New Applications of Approximation Theory to Forecasting of Global Systems Dynamics

Ognyan Kounchev
Institute of Mathematics and Informatics, Bulgarian Academy of Science

October 30, 2006

Abstract

We introduce some new methods of Approximation theory to forecasting and we outline some possible application to scenario management. Research partially supported by DFG project at the University of Bonn.

1 Global Models for Forecasting

Let us first outline the setting to which we will apply forecasting methods of Approximation theory. We describe in general terms the system of human-societal dynamics.

1. We choose some factors which characterize the systems and which may be extracted from the statistical yearbooks or to be estimated by some experts, say by agencies as **Moody's**; for some indicators it is possible to measure by expert opinions their relative change but not their value in an internationally recognized scale:

   (a) *GNP* (Gross national product)

   (b) *Age of the population*

   (c) *Price of labor*

   (d) *Quality of education* (percentage of high-school diploma, university diploma, etc.) with respect to some international scale; changes in its quality in EU, Eastern Europe, US, etc.

   (e) *Foreign investments* - amount and openness for investments; safety of investments (say, China is opened, Iran is not)

   (f) *Speed of introduction of inovations* in industry (say $EU : USA = 1 : 3$)

   (g) Private sector and state sector – percentage in economics

   (h) *Taxes*

(i) *Degree of corruption* of the administration, government and the society in general – important for the reliability of the investments, etc.

2. Second, we create some clustering of the global sphere into geographic and political-economic areas, for which one may collect and aggregate the above indicators, say:

   (a) EU+Bulgaria+Romania

   (b) the rest of the Balkans

   (c) Turkey, Middle East, Persian Gulf

   (d) Ukraine, Russia, and former Soviet Union countries

   (e) China

   (f) Japan, S. Korea, SEAsia

   (g) India

   (h) N. America

   (i) C. America

   (j) S. America

   (k) Australia+New Zealand

   (l) Africa

3. Some balance equations (preservation principles) have to be satisfied, say:

   (a) The whole amount of resources is limited, i.e. we have for every strategic resource the following restriction (oil, gas, water, metals: aluminium, uranium, etc.)

   $$\sum_{j=1}^{N} R_{j,i} \leq R_i$$

   where $R_{j,i}$ is the amount of resource used by the jth "player" in the ith year; here R_i is the whole amount of the resource available in the world for the ith year.

   (b) The markets are also a special kind of "resource" which has some saturation levels and satisfies similar equations as above.

   (c) An analog to the Principle of "*non-arbitrageness*" from Mathematical Finance, i.e. "*there is no empty place in the political space*" or "*there is no free lunch in politics*" is related to the above balances.

For every indicator and for every "player" j we have a time series $J_{j,i}$ which is the measured value at the end of the time period $[t_{i-1}, t_i]$ or a kind of mean value during that period will be taken. It is possible that the data have some **gaps** or they have not been collected for a long time back – short **history**.

We will use splines for modeling the data, for filling the gaps in the data, for the approximation and the forecast of the data – especially forecasting the trends.

The idea is to improve *the rule of thumb* (corresponds to the *linear prediction*) which is used by the practitioners in all spheres of prediction.

We will find **spline-autoregression** approximations S_j which approximate the above data, namely

$$S_{j,i} \approx J_{j,i}$$

providing forecasts of the above time series. A basic point is that we will require that these approximations satisfy at every time interval the conditions of the balance, i.e. if for every time period i we have

$$\sum_{k \in \{area\}} J_{k,i} \leq J_i$$

then we will require also

$$\sum_{k \in \{area\}} S_{k,i} \leq J_i .$$

Another major idea is that *smoothing splines* and the other techniques which we develop are able to *extract the tendencies* in the data and *to remove random elements* in them.

2 What are splines and why splines?

Definition and examples of splines:

Let in the interval $[a, b]$ we have

$$a = t_1 < t_2 < ... < t_N = b$$

1. - linear splines
   - quadratic and cubic splines
   - they are solutions of

$$\begin{aligned} \text{linear}: \ & x'(t) = C_i \\ & \text{for } t \in (t_i, t_{i+1}) \\ \text{quadrat.}: \ & x''(t) = C_i \\ & \text{for } t \in (t_i, t_{i+1}) \\ \text{cubic}: \ & x'''(t) = C_i \\ & \text{for } t \in (t_i, t_{i+1}) \end{aligned}$$

   with *interpolation* (or other type of restrictions):

$$x(t_i) \approx d_i \qquad \text{for } i = 1, 2, ..., N.$$

- More general: solutions to
$$x^{(n)}(t) = F(t)$$
where F has jumps (in part. piecewise constant)
- Problem: For given N data c_j measured at the points t_j find a spline $u(t)$ such that
$$u(t_j) = c_j \qquad \text{for } j = 1, 2, ..., N$$
or
$$u(t_j) \approx c_j \qquad \text{for } j = 1, 2, ..., N.$$
- **Predictions** with splines

2. Why splines? – Splines are a lot more flexible than polynomials. Polynomial splines, exponential splines (used by Vasicek), etc.

   Good for prediction since they "follow the shape" of the data.

3 Discrete time models – ARMA and GARCH

Autoregression, *Moving Average*, and *GARCH*.

1. Autoregression:
$$x_{t+1} = \sum_{i=0}^{p} a_i x_{t-i} + f_t + \varepsilon_t$$
   where f_i is a given vector, and ε_i is random variable, e.g. Gaussian white noise. We generate a sequence $\{x_i\}_{i=0}^{N}$ where $N > p$. Usually the *free term* is taken as $f_i = 0$.

2. Let $\{y_i\}$ be a given data vector. The problem: **Find** the **coefficients** $\{a_i\}$ so that
$$\sum_{i=0}^{N} (y_i - x_i)^2 \longrightarrow \min_{a_i}$$
   Then the **prediction** is
$$x_{N+1} = \sum_{i=0}^{p} a_i x_{N-i}$$

3. Simplest case – linear autoregression: $N = 0$
$$x_{t+1} = a_0 x_t + f_t + \varepsilon_t$$

4. A number of techniques exist for computing AR coefficients. The main two categories are **least squares** and **Burg method**; cf. the monograph of **Hamilton** "Time series analysis", 1995.

5. Main problem: Find the *coefficient-sequence* $\{g_i\}$ such that the solution $\{x_i\}$ **approximates** as nice as possible some data sequence $\{y_i\}$, e.g. in least squares sense.

6. *GARCH* has been created for applications in Finance by the *Nobel prize winner R. Engle* (2003) and by his pupil Bolerslev in the early eighties.

4 Splinified autoregression methods

Recently there has been a growing interest in the application of splines to ARMA and GARCH methods, some has been published by R. Engle.

Independently, we have developed a "splinified" version of the AR method. For simplicity we will restrict ourselves to $AR(1)$.

Definition 1 ***Definition of splinified autoregression of first order*** $sAR(1)$: *We use a subdivision of the discrete interval*

$$1, 2, 3, ..., N$$

into M *subintervals, i.e.* M *integers* d_j *with*

$$1 = d_0 < d_1 < d_2 < ... < d_M = N.$$

We consider the following autoregression equations which define the splinified $sAR(1)$ *method:*

$$\begin{aligned} &x_{t+1} = a_0^{(j)} x_t + \varepsilon_t^{(j)} \\ &\text{for } d_j \leq t \leq d_{j+1} \qquad \text{and } j = 0, 2, ..., M-1 \end{aligned}$$

so the regression changes at the points d_j

For computing the $sAR(1)$ we proceed as follows:

Step 1. For finding the coefficients

$$\left\{ a_0^{(j)} \right\}_{j=1}^{M}$$

we extremize with respect to a **maximum likelihood** criterion. Then the forecast x_{N+1} is given by

$$x_{N+1} = a_0^{(M)} x_N.$$

Step 2. We choose a window and we optimize the procedure with respect to the choice of the subdivision points $\{d_j\}$ for getting a better **forecasting performance**.

5 Application of forecasting of time series to Global Systems Dynamics

In Section **??** we have described the Global system by considering a set of indicators which are described by the time-series $\{J_{k,i}\}_i$ with the time variable i and for every sub-system k. In managing the possible scenarios which might appear we proceed as follows:

Step 1. Assume that the present time is T. For some fixed time T_1 in the future we consider forecasts $\{S_{k,j}\}_{i=T}^{T_1}$ of these time-series $\{J_{k,i}\}_i$ for each k. Note that these forecasts take into account all possible restrictions of the resources available.

Step 2. We apply the methods of analysis of scenarios to the predicted series $\{S_{k,j}\}_{i=T}^{T_1}$ and check for possible bifurcation points in the dynamical system.