Simulation Experiments and Significance Tests

Klaus G. Troitzsch

Abstract The paper uses two formal models of simple processes in artificial worlds—one with and one without an analytical solution—to discuss the role of statistical analysis and significance tests of results from multiple simulation runs. Moreover the paper argues that it is not the sheer existence of an effect of input parameters on simulation results but the effect *size* which is the interesting outcome of a simulation and that significance tests of differences in means are much less important than the distribution of output variables which are more often than not non-normal distributions.

1 Introduction

Simulation models, and particularly agent-based simulation models, often come in the form of stochastic models where at least micro-entities perform their actions not deterministically but only with a certain probability which is often derived from expected utilities of the actions available at the time a decision is made. This leads to the fact that simulation results do not only depend on fixed meaningful parameters but also on the seed of a random number generator, such that results of single simulation runs are often classified as insufficient and a larger number of simulation runs is deemed necessary to explore the simulation model.

When relations between input parameters of such stochastic simulation models and simulation outcomes are analysed the question occurs whether such a relation is “statistically significant”, and more often than not one wonders how many simulation runs are necessary for a valid analysis of the simulation outcome. Obviously, both questions cannot be answered at the same time. Increasing the

K.G. Troitzsch (✉)
Institut für Wirtschafts- und Verwaltungsinformatik, Universität Koblenz-Landau, Koblenz, Germany
e-mail: kgt@uni-koblenz.de

S. Leitner and F. Wall (eds.), *Artificial Economics and Self Organization*, Lecture Notes in Economics and Mathematical Systems 669, DOI 10.1007/978-3-319-00912-4_2,

number of parallel simulation runs (i.e. runs with the same combination of input parameters but with different random number generator seeds) automatically leads to a numerical improvement of the "level of significance" without leading to an improved meaningfulness: a correlation or regression coefficient of 0.1 may be significantly different from 0.0 on a certain level when the number of simulation runs whose results enter into the analysis is above a certain threshold (which, of course, depends on the distribution of the coefficient and the chosen level of significance). Thus any statement about the relation between input parameters and simulation output can be "made significant" no matter what the effect size is, as there is always a fixed relation between effect size, significance level and sample size (here: number of simulation runs).[1]

This paper discusses the role of simulation "experiments" in contrast to laboratory and field experiments and the role of a "sample" of simulation runs in contrast to samples drawn from real populations or universes. We will come to the conclusion that multiple runs of otherwise identical simulations which only differ in the random seed (or other technical or physical means of random number generators) are samples in the mathematical meaning of the term (and in so far superior to samples in the social and economic sciences drawn from real populations which are most often biased by low response rates, self-selection and other obstacles to perfect sampling) but that significance testing is not an adequate means of analysing simulation results. This is partly due to the fact that simulation is not an inductive method but a method of deducting new statements from the "first principles" (theoretical assumptions) incorporated into the simulation model, more or less with the same purpose as deriving a formula of a time dependent probability density function—a macro property—from assumptions about the stochastic micro behaviour. This idea leads to a further result: the average of a coefficient or parameter over a large number of runs of the same simulation model is less interesting than its distribution as a whole, and comparing averages with each other or with zero is sometimes useless.

The paper will use two very different models to illustrate the roles of effect sizes and distribution functions as results of simulation models and wherever possible compare these roles to the role of mathematical analysis. But unfortunately this is often only possible in simple models whereas in economics and the social sciences one is more often interested in more complex models. The first of these two models describes a lock-in process mainly with the methods of mathematical analysis and dates back to the 1970s when Wolfgang Weidlich and Hermann Haken applied methods of statistical physics to social and economic phenomena. This model has meanwhile been superseded by more sophisticated models which do not even have

[1]For an in-depth discussion of "the cult of statistical significance" see [8]. Particularly, Ziliak and McCloskey criticise the use of sentences like "The differences reached the level of statistical significance, by and large." [p. 35]. Their general argument is that it is not enough to "decide '*whether* there *exists* an effect' " [p. 25], but that it is necessary to ask "the scientific question 'How much is the effect?' ".

an approximate analytical solution. The second model (of which several different aspects will be discussed) is a model of the role of minimum wages in low-cost sectors of the economy. The comparison of the two models will show the role of simulation as a deductive method.

2 The Role of Analytically Solvable Models

A simulation model can be seen as a formalisation of a theory, and in terms of the 'non-statement view' it can also be seen as a *full model of a theory* which in turn was specified according to the rules laid down in [2] (see also [4]). Thus simulation is a method of deduction, much like the classical mathematical deduction, for cases when analytical deduction does not lead to a closed solution. As most agent-based models (and, indeed, most simulation models) have stochastical components, one can define the deduction of an estimate of a probability or probability density function as one of the main tasks of a simulation. In relatively simple multilevel models this deduction can be done analytically as the following example (originally provided by Weidlich and Haag [7]) shows which deduces the time dependent probability density function for the proportions of proponents and opponents of a certain decision from a microspecification [3] of the behaviour of the micro-entities which can choose between 'yes' and 'no' with a probability (Eq. 1) depending on the current majority (Eq. 3).

$$\begin{aligned} \mu_{yes\leftarrow no} &= \nu \exp(\delta + \kappa x) \\ \mu_{no\leftarrow yes} &= \nu \exp\left[-(\delta + \kappa x)\right] \end{aligned} \tag{1}$$

$$n = n_{yes} - n_{no} \tag{2}$$

$$x = \frac{n_{yes} - n_{no}}{n_{yes} + n_{no}} \tag{3}$$

From this primary assumption one can derive the probabilities w that the whole population changes its state (measured as $n = n_{yes} - n_{no}$ where traditionally $n_{yes} + n_{no} = 2N$).

$$w[(n+1) \leftarrow n] = w_{\uparrow}(n) = n_{-}\mu_{yes\leftarrow no} = (N-n)\mu_{yes\leftarrow no} \tag{4}$$

$$w[(n-1) \leftarrow n] = w_{\downarrow}(n) = n_{+}\mu_{no\leftarrow yes} = (N+n)\mu_{no\leftarrow yes} \tag{5}$$

$$w[j \leftarrow i] = 0 \qquad \text{for} \quad |i-j| > 1 \tag{6}$$

Combining the probabilities p of the population to be in a certain state n with the probabilities w of changing its state leads to the master equation determining the probability of the population of being in one of its possible states $n \in \{-N, \ldots, N\}$ at time t:

$$\begin{aligned}\frac{p(n;t+\Delta t)-p(n;t)}{\Delta t} = & \; p(n+1;t)w_{\downarrow}(n+1) \\ & -p(n;t)(w_{\uparrow}(n)+w_{\downarrow}(n)) \\ & +p(n-1;t)w_{\uparrow}(n-1) \end{aligned} \quad (7)$$

which, by taking the limit $\Delta t \rightarrow 0$ and further simplification, yields the system of linear differential equations[2] for $2N+1$ functions $p(n;t)$:

$$\dot{\mathbf{p}}(t) = \mathbf{L}\mathbf{p}(t) \quad (8)$$

where $\mathbf{p}(t)$ is a vector of the probabilities $p(n;t)$ for all the possible population states, and $\mathbf{L}$ is a matrix which has non-vanishing elements only in the main diagonal and in the two adjacent diagonals, and all its elements are constant:

$$\begin{array}{ll} l_{ii} = -w_{\downarrow}(i) - w_{\uparrow}(i) & \\ l_{ij} = w_{\downarrow}(j) & j = i+1 \\ l_{ij} = w_{\uparrow}(j) & j = i-1 \\ l_{ij} = 0 & |i-j| > 1 \end{array}$$

This, by the way, leads to $\sum_i \dot{p}(i;t) = 0$ for all t, which also fulfils the condition $\sum_i p(i;t) = 1$.

Equation 8 is a system of coupled linear differential equations and could be solved by analytic means, although it is solved numerically here because for a population size of $2N$ the system consists of $2N+1$ equations. By analytic means, however, the stable equilibrium distribution of populations for $t \rightarrow \infty$ may be calculated approximately, where the approximation is fairly good for population sizes above 50, and its minima and maxima can be depicted as a function of the two parameters δ and κ, see Fig. 2.

κ represents the strength of the coupling of the individuals to the majority and determines whether a population is likely to have a 50-50 distribution of 'yes' and 'no' ($\kappa < 1$ for $\delta = 0$) or is likely to have a strong majority of either 'yes' and 'no' ($\kappa > 1$ for $\delta = 0$). With $\delta \neq 0$ and small κ, the most probable majority in a population would be different from 50 % (and with high κ the probability maxima in the right-hand part of Fig. 1 would be of different height). ν is a frequency parameter but it is of little interest: it affects only the time scale of the structure-building process, because with higher ν the breakthrough of either 'yes' or 'no' comes faster. For $\delta = 0$ and $\kappa > 1$ the distribution of populations develops into a bimodal

[2] A numerical solution of this system of differential equations is a simulation of the macro object 'population' with the vector-valued attribute 'probability of being in one of the possible states'.

Fig. 1 Time-dependent probability density functions of the Weidlich-Haag process for two different κ's

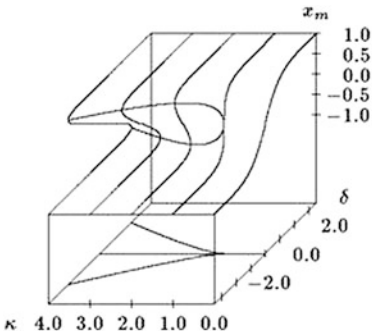

Fig. 2 Dependence of the stationary probability density function of the Weidlich-Haag process for varying δ and κs. x_s is the set of maxima and minima of the stationary probability density function

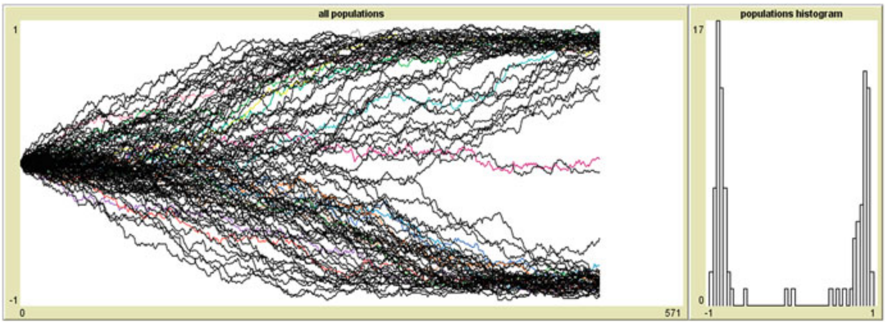

Fig. 3 Trajectories of 100 populations with 200 members each from a simulation of the Weidlich-Haag process for $\kappa = 1.5$s and the histogram of the x distribution

distribution—the probability of finding the population with a strong majority of either 'yes' or 'no' is very high. For $\kappa < 1$ the probability of an evenly split population is very high (see Fig. 1—for $\delta \neq 0$ the threshold for κ is different). Figure 2 shows the dependence of the minima and maxima of the solutions to Eq. 8 on the two parameters κ and δ.

Whereas a stochastic simulation of the process can visualise the time-dependent probability distribution shown in the right-hand plot of Fig. 1—see Fig. 3 which shows a high degree of similarity—an analogue to Fig. 2 is much more difficult to generate from just simulation.

After this short digression into the world of analytically solvable models we can return to strictly formalised models whose solution cannot be found with standard analytical means of what Ostrom [6] called the second symbol system—namely that of mathematics, the first being natural language—but which call for methods taken from the third symbol system, the one of programming languages and computer simulation. Apart from the fact that Fig. 1 had already been output of a numerical simulation (namely the iterative evaluation of Eq. 7) the central difference is now that the numerical values of the $p(n; t)$ can no longer be determined; instead we run a number of realisations of the stochastic process defined in the simulation model and try to reconstruct the shape of the probability or probability density functions from this 'sample' of simulation runs.

3 The Role of Simulation Models

The second model has the same purpose as the first. It deduces the effect of one or two parameters controlling micro behaviour on the outcome which in this case, too, is the probability distribution of one or two macro variables.[3] The model describes the behaviour of employers seeking workers in three different low-cost sectors and of workers seeking employment in one or another of these sectors. In every tick, one employer agent and one worker agent (currently employed or not), both selected at random from the population of this artificial economy, negotiate the wages, and if they find that the wages is above the minimum expectation of the worker and below the maximum expectation of the employer they agree on a contract (in case the worker agent had been employed before it leaves the current employment in favour of the higher wages offered by the new employer agent). Both employer agents and worker agents form their expectations on the base of information they have from other agents in the market such that their expectations range from the minimum to the maximum wages currently paid by or to a selection of their colleagues (and the wages they bid or offer are uniformly distributed between these minima and maxima). The contracted wages, too, are uniformly distributed between bid and offer. Thus a stochastic influence comes into play in three different phases of each simulation step. One external parameter of the model is the minimum wages which serves as a lower bound to both sides' expectations.[4] In the results reported in the next few paragraphs, contracted wages can be as small as the minimum wages which are valid in this artificial economy (even wages near 0 can be contracted, but given that in modern societies something like a basic income is paid to everybody,

[3] This model goes back to discussions between Gregor van der Beek, two master students and the author, was partly documented in the two students' master thesis [5] and extended by the present author.

[4] The current version of the model can be found among the NetLogo User Community Models, http://ccl.northwestern.edu/netlogo/models/community/MinimumWages.

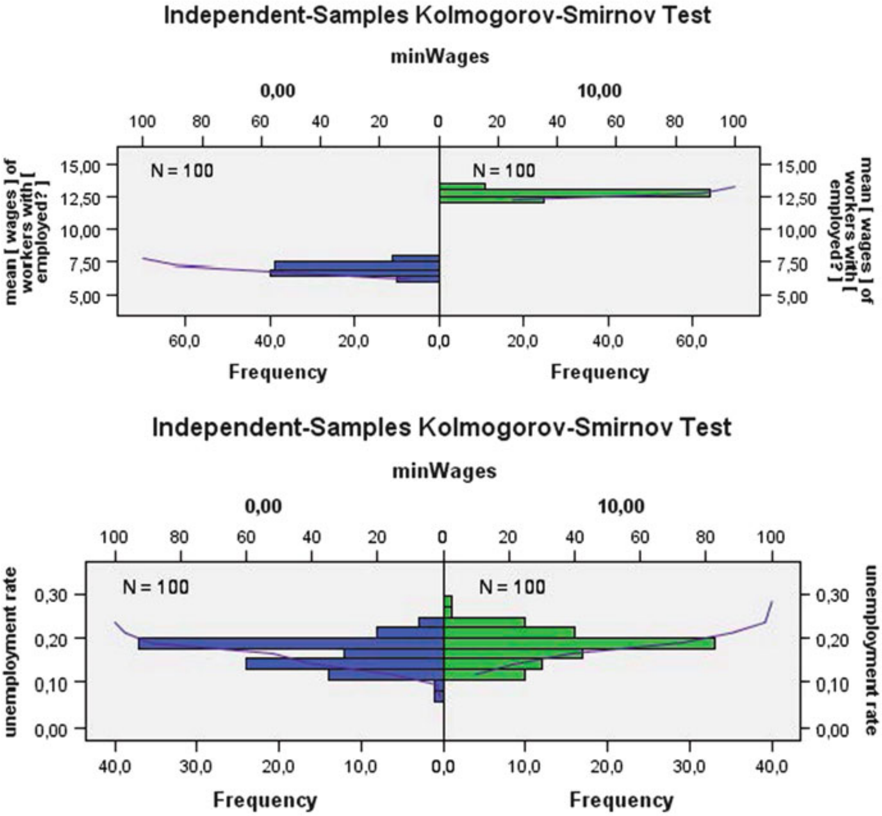

Fig. 4 The effect of minimum wages on average wages and unemployment rate

this makes the following simulation runs slightly unrealistic, for an extension see below).

A first experiment with this model shows the difference between 100 runs each without minimum wages and with a minimum wages of 10 (think of per hour)—see Fig. 4. At first glance one is not surprised to find that the average wages are higher with a minimum wages of 10, but one is a little doubtful whether the minimum wages had an effect on the unemployment rate. The Kolmogorov-Smirnov test says that the null hypothesis that the two distributions are the same should be rejected with an α of 0.037, and a t-test says the same for the comparison of the two means with $\alpha < 0.0005$. Note that the two α's are very different! The left-hand side of Table 1 shows some more information on the two distributions—one should add that the distribution for no minimum wages has two modes for 0.1529 (0.15) and 0.1765 (0.16) with a minimum in between at 0.1647 (0.12). Thus one could believe that the distribution in the no minimum wages case is bimodal, such that the mean is as irrelevant as it is in the case of the lock-in model for higher

Table 1 The effect of minimum wages on the unemployment rate in two different scenarios

	Workers seek ...			
	Random employer Minimum wages		Best employer Minimum wages	
Unemployment rate	0	10	0	10
Mean	0.1667	0.1844	0.1518	0.1521
Standard deviation	0.0327	0.0348	0.0320	0.0298
Mode	0.1765	0.1647	0.1294	0.1412
Median	0.1647	0.1765	0.1529	0.1529
Minimum	0.0706	0.1059	0.0941	0.0824
Maximum	0.2471	0.2824	0.2235	0.2353
Range	0.1765	0.1765	0.1294	0.1529

κ—but with only 85 workers in the current example the two modes with 13 and 15 unemployed and the minimum with 14 unemployed are not convincing. And, moreover, the Kolmogorov-Smirnov test says that the null hypothesis that in both cases the distribution is a normal distribution should not be rejected ($\alpha = 0.300$ and 0.218, respectively).

It goes without saying that the distributions of the means of the contracted wages are very different, as Fig. 4 and later on Fig. 6 show.

This leads to another experiment with a higher number of workers and employers. Instead of 85 workers and 30 employers, these numbers are increased to 300 and 100, respectively. The mean unemployment rate for the 100 runs each for the two values of minimum wages is 0.098667 and 0.149933, respectively, with a standard error of about 0.017 in both cases.

Thus it seems that the unemployment rate is normally distributed over the parallel runs of the simulation with a larger artificial economy, and mean and (perhaps) standard deviation are different for different values of the parameter minimum wages. The question remains whether this statistically significant difference is meaningful at all. The statistical significance estimated from our 100 runs for the two values of the parameter minimum wages means that the hypothesis derived from the simulation model that there is no effect of minimum wages on the unemployment rate has to be rejected at an $\alpha < 0.0005$ (according to the Kolmogorov-Smirnov two-sample test whose test statistics is 6.081 in this case).

To answer this question we return to the original size of our artificial economy (where the difference between the two unemployment rates was considerably smaller (see left part of Table 1) and check the unemployment rates for more values of the parameter minimum wages—besides 0 and 10, also 2, 4, 6 and 8. Figure 5 shows visually that the effect of the minimum wages on the unemployment rate is all but overwhelming.

Analysing this result further, we find that the regression equation between the minimum wages w and the unemployment rate u is $u = 0.162 + 0.002w$, with a standardised regression coefficient of 0.213 and a reduction of variance (R^2) of 0.046 (both of which, too, are significantly different from 0, $\alpha = 0.000195$). Simply formulated this means that a percentage point increase of 1 in the minimum wages

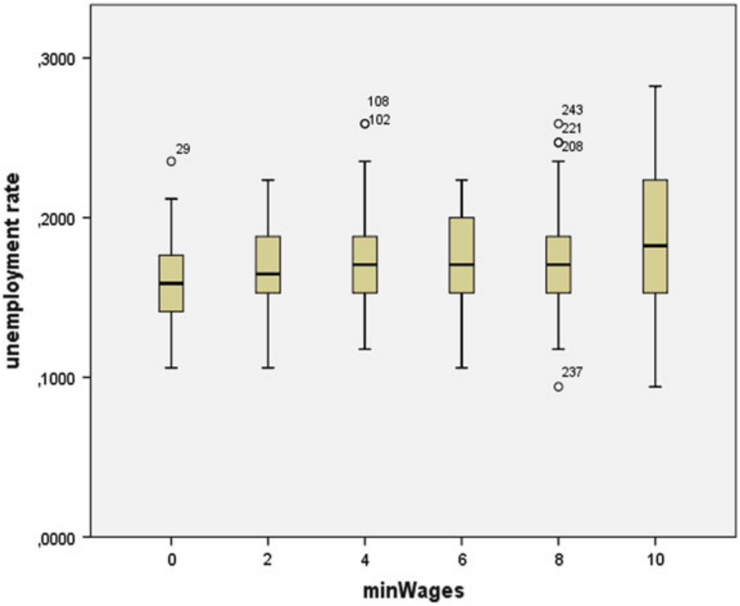

Fig. 5 The effect of different values of the parameter minimum wages on the unemployment rate

leads to an increase of the unemployment rate of 0.2 % points. 95.4 % of the variance of the unemployment rates in the $6 \times 50 = 300$ runs remains unexplained by the minimum wages parameter and can only be explained by the random noise in the model.

The conclusion from this first model setting is: There is a statistically significant effect, but it is small.

Now let us change the setting a little. In the runs discussed so far workers and employers met randomly and started their negotiations, and both only knew the range of wages contracted in the past among their colleagues, formed their ask and bid wages by a random selection of a number within the respective range, and when the employer offered more than the worker had expected, the contract was executed. In what follows we will be a little more realistic: The worker agent randomly selected in every time step knows beforehand how much the individual employers offer and selects the employer with the highest offer (which could perhaps be less than the worker's expectation, but in this case no negotiation is started). Again the contractual wages will be a random value, uniformly distributed between the wages expected by the worker and the offer of the employer—which is the best currently offered. One stochastic element is removed, as compared to the former version, namely the random assignment between workers and employers. This leads to the following result (for an easier comparison documented in the right-hand side of Table 1).

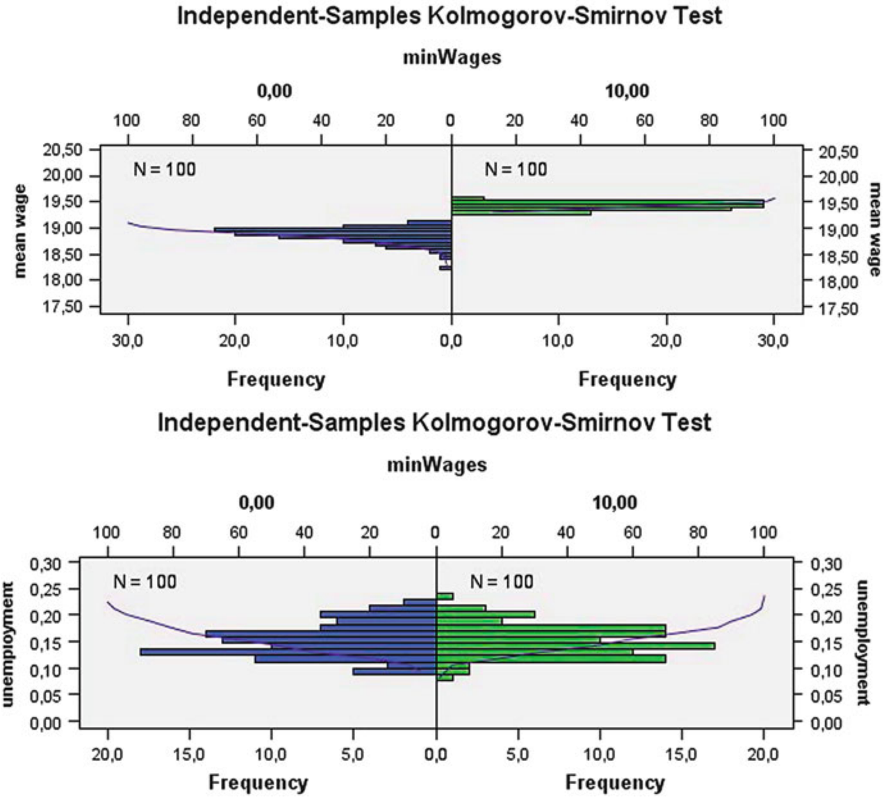

Fig. 6 The effect of minimum wages on average wages and unemployment rate

The median unemployment rate is exactly the same in this setting both with and without minimum wages, and the two distributions also seem to be the same, as the Kolmogorov-Smirnov test suggests to retain this hypothesis ($\alpha = 0.994$, test statistic 0.424, see also Fig. 6 which also shows that the distributions of the average wages for the two values of the minimum wage parameter are different). The hypothesis that the distributions of the unemployment rates are normal again cannot be rejected ($\alpha = 0.77$ and 0.96, respectively), but the histograms are not very similar to a histogram of a normally distributed variable.

Thus under slightly more realistic conditions, the simulation reveals that the introduction of minimum wages does not influence the unemployment rate. This does, of course, not exclude that with a higher number of simulation runs or with a larger size of the employer and worker populations the situation changes. As with the current results the probability distributions over the parallel simulation runs seems to be a normal distribution—which does not come as a surprise, as the random effects included in the simulation model are more or less of the same type and only linearly related—one can expect that the statistical significance, measured as α of the small differences between means will change and that the difference of 0.000353 between the two means will become "statistically significant" when the number of runs is increased to 2,000 instead of 100 for each parameter value. But the question remains whether it is reasonable to run the model twenty times as often to make an unemployment rate percentage point difference of 0.0353 "significant" on the $\alpha = 0.05$ level.

4 Conclusion

Our examples showed that a stochastic simulation model can be used to find out what the distribution of result variables is like and to determine its parameters.

In the case of the first example (where an approximate analytical solution is available in parallel to a multitude of simulation runs) the simulation showed that under certain parameter combinations a bimodal distribution emerges from the assumptions of the model, but it would perhaps be difficult to find out for which κ threshold the bifurcation into a bimodal distribution happens. And it is perhaps impossible to claim from simulations that this threshold is $\kappa = 1.0$—perhaps the best guess from simulation runs with 500 steps as in Fig. 3 is $0.7 < \kappa < 1.3$. This shows that wherever an analytical solution is possible simulations should only be used for illustrative visualisations. Thus one can still support Alker's verdict that simulation is inelegant mathematics [1]. Moreover one cannot decide how many simulation runs are necessary to find out that for $\kappa > 1$ a bimodal distribution emerges—one run is certainly not enough but would only yield the information that the population will quickly show a strong majority of one of the two alternatives.

In the case of the second example where not even an approximate analytical solution is available to the stochastic process and where the standard macro method of finding an equilibrium at the intersection of the demand and supply curves makes one believe that the introduction of minimum wages leads to a higher unemployment rate, we see that the simulation supports the macro derivation in so far as with actual wages between 0 and some maximum and no directed search for an employer agent which offers the highest wages, the unemployment rate is slightly less than in the more realistic version where workers do not contract randomly but only with the employer who currently offers the highest wages. Here the median was identical between the simulation runs with and without minimum wages, and all other comparisons showed no "significant" differences. The distribution over 100 runs for each parameter set shows that two single simulation runs are by no means sufficient to evaluate the difference. In a way the situation in this case is the same as the comparison between two regions of the same country in one of which minimum wages were introduced while in the other region this was not the case. The two regions would certainly show two different rates of unemployment, but given that unemployment rates are influenced by so many other environmental variables this simple "field experiment" comparison should not convince anybody. The multitude of simulation runs, however, allows for a comparison not of two single realisations of two stochastic processes but of two distributions of two stochastic processes.

But what does "significant" mean in this context? For the third comparison of the second example we expect that in 1 out of 2,000 simulation runs (for both values of the minimum wages) we would have found a difference in the mean which a standard t-test would have identified as significantly different from 0. The probability that by pure chance the one and only simulation run had this extraordinary outcome, and the probability that we had restricted ourselves to this singular simulation run is even much less (as one simulation run on a standard PC takes less than 2 min). Thus if we compare significance considerations between analyses of empirical data (where we usually have only one potentially biased sample) and analyses of simulation output (where we can arbitrarily increase the number of usually unbiased samples) we see that significance is not significant in the latter case.

References

1. Alker HR Jr (1974) Computer simulations: inelegant mathematics and worse social science? Int J Math Educ Sci Technol 5:139–155
2. Balzer W, Moulines CU, Sneed JD (1987) An architectonic for science. The structuralist program. Volume 186 of synthese library. Reidel, Dordrecht
3. Epstein JM, Axtell R (1996) Growing artificial societies – social science from the bottom up. MIT, Cambridge MA
4. Ihrig M, Troitzsch KG (2013) An extended research framework for the simulation era. In: Diaz R, Longo F (eds) 2013 Spring simulation multiconference on emerging M&S applications in industry and academia symposium and the modeling and humanities (EAIA and MathH 2013) (SpringSim '13). Simulation Series, vol 45(5). Curran Associates, Red Hook, pp 99–106
5. Jazayeri P, Tohum MH (2012) Auswirkung der Einführung eines Mindestlohns auf den Arbeitsmarkt, anhand eines Simulationsmodells. http://kola.opus.hbz-nrw.de/volltexte/2012/759/
6. Ostrom T (1988) Computer simulation: the third symbol system. J Exp Soc Psychol 24:381–392
7. Weidlich W, Haag G (1983) Concepts and models of a quantitative sociology. The dynamics of interacting populations. Springer series in synergetics, vol 14. Springer, Berlin
8. Ziliak ST, McCloskey DN (2007) The cult of statistical significance. The University of Michigan Press, Ann Arbor

http://www.springer.com/978-3-319-00911-7

Artificial Economics and Self Organization
Agent-Based Approaches to Economics and Social Systems
Leitner, S.; Wall, F. (Eds.)
2014, XV, 256 p. 88 illus., 49 illus. in color., Softcover
ISBN: 978-3-319-00911-7