Business Process and Performance Indicator Simulation for Business Intelligence

Veit Köppen

Freie Universität Berlin
Institut für Produktion, Wirtschaftsinformatik und OR
Garystr. 21, 14195 Berlin
koeppen@wiwiss.fu-berlin.de

Summary. Companies are nowadays controlled by a mass of business figures. These business figures build a system and are typically dependent upon restrictions that arise from balance equations. The introduction of uncertainty within these figures dismisses standard mathematical formulations. Uncertainty is imposed by forecasting or estimation, but also by evaluation of stochastic process. This article gives an brief overview of discrete event simulation on the one hand and stochastic business figure controlling on the other hand. Hereby two different simulation methods are presented that are both essential for Business Intelligence to control a business efficiently.

1 Simulation

Simulation is an established scientific method. If the analyzed part of the real world is too complex or experiments are not possible, simulation can be used. The Cambridge dictionary has the following definition for the verb simulate:

"to do or make something which looks real but is not real".

In sciences simulation is extensively used. It "describes a wealth of varied useful techniques, all connected with the mimicking of the rules of a model of some kind."(MORG84, p. 1) The techniques implemented on a computer are frequently exhibit satisfactory success. (HART96) The reality is simplified and described in a model. A model abstracts from reality by focussing on the parts of interest. However, in order to obtain valuable results it is important that the model is close to the analyzed part of the reality. A model is a simplification of a real world system or process. It is adopted for explanation of the workings and often relies on a mathematical representation. The process of model generation is called abstraction because unnecessary information is cut

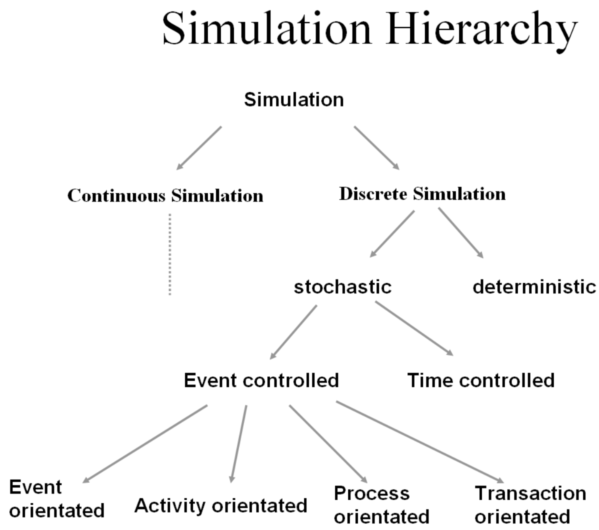

Fig. 1. A Hierarchy of Simulation

out. A model should also be validated, to ensure that all necessary information is included in the model.

Fig. 1 provides a possible classification of simulation. A first differentiation is done by separating the state space. In this level a partitioning into continuous and discrete elements of the simulation is done. In both cases a further partitioning into stochastic and deterministic is taken in the next level of classification. But in Fig. 1 only the discrete branch is visualized. Deterministic means that every state in a simulation step can be predicted. When some states are achieved at random the simulation becomes stochastic. On the next level a differentiation of the simulation trigger is done. On the one hand a simulation can be seen as time driven, where a change of state is performed only dependent on time. On the other hand, events can be responsible for state transitions. An additional level of classification is viewed with respect to the focus. Here typical representatives are the events that are evaluated with the simulation or the activities that have the main focus of the simulation study. On a more aggregated view the activities are put together into processes and these are of the main focus. Another possible target of a simulation study can be transactions. In Fig. 1 only the discrete branch is visualized for reasons of clarity. But the same branch can be copied to the continuous node.

The simulation itself is a process that can be divided into preparation phase, execution phase and analysis. Fig. 2 shortly describes the phases.

At the beginning a problem has to be formulated. This should be done as precise as possible to guarantee a solution. In the next step the problem is analyzed and a abstraction is done to focus on the main aspects. This is called model creation and can be divided into model design and model formu-

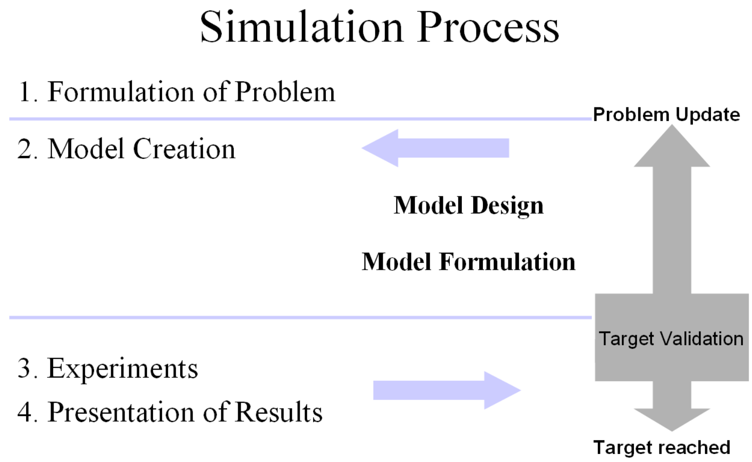

Fig. 2. A Process Model for Simulation

lation. In the model design the elements of simulation are chosen. The level of abstraction from reality is also decided at this point. The model formulation follows to this phase. Here natural language descriptions and requirements have to be transferred into formalizations. A validation of the model can also be performed. At the end a simulation model with all necessary information is created. The preparation phase comprise all of these procedures.

The execution phase uses the simulation model. A program called simulator is the implementation of this model. One simulation run is called experiment. How often a simulation is performed is dependent on the simulator and its parameters.

The next step is the presentation of the results. This can be done by using graphics or presenting the simulated values. Then a decision has to be made if the problem is indeed solved and the targets are fulfilled. If the simulation results are not satisfactory, an update of the problem should be initiated. A new model has to be created and the simulation procedure is performed again.

Physical simulation is used when real objects are substituted by a cheaper or smaller object. Interactive simulations are aligned to the human behavior. The human in the simulation loop activate human operations. Computer simulations have been used since the early 1940s. It is an attempt to model a real situation so that it can be studied. The behavior of the investigated system is analyzed with changes in the variables and usage of predictions. A set of initial parameters in a given environment determines the start of a simulation run. Evaluation and prediction are the targets with a simulation study. The simulation model specifies the relationships between objects within the system.

Simulation as a method and tool in Business Intelligence can be manifold. The two most important simulations are in the area of key performance indicators and business process optimization. While in the case of key performance indicators, prognosis and fraud detection are important, business process optimization uses simulation in developing new processes or detect inefficiencies. In the following simulation in both areas is described.

2 Business Process Simulation

A continuous improvement of the business and its processes is critical for the business success. Technical improvements, permanent evolving market conditions and public environment make a continuous business optimization necessary. Performance measures are used to describe the business in an aggregated view. Most of the performance measures regarding business processes can be assigned to one of the following categories:

- costs,
- quality,
- time and
- service.

Three different application areas exist: test of non existing business processes, analysis and evaluation of existing processes and what-if analysis of existing business processes with changes in some variables.

For checking a simulation model prior to experimental analysis different phases are possible. The simulation model must be checked for building the right model and building the model right. This is performed with validation and verification. Validation means testing if the model behaves with satisfactory accuracy consistent with the objectives of interest. Verification checks the accuracy of transforming a problem into a model.

The initialization of model variables can be quite difficult and will deviate from the desired equilibrium in most cases. Fig. 3 shows how the initial value

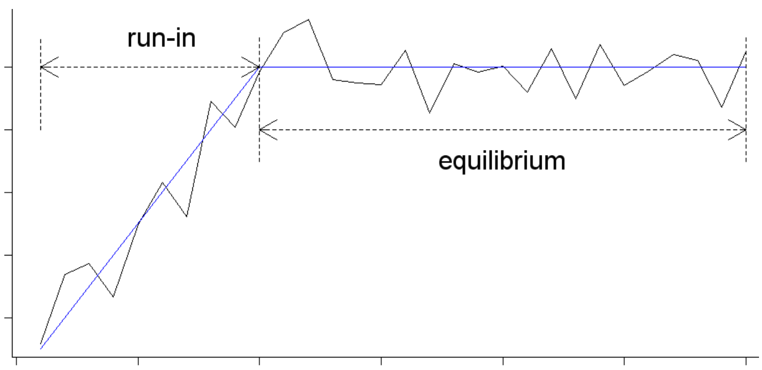

Fig. 3. Run-In and Equilibrium

influences the simulation run. An increase of the observed objective can be seen in the run-in phase. While reaching the plateau the simulation converges to its equilibrium.

Simulation—being one out of several methods for evaluation of business processes—exhibits the following advantages: less cost-intensive due to the fact that try and error are done in a virtual surrounding, computer simulation reduces the time of experiments, and communication of experimental results is easy. Cons for use of simulation are: cost of software, expensive generation and implementation of simulation models, data collection because expert estimation and real data are not always available. But the main problem is over-interpretation of the simulation results.

2.1 The Process of Business Process Simulation

Business Process simulation consists of eight steps (HLUP98). Although the steps are sequential, an iteration is necessary to obtain a suitable outcome.

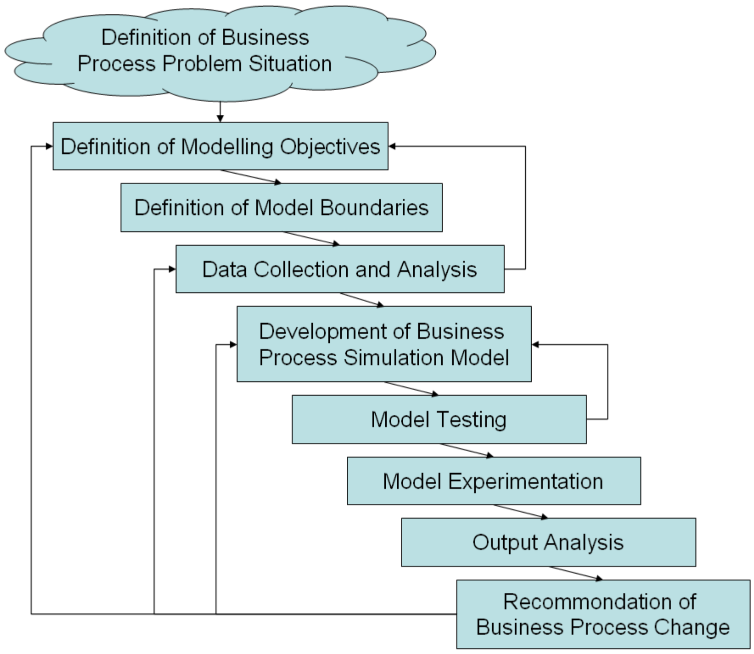

Fig. 4. Process of Business Process Simulation

Fig. 4 gives the complete view of the process of business process simulation. A simulation is implemented due to the fact that a problem occurs. This problem situation has to be defined and business process simulation can begin. In a first step, the objectives of the simulation study are defined. Corresponding to that, the boundaries are stated in a second step. Here the focus of the essential parts of the simulation study is done. Time and resource restrictions, the suitability of processes and importance of processes play an important role in this step. This is followed by data collection, which can be expert knowledge or real data. A huge amount of data is required to bring the simulation model close to reality. The constants of the model, initial values for variables and stochastic distributions have to be estimated or given by experts. In the case that not all necessary data is available, an adjustment in model objectives is needed and a new iteration begins. After a successful data collection, the development of the simulation model begins. A software package is use to create the simulator. The next step is model testing and evaluation of the developed model. Normally the development and testing are alternatively iterated. Starting with an initial easy model both steps refine and expand the simulation model up to satisfactory result. In the next step the experiments are performed. To minimize the random effects a plan is necessary. The results of the simulation runs should be statistical sound without restrictions. In the output analysis standardized statistical methods are used to interpret the simulation results. Recommendations can be given if the evaluated results are satisfactory and process changes are reasonable.

2.2 Discrete Event Simulation

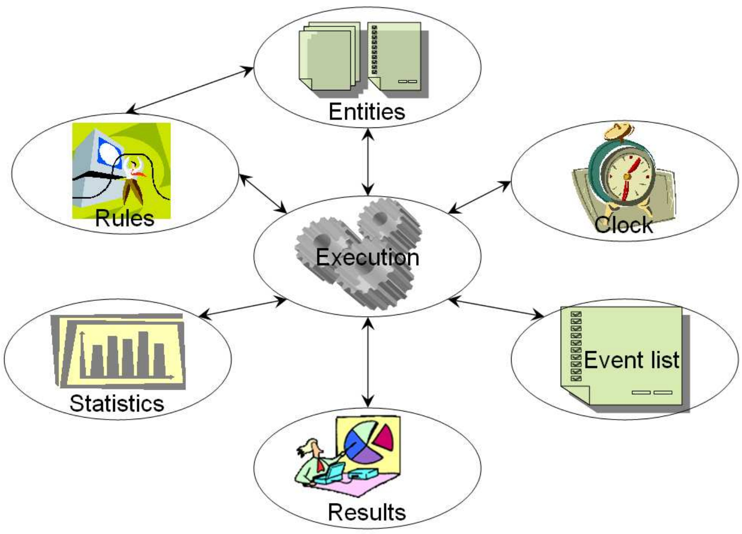

Fig. 5. Elements of a Discrete Event Simulation

The model in the Business Process Simulation is a discrete event model. The system is described in a chronological sequence of events. Each event marks a system change. According to (KREU86), a discrete event simulation model consists of two major concepts: entities and rules system or logical statements. Entities are tangible elements, that have a counterpart in reality. Entities can be further differentiated into temporary and permanent entities, durable and non-durable. The rule systems links the entities together. The logical relationships are the key element in the simulation model, because the rules define the behavior of the complete simulation systems. A huge quantity and variation of rules exist so that the easy rules sum up to a complex system. A rule is performed if the requirements are fulfilled. These requirements can be events that have occurred and entities that are necessary. In the simulation execution the controlling of events is done. An event list and a clock keep track of time and necessary rules are followed. The execution provides the dynamics and time based behavior of the complete simulation environment. Two other components complete the simulation environment. Both are necessary for all simulations not only for the discrete event simulation. The statistics component is responsible for the random number generation. Several different distributions describe typical real world behavior. The second component is the result presentation. Here results are displayed, important measures computed and meaningful analysis provided. Charts, histograms and tables are typical tools in this area.

3 Business Indicator Simulation

Controllers and decision makers base their activities on performance measures. These indicators are computed with an equation system to aggregate the operational data into a comprehensive format. The operational data might be noisy due to measurement errors or estimation. As a result of error propagation, all aggregated data have errors too. Different approaches exist in this case. Fuzzy set theory or Gaussian techniques are able to handle such indicator systems with restrictions,(KOEP06). However, dependencies between the variables are neglected. In this case, simulation with error in the variables is the only choice of handling an indicator system. In the following this is briefly explained.

The two components of an indicator system are indicators and relationships. These relationships can be described by an equation system. This is true due to the fact that the aggregated indicators are derived from lower level indicators. Fig.6 shows the DuPont System of Financial Control which is an easy indicator system. The only operators in this system are +, -, * and /. These are the typical operation in a business indicator system. Due to multiplication and division, nonlinearity occurs, which makes it difficult to handle such an equation system. A further model description can be found in (LENZ91).

Dependencies between variables are described with a multivariate distribution function. Due to the fact that sampling from a multivariate density

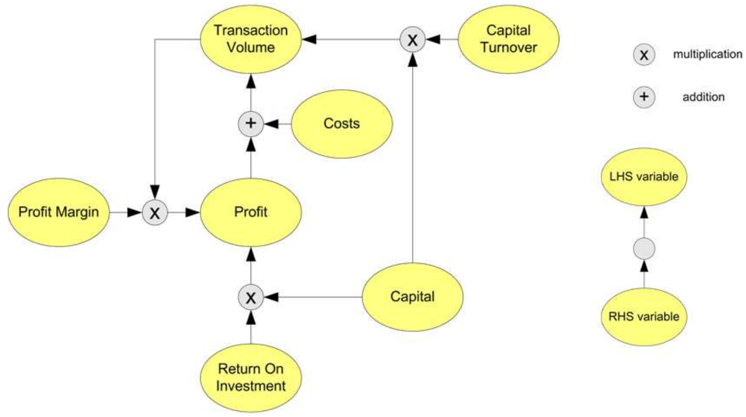

Fig. 6. DuPont System of Financial Control

function is not a trivial task, the Metropolis-Hastings algorithm (HAST70; GAME06) is the appropriate choice for sampling from these distributions.

Additionally, the equation system can be used to compute further estimations of the variables. In such a case, different estimations for a variable exist. These estimations have to be merged and a decision has to be made whether or not the estimates match.

The complete procedure to simulate a indicator system is given in the SamPro algorithm, see Alg. 1. The properties of this algorithm are estimating missing values of variables on the one hand, and improving the estimations of variables on the other hand. In this algorithm sampling from the given information on the indicators is done as a first step. A MCMC technique (the Metropolis Hastings algorithm) is used. With this technique samples can be drawn from any L1-norm function, because the normalising constant is factored out.

In the equation $z = x_1 + x_2$, z is the left hand side (LHS) variable and x are the right hand side (RHS) variables. With the help of the equation system all LHS variables can be computed in the algorithm in the second step. Note, that most operations in a business indicator system are separable and so a huge number of equations exists within a equation system.

When more than one estimate per variable exists a merge of information is done. Here two possibilities exist. One way is that the estimates are disparate. In this case the algorithm stops and a inconsistency message will be displayed. In the other case all estimates per variable build the joint distribution which is restricted on all estimates with the same value. Here a variance reduction and shift in mean according to the variance is achieved, see (MUEL07).

Fig. 7 shows the results of a simulation with the DuPont system. Here the Return On Investment (ROI) is displayed with given and simulation result data. A shift in mean and a variance reduction is evident.

Algorithm 1 SamPro algorithm for simulation of Business indicator systems

Require: equation system M, observation vector x and z
Ensure: estimates for all variables
1: resolve (set $LHS \equiv RHS$) for each variable in all equations
2: simulate samples for all RHS variables
3: compute LHS variables with the equation system M
4: estimate quantiles $\underline{q_{max}}$ and $\overline{q_{min}}$ for each variable where $\underline{q_{max}} = max\{\alpha$ quantiles per variable$\}$ and $\overline{q_{min}} = min\{(1-\alpha)$ quantiles per variables$\}$
5: **if** $[\underline{q_{max}}, \overline{q_{min}}] = \varnothing$ **then**
6: inconsistent system and algorithm stops
7: **else**
8: compute the joint distribution of $\hat{f}_{x,z}$ restricted by the subspace $x - z = 0$
9: compute estimates for each variable
10: **end if**

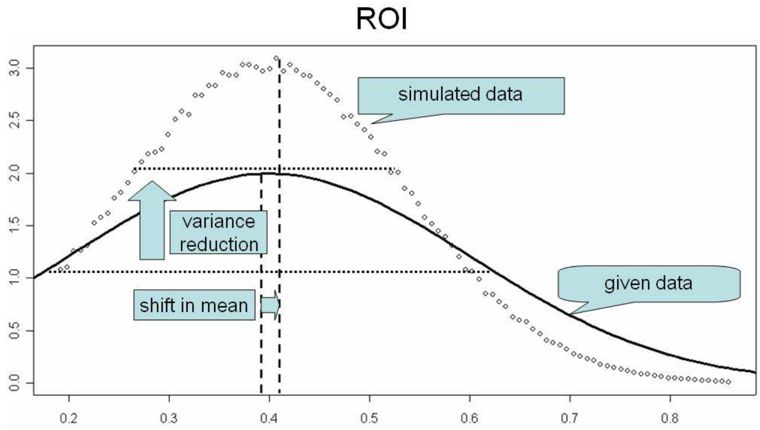

Fig. 7. SamPro result of indicator ROI

4 Conclusion

Simulation is an approach to analyze static and dynamic systems. In the area of Business Intelligence simulation is very powerful tool to obtain new insights for business analysis, controlling, optimization and decision support.

Mathematical or rules based systems are the fundamentals for business simulation. The generation of simulation samples can be done with Monte Carlo techniques. But in the case of multivariate data appropriate techniques are very complex. Markov Chain Monte Carlo techniques are in most of the cases the only choice of resources.

Building a discrete event simulation model can help us understand the business processes and find bottlenecks or unnecessary waste of time and profits.

Using business indicator simulation might reveal fraud. It improves the quality of stochastic indicators with shift in means and variance reduction as well.

References

[GAME06] Gamerman, D. and Lopez, H.F.: Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference. 2nd edn. Chapman & Hall / CRC (2006)

[HART96] Hartmann, S.: The World as a Process: Simulations in the Natural and Social Scienes. In: Simulation and Modelling in the Social Sciences from the Philiosphy of Sciences Point of View. Kluwer, Dortrecht (1996) 76–100

[HAST70] Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. Biometrika **57** (1970) 97–109

[HLUP98] Hlupic, V. and Robinson, S.: Business process modelling using discrete-event simulation. In Madeiros, D., Watson, E., Karson, J., Manivannan, M., eds.: Proceedings of the Winter Simulation Conference WSC'98, Washington DC, SCS (1998) 1363–1369

[KOEP06] Köppen, V. and Lenz, H.J.: A comparison between probabilistic and possibilistic models for data validation. In Rizzi A, V.M., ed.: Proceeding in Computational Statistics COMPSTAT 2006, Rome (2006) 1533–1541

[KREU86] Kreutzer, W.: Systems Simulation - Programming Styles & Languages. Addison Wesley, Reading (1986)

[LENZ91] Lenz, H.J. and Rödel, E.: Statistical quality control of data. In Gritzmann, P., Hettich, R., Horst, R., Sachs, E., eds.: 16th Symposium on Operations Research, Physica Verlag, Heidelberg (1991) 341–346

[MORG84] Morgan, B.T.: Elements of Simulation. Chapman and Hall, London, New York (1984)

[MUEL07] Müller, R. M.: A Fuzzy-based approach for long-range forecasting In Köppen V., Müller, R.M., eds.: Business Intelligence: Methdos and Application, Verlag Dr. Kovac, Hamburg, (2007)