PREDICTION IN SHORT AND EXTREMELY SHORT TIME SERIES

Jaromír Antoch

MODELLING SMART GRIDS 2018

Goals

- Prediction of the next value
- Estimation of the probability of exceedance over given threshold

for

- Very short time series, i.e. $7-10$ observations
- Extremely short time series, i.e. $4-6$ observations

and illustrate the idea on real data set

Goals

- Prediction of the next value
- Estimation of the probability of exceedance over given threshold

for

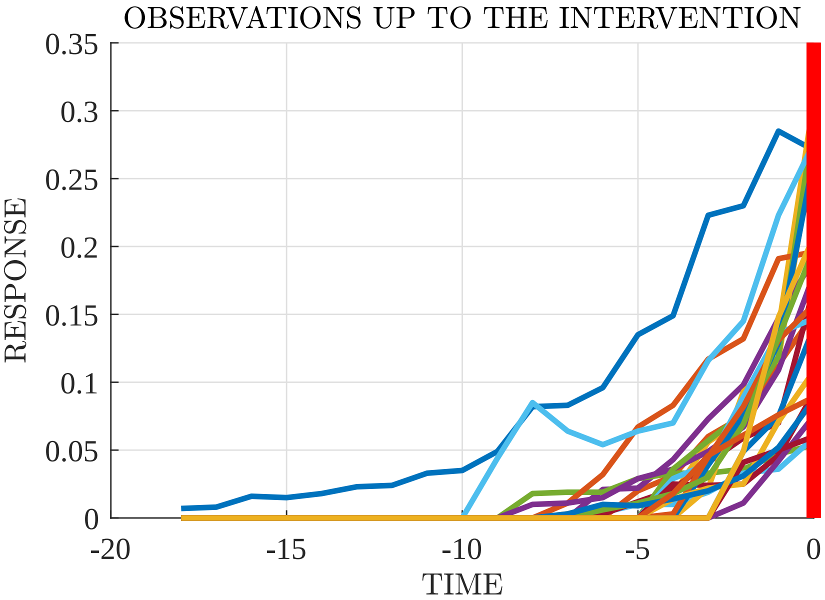
- Very short time series, i.e. $7-10$ observations
- Extremely short time series, i.e. $4-6$ observations

and illustrate the idea on real data set

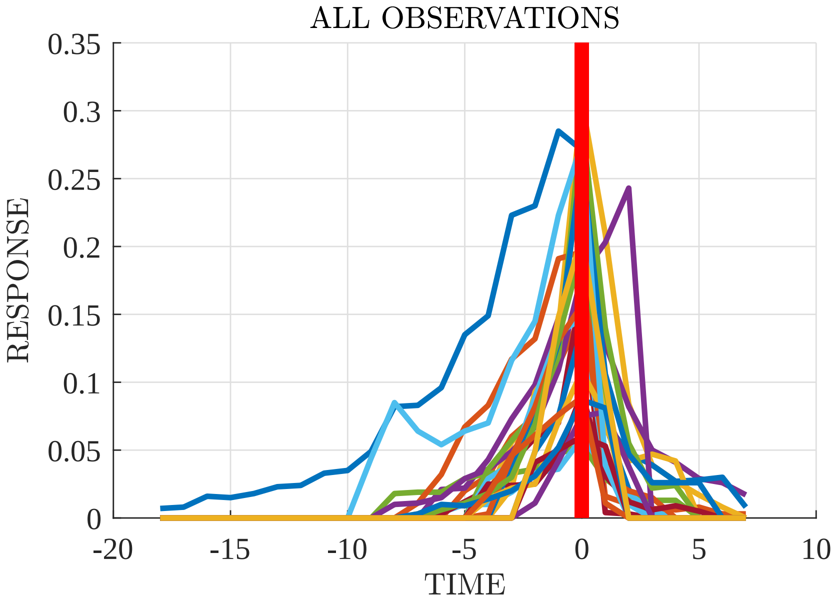
Basic ideas are

- To use growth curves
- To use prior idea about the character of data

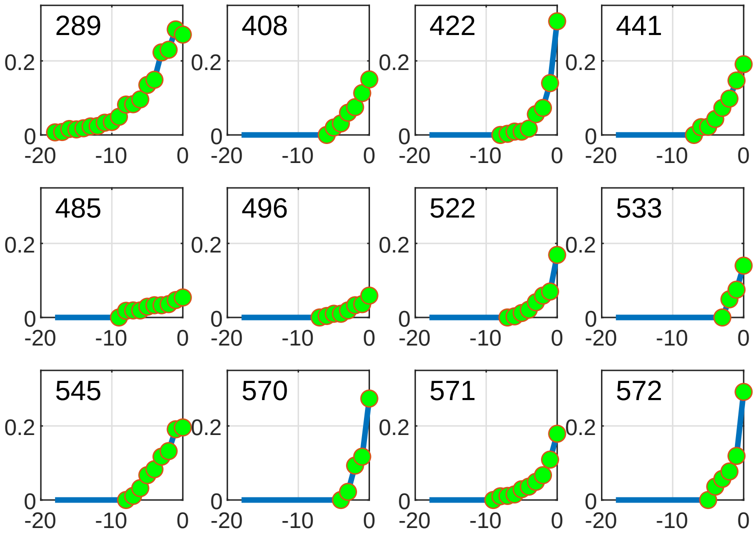
Data


OBSERVATIONS UP TO THE INTERVENTION
RESPONSE
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-20
-15
-10
-5
0
TIME

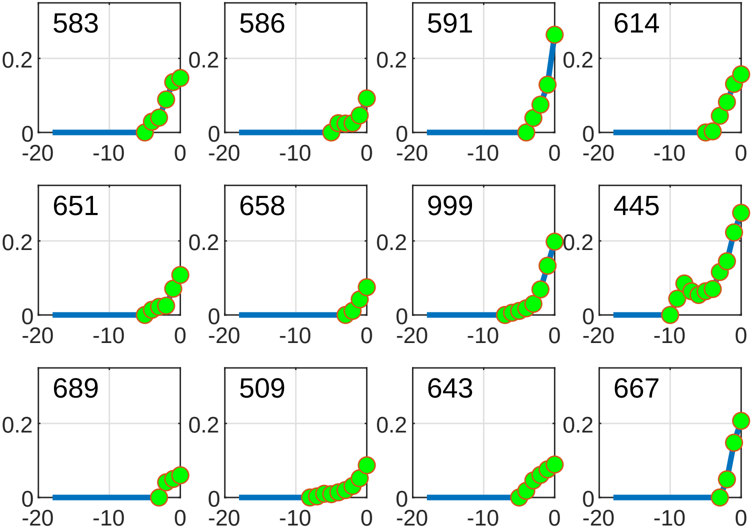
Data


ALL OBSERVATIONS
RESPONSE
TIME
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-20
-15
-10
-5
0
5
10

Data



Data



Models used

1. Additive allometric model
2. Multiplicative allometric model (time-power model)
3. Exponential model
4. Exponential model with additive error
5. Quadratic model
6. Autoregressive model
7. Gompertz' model.

Additive allometric model

- Has (in general) the form

$$y = \gamma + \alpha x^{\beta} + \varepsilon \tag{1}$$

- For $\gamma = 0$ and $\beta = 2$ we get quadratic model without absolute term, i.e.

$$y = \alpha x^2 + \varepsilon$$

Additive allometric model

- Has (in general) the form

$$y = \gamma + \alpha x^{\beta} + \varepsilon \tag{1}$$

- For $\gamma = 0$ and $\beta = 2$ we get quadratic model without absolute term, i.e.

$$y = \alpha x^2 + \varepsilon$$

- Griffits and Sandland, Biometrics (1984), discuss also allometric models with more than two variables and provide references to the literature of allometry.
- Recall that this type of models has been studied for over 50 years and, as a first approximation, it seems to fit a number of growth processes quite well.

Additive allometric model - cont.

- In study of allometry, an attention is focused on differences in shape associated with size or time.
- Instead of relating two size measurements x and y (e.g. length of bones) to time, we may be interested in relating them to each other.

- Suppose that two relative growth rates are given by

$$\frac{1}{x}\frac{dx}{dt} = k_x \qquad \text{and} \qquad \frac{1}{y}\frac{dy}{dt} = k_y$$

Defining $\beta = k_y/k_x$ and cancelling out dt, we get

$$\frac{d_y}{y} = \beta\frac{dx}{x}$$

Making the strong assumption that β is constant, we can integrate the above equation to obtain

$$\log\, y = \log\, \alpha + \beta \log\, x \qquad \text{or} \qquad y = \alpha x^{\beta}$$

Additive allometric model

Our personal experience is that it has a sense to consider separately both three parametric model

$$y = \gamma + \alpha x^{\beta}$$

and two parametric model

$$y = \alpha x^{\beta} + \varepsilon \tag{2}$$

Multiplicative (time-power) allometric model

- Has (in general) the form

$$y = \alpha x^{\beta} \varepsilon. \tag{3}$$

- Using logarithmic transformation we get

$$\log\ y = \log\ \alpha + \beta \log\ x + \log\ \varepsilon \tag{4}$$

respectively

$$\log\ y = \alpha^{\star} + \beta \log\ x + \varepsilon^{\star} \tag{5}$$

where $\alpha^{\star} = \log \alpha$ and $\varepsilon^{\star} = \log\ \varepsilon$

Exponential models

- Exponential model has (in general) the form

$$y = e^{\alpha+\beta x+\varepsilon} \tag{6}$$

Using logarithmic transformation we get linear model of the form

$$\log y = \alpha + \beta x + \varepsilon \tag{7}$$

- Exponential model with additive error has the form

$$y = e^{\alpha+\beta x} + \varepsilon. \tag{8}$$

Notice different placements of the error term, which implies different interpretation of the error term.

This model can be complemented also by a shift parameter, where needed.

Quadratic model

- Quadratic model has (in general) the form

$$y = a + bx + cx^2 + \varepsilon. \tag{9}$$

which can be also obtained as a special case of the additive allometric model (8).

Autoregressive model

$$y_{n+1} = \varrho y_n + \varepsilon_n, \quad \varrho > 1. \tag{10}$$

Parameter ϱ was in each step estimated using

$$\widehat{\varrho}_n = \sum_{i=2}^{n} y_i y_{i-1} \Big/ \sum_{i=2}^{n} y_{i-1}^2.$$

In general, this model again describes exponential growth and for $\varrho = 2$ it corresponds to the model of splitting. **Remarks:**

- Autoregressive model can be considered as a complement of the exponential model because it shows the speed of splitting, (partitioning, halving, division,...).
- For our data model (10) offered practically the same predictions as exponential model (6).
- In the literature is often suggested also Gompertz' model. In the initialization phase it practically coincides with the exponential model.

Assumption concerning errors

We assumed following cased

a. Errors are independent
b. Errors are dependent and form autoregressive sequence $AR(1, \rho)$

Assumption concerning errors

We assumed following cased

- a Errors are independent
- b Errors are dependent and form autoregressive sequence $AR(1, \rho)$
- c For construction of the prediction errors and/or estimation of the probability of exceedance over given threshold we assumed independence and normality of the errors

Remark: Comparison of the results when assuming independent errors, respectively dependent errors, confirmed our expectation that dependence does not play important role due to the short length of considered series.

General modes of behavior

We assume two general types of the behavior of observed series:

a "Slow" (gradual), corresponding in medical applications to the local relapse

b "Fast", corresponding corresponding in medical applications to the generalization of the disease

Decision between "slow" and "fast" type of the behavior depends on past experience of the user and all available prior information about the problem

GENERAL REMARKS

A For the case of "slow" (gradual) change additive allometric models seems to be most suitable

B For the case of "fast" change exponential model seems to be more suitable

C Generally, while prediction based on additive allometric model can be considered **conservative**, prediction based on exponential model rather **liberal** !

General remarks

- **A** For the case of “Slow” (gradual) change additive allometric models seems to be most suitable
- **B** For the case of “Fast” change exponential model seems to be more suitable
- **C** Generally, while prediction based on additive allometric model can be considered **conservative**, prediction based on exponential model rather **liberal** !

Representatives of both these models can be used for a prediction of the next value. However, this does not mean NEITHER the true value NOR next observation MUST be between these two predictions.

Moreover, one must always keep in mind that prediction more than one step ahead can be both non precise and meaningless .

Estimation of the exceedance probability

Aside prediction of the next value we were also interested in the estimation of the given exceedance probability, say 0.2 in my example. In the case of exponential model (6) we have

$$P\big(y_{n+1} > 0.2\big) \approx 1 - \Phi\Bigg(\frac{\log 0.2 - \widehat{\alpha} - \widehat{\beta} \cdot (n+1)}{\sigma\sqrt{1 + v_E^2}}\Bigg), \tag{11}$$

where we set $x_i = 1, 2, \ldots, n,$, $v_E^2 = 2(2n+1)/n(n-1)$ and $\Phi(x)$ denotes cdf of N(0,1).

Estimation of the exceedance probability

Analogously for multiplicative allometric model (3) we get

$$P\big(y_{n+1} > 0.2\big) \approx 1 - \Phi\Bigg(\frac{\log 0.2 - \widehat{\alpha^\star} - \widehat{\beta} \cdot \log(n+1)}{\sigma\sqrt{1 + v_{MA}^2}}\Bigg), \qquad (12)$$

where

$$v_{MA}^2 = \frac{n\sum_{i=1}^n (\log\, i)^2 - 2n\log(n+1)\sum_{i=1}^n \log\, i + n^2\big(\log(n+1)\big)^2}{n^2\sum_{i=1}^n (\log\, i)^2 - n(\sum_{i=1}^n \log\, i)^2},$$

where we set $x_i = 1, 2, \ldots, n,$, $\Phi(x)$ denotes cdf of $\mathrm{N}(0{,}1)$ and $\widehat{\alpha}$ and $\widehat{\beta}$ are estimates of the regression line fitted to the points $(\log\, x_i, \log\, y_i)$, $i = 1, \ldots, n$. Analogously as above we can estimate σ using historical data (experience).

Recommendations

pacient	age	obs.	model	$\widehat{y}_{n+1}$	$\widehat{p}_{0.2}$	model	$\widehat{y}_{n+1}$	$\widehat{p}_{0.2}$
289	74	19	aloA2	0.341	1.000	—	—	—
408	67	6	exp	0.248	0.790	aloA3	0.199	0.499
422	71	8	exp	0.536	1.000	aloA2	0.528	1.000
441	57	7	exp	0.315	0.961	aloA3	0.246	0.963
485	64	9	aloA3	0.062	0.000	—	—	—
496	62	7	exp	0.096	0.002	aloA3	0.072	0.000
522	63	7	exp	0.323	0.969	aloA2	0.237	0.894
533	57	3	???	—	—	málo dat	—	—
545	72	8	aloA3	0.261	0.996	—	—	—
570	69	4	exp	0.632	1.000	—		

Table 1. Models suitable for different patients.