Interdisciplinary Description of Complex Systems 2(1), 88-94, 2004

COMPLEXITY AND UNCERTAINTY IN THE FORECASTING OF COMPLEX SOCIAL SYSTEMS

Péter Alács*

Futures Studies Centre, Budapest University of Economic Sciences
and Public Administration
Budapest, Hungary

SUMMARY

The better the model, the more features of the problem it explains. However, showing that the model has similarities to that of a phenomena is often less significant in applications due to lack of data. Forecasting, as special application of modelling, is neither an exception: besides statistical data one should use several types of subjective assumptions about the present and the future state of the model. In case of complex models, this fact is extremely important, because these models use often unobservable, hidden or – regarding its future evolution – uncertain variables. We developed a simple mathematical approach how these uncertainties can be managed in the model. We shall also show how these uncertainties can influence the behaviour of modelled variables, and how an approximate for time horizon of forecasts can be calculated.

KEY WORDS

complex systems, futures studies, foresight, modelling, time horizon

CLASSIFICATION

ACM Categories and descriptors: J.4 [Computer Applications]; Social and behavioral sciences,

APA: 3040, 4010,

JEL: A0,

PACS: 87.23.Ge, 89.65.-s

*Corresponding author, η: peter.alacs@bkae.hu; +36 1216 2016;
Budapest University of Economic Sciences and Public Administration, Fővám tér 8.,
H-1093 Budapest, Hungary

INTRODUCTION

It must have been an early discovery that understanding social phenomena needs complex modeling approach. Modeling difficulties, especially in the case of formal modeling, however, seemed to be much harder to surmount as explaining the results of the partial, noncomplex models. E.g. in economics, utility functions are an easy and capable tool for explaining several types of economic behavior, but formal model builders using this tool had never even been aimed to explain why and how these utility functions were chosen. Of course, all of them know very well, that the tool they use is the contact point to many other social behavior and phenomena. But, believing in traditional scientific separation of disciplines, these contact points are regarded as curtains indicating the border, the things behind which they simply do not want to care. Complex model approach concentrates on the targeted phenomena, uses interdisciplinary components if necessary. But as complex model approach covers a wider range of processes, it also demands a wider range of data, which is often inaccessible. Sometimes inaccessible data is only because of the fact that statistical offices supply only traditional models with aggregate macro data that is not detailed in the demanded complex way. Further research and experiments may access this data. But these research and experiments can also be very expensive or probably unfeasible: the data remains inaccessible, fitting the model needs further assumptions.

UNCERTAINTY

In the following, we shall show on a very simple example how uncertainty can be managed within a very simple model. Let us consider a certain social quantity x that can be observed continuously. We are certain about the growth rate of this quantity α, that should be considered to be constant even on the long run. The evolution of this quantity is given by

$$x(t) = x_0 \mathrm{e}^{\alpha t}. \tag{1}$$

where x_0 denotes the initial value of the variable $x(t)$ at time 0. We are aware that most social processes are observed with more complex dynamics and structure, but here in this simple example we chose also a simple dynamics. In fact, here parameter α plays only a symbolic role, that shows the models self-deterministicity, as $1/\alpha$ denotes how strong the model depends on its own parameters: e.g. if α is negative, that large α values mean short-time memory of the process.

Regarding the small increments of x we write the previous form of the evolution of the process as:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \alpha x(t)\,. \tag{2}$$

We assume, that all type of uncertainties are due to an additive term to this expression:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \alpha x(t) + f(t)\,. \tag{3}$$

where $f(t)$ denotes some type of "disturbing" processes, that are not managed by the model. Of course as $f(t)$ is an external variable, we do not know much about its behaviour, and we can have basically two types of assumptions. In the practice of data-analysis some probabilistic behavior is often assumed for uncertainty. The simplest example for this type is the white noise, which means in words that the uncertain variable $f(t)$ is independent from the process x, but also from its previous values (uncorrelated), and the "size" of the uncertainty (variance of $f(t)$) does not change in time. Note that, as above in case of parameter α, the choice of white noise is also an illustration of the stochastic variables that may appear in the model. Another way to manage uncertainty in the model is to choose $f(t)$ from a set of functions without any probabilistic assumptions. In this case (3) must be solved for all element of this set, and the future value, in case this type of uncertainty is the only one present, is the set of future value of all possible evolution of the process. The simplest example for this type of uncertainty is to assume $f(t)$ to be small: $|f(t)| < \varepsilon$ for all t. Note that modeling using interval dynamics e.g. multifunctions is not unknown for economists (Debreu became a Nobel-prize for his research in this field [3]). Now, we solve for all "small" (measurable) functions (3). The uncertainty under these assumptions can be measured by the diameter of the interval of the future value of x:

$$E_{\text{Det.}}(x(t)) = \max_{|f|<\varepsilon} x_{f(t)} - \min_{|f|<\varepsilon} x_{f(t)}, \tag{4}$$

where x_f denotes the solution of (3) for the measurable function f.

Fortunately, we have an analytical solution of (3) for all (measurable) functions:

$$x_{f(t)} = x(t) = \int_0^t \mathrm{e}^{\alpha(t-s)} f(s)\mathrm{d}s, \tag{5}$$

In particular, if f contains stochastic and a deterministic part (a white noise and a "small" function), the solution can be written in the following form:

$$x(t) = \sigma \int \mathrm{d}W(s) \cdot \mathrm{e}^{\alpha(t-s)} + \int \mathrm{d}s \cdot f(s)\mathrm{e}^{\alpha(t-s)}, \tag{6}$$

where $\mathrm{d}W$ denotes the white noise, σ is the measure of the stochastic uncertainty and f is a "small" function. We see, that in our special simple case, the two types of uncertainties (the stochastic and the deterministic one, i.e. the first and the second part of the solution in (6) can be calculated independently. Note that this is now a very favorable case to consider, because in general even if analytical solutions are supplied, it is not ensured that the evolution of the system (e.g. in case of chaotic dynamics) will result contiguous sets for the deterministic disturbance.

The interpretation of equation (6) can be approached from either the stochastic or the deterministic side. On the one hand, let us fix the deterministic disturbance function f. Now we get a stochastic process, and the future values of variable x are stochastic variables that can be managed by the usual probabilistic methods. If another possible disturbance function is chosen, we get another stochastic process. It is clear that one possible way to interpret the solution is a set of stochastic processes, and according to this the interpretation of the future variable x is a set of random variables. On the other hand, let us realize the white noise, and calculate all the results of the deterministic disturbance function f. This yields a set function where for each t the future value of x is a contiguous set (interval). If another realization is taken, we get another set function and intervals for the future value. In this sense the future value of x is a "random interval variable", which gives intervals as a result when realized.

Let us analyse the solution (6) by calculating for any given time t the measures of uncertainty for the two terms. The measure of uncertainty of the first, stochastic term is the variance:

$$\mathrm{Var}\left[\int_0^t \mathrm{d}W(s)\mathrm{e}^{\alpha(t-s)}\right] = \int \mathrm{d}s\left[\mathrm{e}^{\alpha(t-s)}\right]^2 = \frac{\mathrm{e}^{2\alpha t}-1}{2\alpha}.$$

The measure of the "deterministic uncertainty" is the measure (i.e. length) of the interval, which can be calculated

$$\mu_L\left\{\int \mathrm{d}s f(s)\mathrm{e}^{\alpha(t-s)} : f \in F_\varepsilon\right\} = \varepsilon\frac{\mathrm{e}^{\alpha t}-1}{\alpha}.$$

where μ_L denotes the Lebesgue-measure and F_ε the set of real functions that maps to $[-\varepsilon, \varepsilon]$.

Comparing the measures of the two different type of uncertainty we see, that for small time-scales ($t << 1/\alpha$), the stochastic one behaves as the square root function, while the deterministic one is linear. For large time-scales ($t >> 1/\alpha$) however both expressions behave as $\mathrm{e}^{\alpha t}$. This means that for small time-scales always the stochastic type of uncertainty is dominant. For large time scales the two type of uncertainty follows the same law, however, it depends on the parameters which is the larger, Figs 1 and 2.

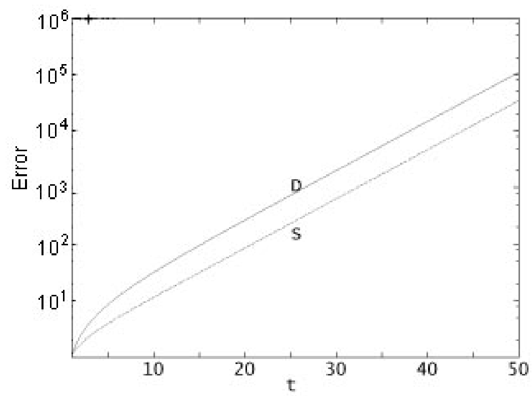

Figure 1. The two types of uncertainty follow the same rule on the long-run (illustration).

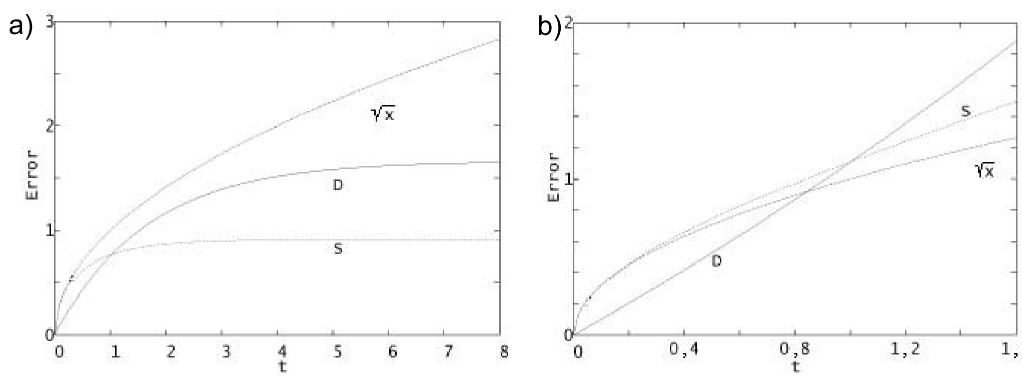

Figure 2. The two types of uncertainty on short-run. Parameter α is a) negative, b) positive.

Of course, the most important question is, what is the interpretation of a future variable with mixed types of uncertainty.

Nowadays it is usually accepted to have stochastic uncertainty in a model. It is also accepted, that different future scenarios are worked out to manage easier the future complex behavior of the system. It is however not clear how to interpret future scenarios with stochastic variables (see table 1). Now we are facing a similar problem in our model.

Table 1. The impact of different types of uncertainty.

Type of uncertainty	None	Stochastic
none	deterministic value	Statistics
deterministic	scenario	?

If we have small stochastic uncertainty than we can use an interpretation similar to the scenarios. If we have small deterministic (external) impact, we can use stochastic models (even on long run, as proven above). The problem arises when the two type of uncertainty have about the same impact. Our opinion is, that it is impossible to put a question about the future variable in this case. The main problem with this kind of mixed uncertainty is that the stochastic noise is large enough to make the "deterministic noise" (intervals) overlap, but not large enough to disperse them disjunctly. Therefore both approaches to interpret the future value fail. In fact, this is not the case, in which we do not know anything about the system. But it is the case, in which the two types of uncertainty together results in a much higher level of uncertainty compared to the accepted level. As we see in our simple model, stochastic uncertainty is always dominant for short time-scales. It depends on the parameters of the model, whether or not the deterministic uncertainty takes over for long time-scales. If so, then there exists a point where the usual probabilistic methods fail due deterministic perturbance. We define this point as the point where the measures of the two types of uncertainty are the same. According to the above, we regard this point as a goodness of the forecast: the larger this value is, the more stability our model has, against not modeled (e.g. external) perturbance. We call this point time horizon of the forecast, which is not to be regarded as a cutting point of our forecast to throw anything away behind this point, but as a characteristic value for our model, where increasing the value the model improvement should keep an eye on. Time horizon approximations are recently developed on complex models of demographic processes [1, 2].

Now, we calculate the time horizon of our simple model. Firstly we determine the condition of the existence of the time horizon. As mentioned previously, the deterministic noise can be small enough to be smaller for the whole evolution of the model. In this case there is only one solution of the equation

$$\varepsilon \frac{e^{\alpha t} - 1}{\alpha} = \sigma \sqrt{\frac{e^{2\alpha t} - 1}{2\alpha}}, \tag{7}$$

namely $t = 0$. So, the condition is

$$\left| \frac{\varepsilon}{\sigma} \right| \leq \sqrt{\frac{\alpha}{2}}.$$

If the deterministic noise (i.e. ε) is small, stochastic models are appropriate also for the long time-scales. If this condition does not hold, the time horizon should be considered, so the (only) positive solution of (7) should be calculated:

$$\left| \frac{\varepsilon}{\sigma} \right| = \sqrt{\frac{\alpha}{2} \cdot \frac{e^{\alpha t} + 1}{e^{\alpha t} - 1}}.$$

Thus, how far we can predict with our simple model depends on how strong the deterministic noise is present in the model (compared to the stochastic noise), ε/σ and on how strong the model is determined through its inner dynamics, α.

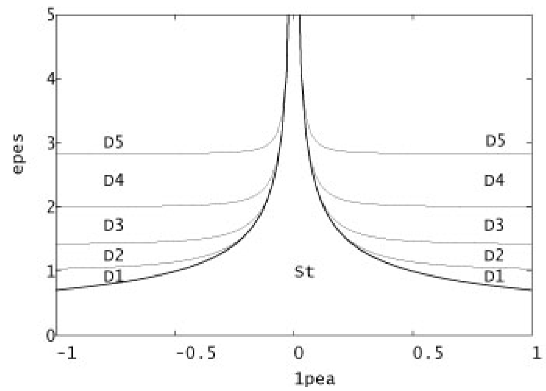

Figure 3. Regions for different time horizons (D_1, …,D_5), and the stochastic stable region (Stoch.).

In figure 3 we can see several regions of the model parameters defined by the different time horizons (not that in the figure not α but $1/\alpha$ is plotted for a better overview). It is clear that in case of strong internal dynamics of the model, the deterministic noise has little impact on the process, but for each finite α a certain threshold can be given above which the influence is significant and time horizon should be considered. The regions of D_1, …, D_5 are separated by the iso-time-horizon lines for time horizons of 0,125, 0,25, 0,5 and 1. It is interesting to see, that the real "model risk" for forecast is the parameter α, the strength of internal determination of the model. Two similar models with little difference in internal determination can have larger difference in time horizons than two similar models with little difference in deterministic noise. If the parameter $1/\alpha$ is negative, it is also to be interpreted as the memory of the system. If the system has large memory stochastic stability can be disrupted by small deterministic noise. In case of large memory, the time horizon is mainly the function of the deterministic noise. The real model risk in this case is similar as above: at a certain memory, a little change in deterministic noise can have large impact on the time horizon. In words, we would express this property as above this threshold the model would be depend stronger on other non-modeled (e.g. external) processes rather that its own dynamics.

CONCLUSIONS

The time horizon is defined for forecasting models of social phenomena as a test of the dependency of non-modeled or external processes. As almost all social phenomena is embedded strongly in a complex background, non-modeled processes should be quite often considered. Often these processes are not included in the model, because it is hard or impossible to obtain information about them. Stochastic models are used quite often especially in economics. These models proved to be unstable if other non-modeled processes (e.g. politics) can have large influence on the process. The question whether the impact of these processes from the complex background should be considered as large or small depends also on the internal dynamics of the system. Systems with small memory or characteristic internal dynamics are less sensitive to these effects.

ACKNOWLEDGMENT

This paper is based on the results of a research program which is on at the *Complex Futures Studies Research Group of HAS-BUESPA*.

REFERENCES

[1] Alács, P.: *Evolúciós Demográfiai Előrejelzés Evolúciós Modellel*. Hideg, É., ed.: Evolúciós Modellek a Jövőkutatásban, 2001,

[2] Alács, P.: *Evolutionary Modeling of Demographic Processes*. Hideg, É.; Martinás, K.; Moreau, M. and Meyer, D., eds.: *Complex Systems in Natural and Social Sciences*, 2003,

[3] Kováts, G.; Muresan, M. and Vizivári, B.: *Viability Results in Control of One-Dimensional Discrete Time Dynamical Systems Defined by a Multi-Function*. Hideg, É.; Martinás, K.; Moreau, M. and Meyer, D., eds.: *Complex Systems in Natural and Social Sciences*, 2003.

SLOŽENOST I NESIGURNOST U PREDVIĐANJU SLOŽENIH SUSTAVA

Péter Alács

Centar za proučavanje budućnosti, Sveučilište za ekonomske znanosti i javnu upravu
Budimpešta, Madžarska

SAŽETAK

Što je bolji model, više svojstava objašnjava. Međutim, dokazivanje sličnosti modela i pojava obično je manje važnosti zbog nedostatka podataka. Predviđanje, kao posebna primjena modeliranja, pritom nije izuzetak: osim statističkih podataka potrebno je primijeniti nekoliko vrsta subjektivnih pretpostavki o sadašnjem i budućem stanju modela. U slučaju složenih sustava ova je činjenica posebno značajna, jer njihovi modeli upotrebljavaju obično nemjerljive, skrivene, ili – obzirom na njihovu buduću evoluciju – nesigurne varijable. Postavljen je jednostavni matematički pristup baratanju s nesigurnostima u modelu. Dodatno je pokazano kako nesigurnosti mogu utjecati na ponašanje modeliranih varijabli i kako se određuje približni iznos horizonta vremena predviđanja.

KLJUČNE RIJEČI

složeni sustavi, proučavanje budućnosti, predviđanje, modeliranje, horizont vremena