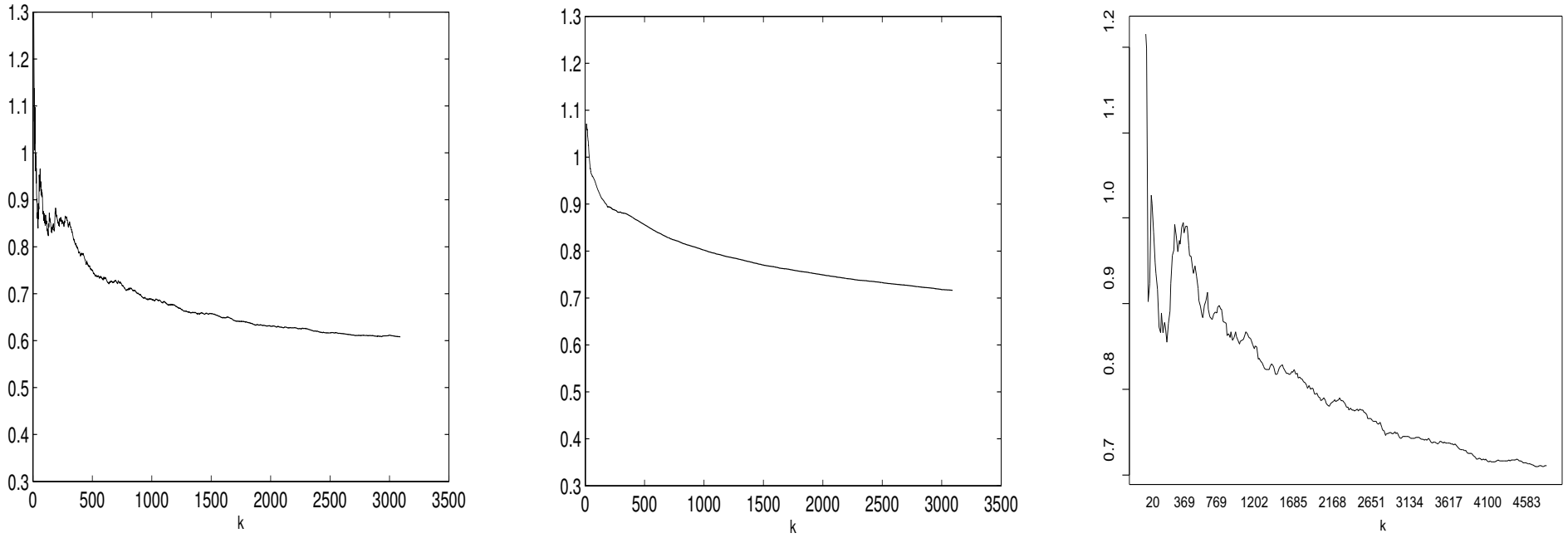


Figure: Sample paths of Hill (left), moment (center) and maximum likelihood (right) estimators of η.

the estimate is about 0.85; however, due to some stability also around 0.75, we consider more than one scenario: $\widehat{c} = 0.85$, $\widehat{c} = 0.8$ and $\widehat{c} = 0.75$.

Step 3: Capture innovations $\mathbf{Z}$ $(X_i > X_{i-1}^{\widehat{c}})$ and test $Z \in \mathcal{D}(G_\gamma)_{\gamma \geq 0}$

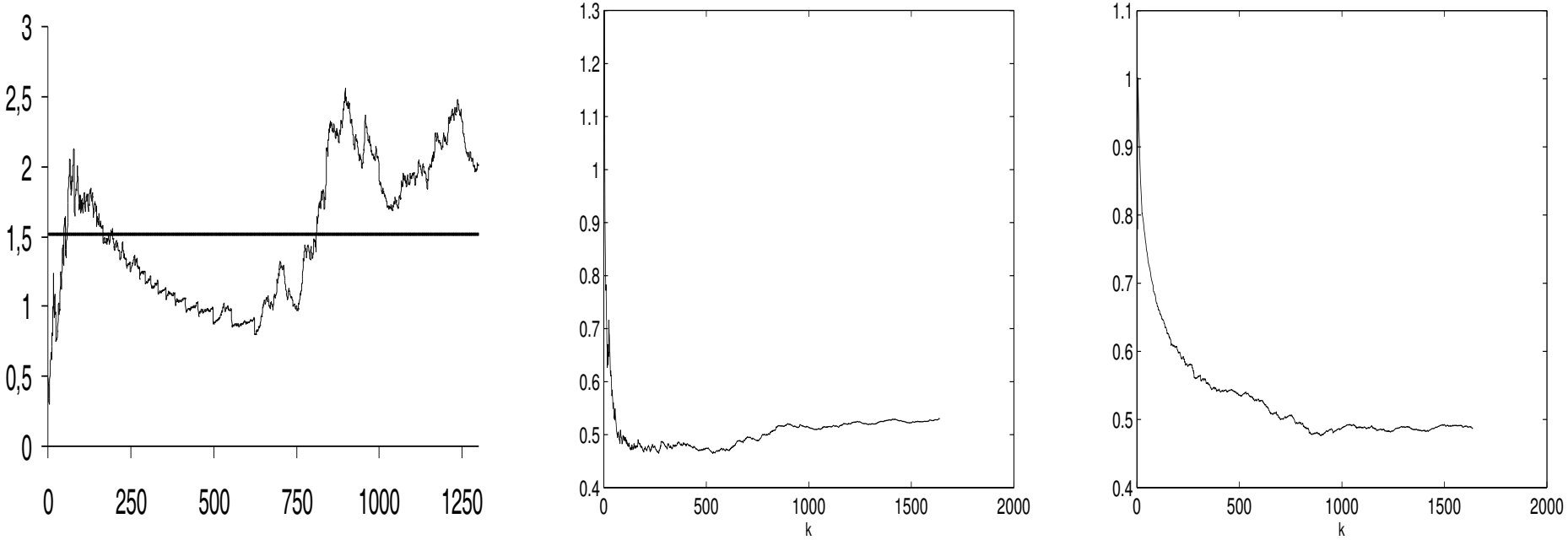


Figure: Left: sample path of the test statistic, $PE_{k,n}$, for the innovations, Z, captured from $\mathbf{X}$ on step 3.; sample paths of Hill (center) and moment (right) estimators, for Z.

Steps 4 and 5: Capture U (if $X_i < X_{i-1}^{\widehat{c}}$ and $X_i \in \mathcal{B}_\lambda$ then $U_i = X_i / X_{i-1}^{\widehat{c}}$; $\lambda = 0.05, \ldots, 0.5$, $\widehat{c} = 0.85, 0.8, 0.75$) and test $U \frown Beta(1/(0.5 * \widehat{c}) + 1, 1)$ (K-S test)

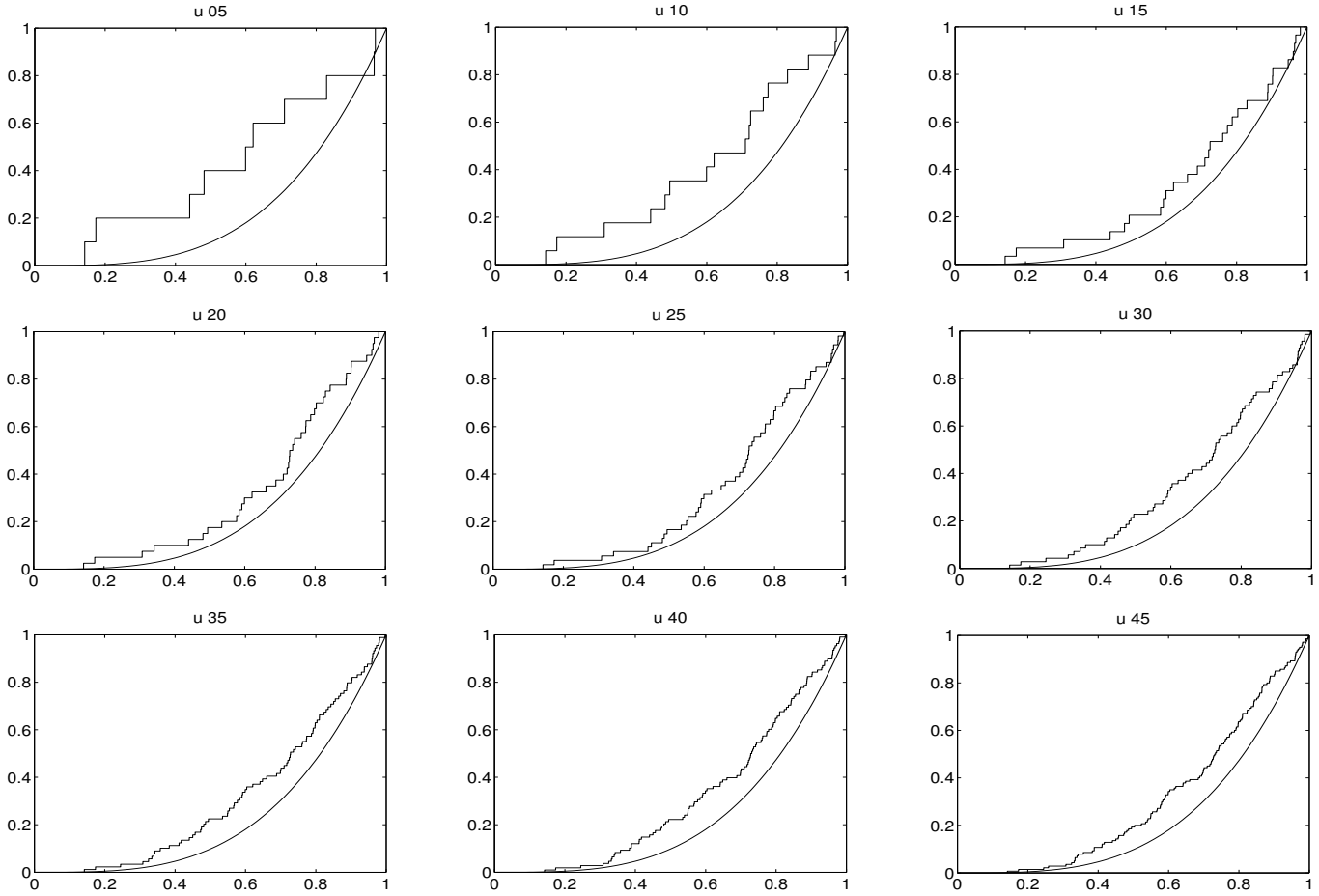


Figure: Empirical and theoretical d.f.'s of the captured U for $\widehat{c} = 0.85$.

rejection for $\lambda \geq 0.20$; $\lambda = 0.15$ matches the simulation study (with 29 obs. captured).

Taking $\widehat{c} = 0.8$ (less catches), then $Beta(1/(0.5 * 0.8) + 1, 1)$ is rejected for $\lambda \geq 0.3$, with the best fit occurring for $\lambda = 0.2$, matching once again the simulation results.

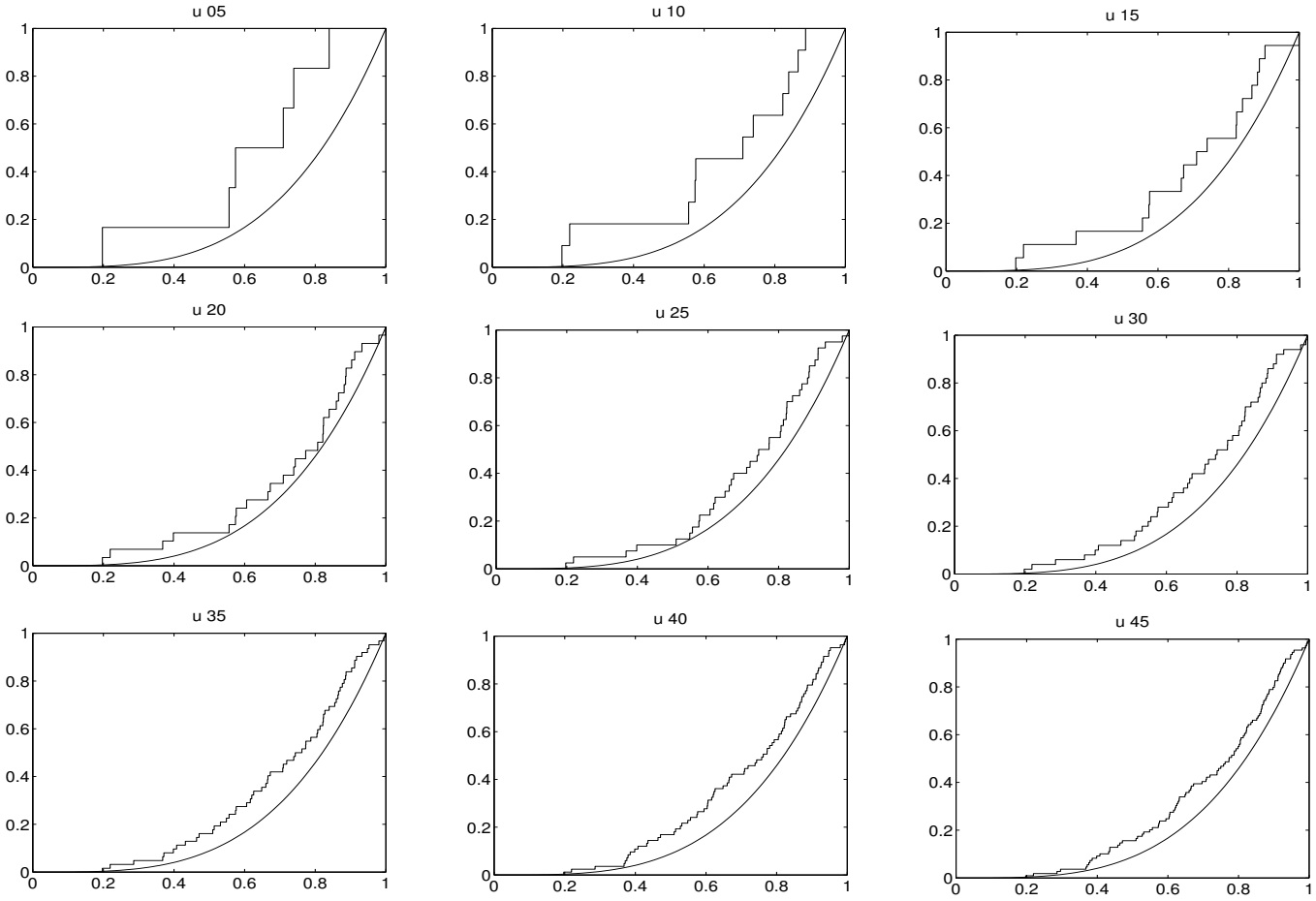


Figure: Empirical and theoretical d.f.'s of the captured U for $\widehat{c} = 0.8$.

For $\widehat{c} = 0.75$ (even less catches), only rejects $Beta(1/(0.5*0.75)+1,1)$ for $\lambda = 0.5$.

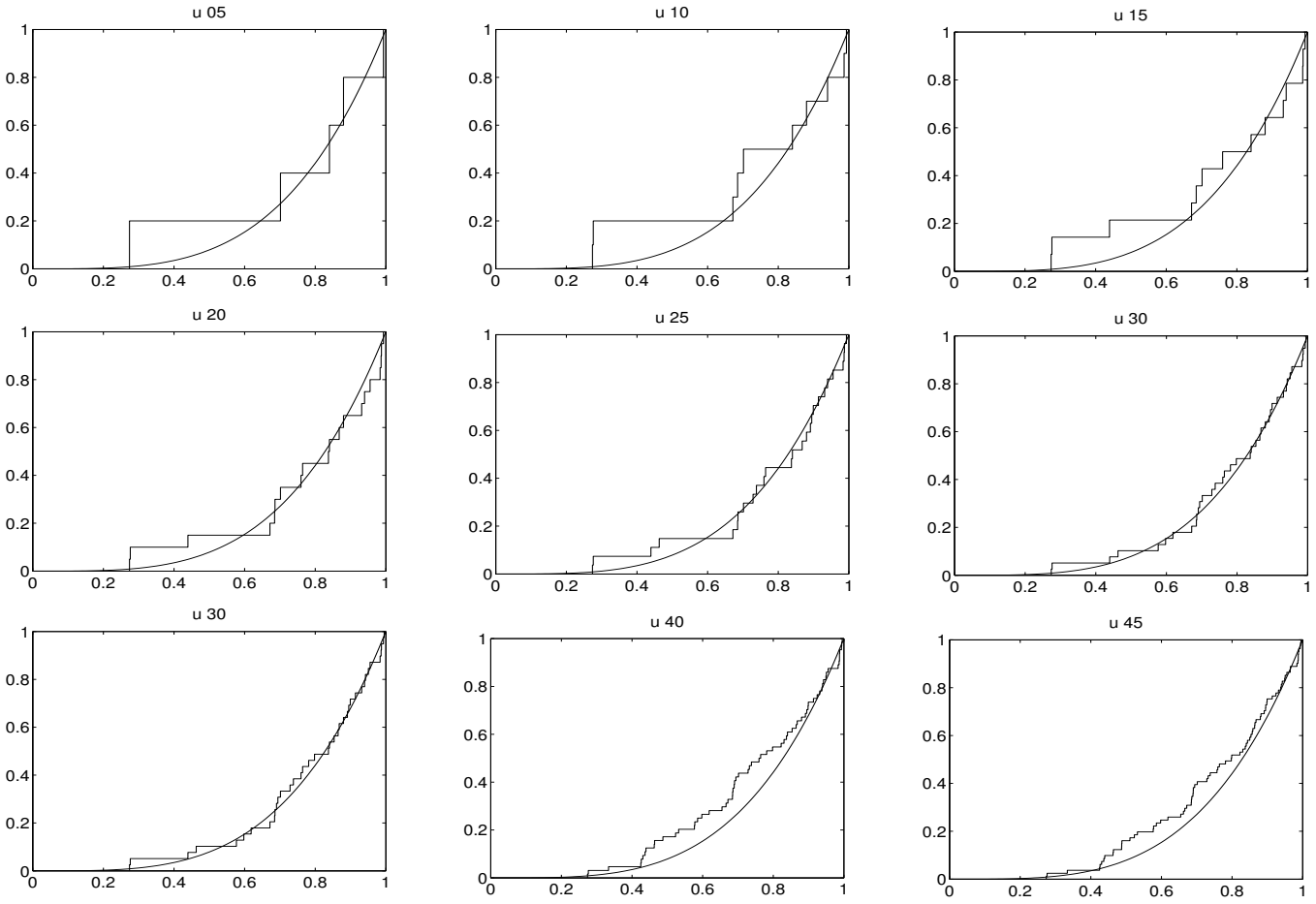


Figure: Empirical and theoretical d.f.'s of U captured with $\widehat{c} = 0.75$.

Hence a pRARMAX can be a good option for the modeling of the transformed data $\mathbf{X}$.

Estimating the probability that the maximum volatility exceeds a risky amount (e.g. 0.2), using $P\left(\bigvee_{i=1}^{n} X_i \leq u_n\right) \sim \left(K(u_n)\right)^{n\theta}$ with $\theta = 1$ (the true value for pRARMAX) and θ replaced by the pre-asymptotic $\theta(u) = 1 - u^{\frac{1}{\gamma}(1-1/c)}$; for $\gamma = 0.5, 0.45, 0.4$ ($\widehat{\gamma} = 0.4$, when considering data only until the day before the big market crash) and the 3 scenarios for parameter c (0.85, 0.8 and 0.75)

	$\gamma = 0.5$	$\gamma = 0.45$	$\gamma = 0.4$
$c = 0.85$	0.053222	0.010474	0.002881
$c = 0.8$	0.057372	0.015703	0.003028
$c = 0.75$	0.059219	0.01609	0.00308
$\theta = 1$	0.060295	0.016281	0.003101

the probability estimates decrease significantly with the decrease of γ, but very small changes with c (γ is crucial)

the higher the γ and the c, the greater the differences in estimates

Dietrich, D., de Haan, L. e Hüsler, J., Testing extreme value conditions. *Extremes* **5**, 71-85, 2002.

Ferreira, M., Canto e Castro, L., Asymptotic and pre-asymptotic tail behavior of a power max-autoregressive model. Notes and communications No. 20. Dept of Statistics and Operations Research, University of Lisbon, 2008. (submitted)

Ferreira, M. e Canto e Castro, L., Tail and dependence behaviour of levels that persist for a fixed period of time. *Extremes* **11**, 113-133, 2008.

Ferreira, M. and Canto e Castro, L., Modeling Rare Events Through a pRARMAX Process. Notes and Communications No. 3. Dept of Statistics and Operations Research, University of Lisbon, 2009. (submitted)

Leadbetter, M. R. e Nandagopalan, S., On Exceedance Point Processes for Stationary Sequences under Mild Oscillation Restrictions. *Extreme Value Theory.* (J. Hüsler e R.-D. Reiss, eds.), Springer-Verlag, 69-80, 1989.

Ledford, A. e Tawn, J. A., Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169-187, 1996.

Ledford, A. e Tawn, J. A., Modelling Dependence within joint tail regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **59**, 475-499, 1997.

Rohatgi, V.K., *An Introduction to Probability Theory and Mathematical Statistics*. Wiley, 1976.

Storey, J.D., The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value. *Ann. Statist.* **31**, 2013-2035, 2003.