# Validation and Functional Complexity

**Robert E. Marks**
Australian Graduate School of Management
University of New South Wales
Sydney NSW 2052
Australia

bobm@agsm.edu.au

ABSTRACT: This paper provides a framework for discussing the validity of computer simulation models of market phenomena. It defines functional complexity and derives measures of this for a well known agent-based simulation model and suggests methods to overcome the obstacle of complexity in validating such models..

*Acknowledgments:* I thank Yaneer Bar-Yam, Ian Wilkinson, and three referees for their assistance.

### 1. Introduction

One possible reason for the relative slowness of the economics discipline to embrace computer simulation models of economic phenomena might be lack of confidence in the behaviour and results exhibited by such models. Even if there are other reasons, better validation of such models would reduce the skepticism about their results. Leombruni et al. (2006) go further, arguing that a common protocol for conducting and reporting agent-based social simulations is necessary, including five types of validation (see below). Midgley et al. (2007) also argue for a common approach for establishing the "assurance" (verification, or program validity, and validation) of agent-based models, introducing a five-step process for modellers.

This paper discusses the general issue of validation (for whom? with regard to what?) and its relationship to the use of computer models for explanation in Section 2. Section 3 delineates agent-based computer models from other simulation techniques. Section 4 discusses notions of complexity, and argues that functional complexity is the appropriate measure for the complexity of computer simulation models. We calculate the functional complexity of a well known simulation model, and conclude that such models are too complex to be properly tested. Section 5 discusses possibilities for overcoming this obstacle and concludes.

### 2. Validation

What is a good simulation? The answer to this question must be: a good simulation is one that achieves its aim. But just what the aim or goal of a simulation might be is not obvious. There are several broad possibilities.[1] A simulation might attempt to explain a phenomenon; it might attempt to predict the outcome of a phenomenon; or it might be used to explore a phenomenon, to play, in order to understand the interactions of elements of the structure that produces the phenomenon.

Explanation should result in arriving at sufficient conditions for an observed phenomenon to occur. Figure 1 presents prices and quantities of branded coffee sales by week in a single supermarket chain in the U.S. Midwest in the 1980s, from supermarket scanner data. Several characteristics are obvious from the graph: first, there is a great deal of movement in the market: prices and quantities of at least some brands are not at all stable, and are experiencing great week-to-week changes in their quantities sold. Second, some brands are not altering their prices much at all. Third, for the first group the norm seems to be high prices (and low sales), punctuated every so often by a much lower

---

1. Haefner (2005) lists: usefulness for system control or management, understanding or insights provided, accuracy of predictions, simplicity or elegance, generality (number of systems subsumed by the model), robustness (insensitivity to assumptions), and low cost of building or maintaining the model.

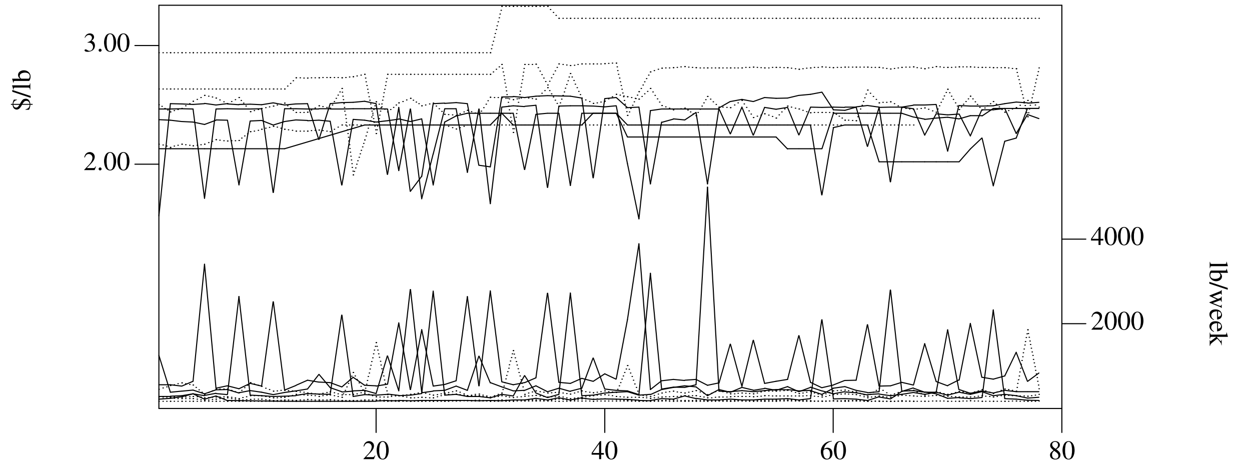


**Figure 1: Weekly Prices and Sales** (Source: Midgley et al. 1997)

price and much higher weekly sales.

What is causing these fluctuations? Is it shifts in demand, whether in aggregate or for specific brands (perhaps in response to non-observed actions such as advertising)? Is it driven by the individual brand managers themselves? If so, are they in turn responding to the previous week's prices? Or it might be that the supermarket chain is moderating the behaviour of the individual brands. A further cause of the behaviour might be the non-price, non-advertising actions of the brands: aisle display and coupons.

Explanation here would occur if the profit-maximising behaviour of the simulated brand managers, together with some external factors or internal factors, led to behaviour qualitatively or quantitatively similar to the behaviour of the brands' prices and quantities sold seen in the figure. Note, following Durlauf (2005), that by making the assumption of profit maximizing, we are going beyond merely seeking a set of equations exhibiting periodicity similar to the "rivalrous dance" of the brands in the figure.

Having sought patterns in the past, and calibrated a model, it becomes possible to predict — to answer the question: what will happen if such and such are the prevailing exogenous conditions? Here, the endogenous variables might include the prices and other marketing actions of one brand manager, or of a set of them, or perhaps of all of them in this market.

There is no need to know which if any alternate conditions will also lead to the observed endogenous behaviours. That is, prediction does not require an understanding of necessity of the underlying exogenous variables. This might explain, as Friedman (1953) argued, that economic actors can behave as though they have solved complex optimisation problems, even though they remain ignorant of the formal representation of the problem or its solution.

Exploration is perhaps the most interesting example of what can be done with simulation models. What are the limits of this behaviour? Under what conditions does it change to another general form of behaviour? Just what ranges of behaviour can the system generate? How sensitive is the model behaviour (and hopefully the real-world behaviour) to changes in the behaviour of a single actor, or of all actors, or of the limits of interactions between players?

The aim of the simulation depends partly on who is simulating (or who is the client), and who will be presenting the results. The first step to convince others that your simulation is appropriate is to convince yourself. Resnick (1996) provides a clear example of a student using a NetLogo simulation to explore emergent behaviour. Barreteau (2003) discuss the value of including the stakeholders in the model's use in the validation of the model.

Leombruni et al. (2006) list five types of validity that theory- and data-based economic simulation studies must consider: theory (the validity of the theory relative to the simuland), model (the validity of the model relative to the theory), program (the validity of the simulating program relative to the model), operational (the validity of the theoretical concept to its indicator or measurement), and empirical (the validity of the empirically occurring true value relative to its indicator). Throughout the paper, we focus on their empirical validity (Manson's (2002) "output validation"): how successfully the simulation model's output exhibits the historical behaviours of the real-world target system, rather than their program validity (Manson's "structural validation"): how well the simulation model represents the (prior) conceptual model of the real-world system. Of course, if the work lacks theory or model or program validity, then it will in general be very difficult to obtain empirical validity.

### 3. Agent-based Computational Economics (ACE)

ACE models are being used more often in the social sciences in general (Gilbert & Troitzsch 2005) and economics in particular (Tesfatsion & Judd 2006). They are also being adopted in marketing research (Midgley et al. 1997), in political science, in finance (Sharpe 2006) and in the multidisciplinary world of electricity markets (Bunn and associates).

In economics, the first ACE models used Genetic Algorithms as a single population of agents. This was acceptable when the players were not differentiated and when the flow of information from parents to offspring at the genotype level was used as an issue (Vriend 2000), but when the players are modelling heterogeneous actors — in realistic coevolution, for instance — each player must be modelled with a separate population.

There has however been a reluctance in economics to embrace simulation in general or ACE modelling in particular. This aversion — or disdain — is mirrored in the discipline's approach to viewing the economic system as a whole — or parts of it such as markets — as complex systems, despite the recent publication of four papers in the June 2005 issue of the *Economic Journal* and the Tesfatsion and Judd *Handbook* (2006).

### 4. Complexity of Agent-Based Simulations

By simulating bottom-up behaviour, agent-based models of such social interactions as market exchanges more closely represent the way phenomena at the macro level, such as prices and aggregate quantities, emerge from behaviour at the micro level than do reduced-form simulations using systems of equations. But there is a trade-off: complexity.

Some would argue that since the world, or at least those phenomena in the world we wish to model, is complex, the models must be complex too. But with complexity comes a challenge: to validate.

Can we put a number to the complexity of such models? Yes. In fact there are several definitions of complexity (Bar-Yam 1997): algorithmic complexity, in which the complexity of the procedures in transforming inputs to outputs is measured; behavioral complexity, which is a measure of the possibilities of the outputs exhibited by the model; descriptive complexity, the amount of information necessary to describe it; and functional complexity (Bar-Yam 2003), the complexity of possible mappings from inputs to outputs of the model. When focussing on validation — to what extent combinations of inputs result in the model exhibiting "correct" outputs (where there are levels of correctness, and contours of the probability that a specific conjunction of inputs will exhibit a target set of outputs), the appropriate measure of complexity is that of functional complexity.

Bar-Yam (2003) defines functional complexity as the relationship of the number of possible inputs and the number of possible outputs:

$$C(f) = C(a)2^{C(e)} \tag{1}$$

The complexity of the function of the system, $C(f)$, equals the complexity of the actions of the system $C(a)$ times two raised to the complexity of the inputs variables of the system, $C(e)$. The three complexities, input $C(e)$ and output $C(a)$ and functional complexity $C(f)$, are defined by the logarithm (base 2) of the number of possibilities, or equivalently, the length of its description in bits.

Bar-Yam argues that this follows from recognizing that a complete specification of the function is given by a table whose rows are the actions ($C(a)$ bits) for each possible input, of which there are $2^{C(e)}$.

Equation (1) is similar to a derivation of mine in 1989 (see Marks 1992) when modelling a $p$-player repeated game with players as stimulus-response automata, responding to the state of play as they know it: with $p$ players, each choosing from $a$ actions and remembering $m$ rounds of play, the number of possible states each must be able to respond to is $a^{mp}$. When modelling players as bit strings with unique mappings from state to action, what is the minimum length of string? The string would require Ceiling[$\log_2(a)$] bits per state, or a total of Ceiling[$\log_2(a)$]$a^{mp}$ bits per player.

For a player in an iterated Prisoner's Dilemma, $a$ is 2, $m$ is 1, and $p$ is 2, resulting in 4 possible states. With only two possible actions, $\log_2(a) = 1$, and the minimum-length bit string is 4. Using equation (1), $C(e) = \log_2(2^2) = 2$, $C(a) = \log_2(2) = 1$, and so $C(f) = 1 \times 2^2 = 4$, nothing more than the minimum-length bit string. More generally, $C(e) = \log_2(a^{mp})$, $C(a) = \log_2(a)$, and so log functional complexity is given by $C(f) = \log_2(a) 2^{\log_2(a^{mp})} = a^{mp} \log_2(a)$ per player.

Functional complexity per player or agent (or the minimum bit-string length for each) is not the model's functional complexity, however. The interactions of the agents may lead to the emergence of higher-level patterns, which is often the object of interest of such models. The log functional complexity of the model will simply be $p$ times the agent's log functional complexity. In the iterated Prisoner's Dilemma, the simulation model has log functional complexity of 8. More generally, the log functional complexity of models of $p$ agents choosing from $a$ actions and remembering $m$ previous rounds of play of a repeated interaction is $pa^{mp} \log_2(a)$, for state-based stimulus-response games.[2]

Bar-Yam further adds that this theorem applies to the complexity of description as defined by the observer — apart from knowing possible inputs and outputs, the function could be a black box — so that each of the quantities can be defined by the desires of the observer for descriptive accuracy. This dovetails well with our purpose here.

Because of the exponential nature of the definition of functional complexity, if the complexity of the inputs or environmental variables to the model is larger then, say, 100 bits, the functional complexity will be huge, and cannot reasonably be specified, as Bar-Yam (2003) remarks.[3]

But all is not lost: for a whole class of micro models the same macro behaviour will emerge, which means that the complexity at the micro scale, although real enough, need not preclude validation of models to exhibit target macro behaviour: with regard to the macro behaviour, many variables are irrelevant. This is known as universality (Bar-Yam 2003). Another way of putting this is that the emerging behaviour at the macro scale is insensitive to fluctuations in the values at the micro scale.

Moreover, before the measure of complexity can have any practical meaning in validation, we must specify the scale of discussion (is it a bitwise comparison or is it at a more macro level?) or the order of magnitude of the accuracy of validation. For instance, any agent-based model can be represented to any desired accuracy by a differential-equation model, and vice versa, and a differential-equation model has infinitely many bits of information, so the scale of accuracy for validation of simulation model is crucial.

The observation that a complex system is frequently comprised of interrelated subsystems that have in turn their own subsystems, and so on until some non-decomposable level of elementary components is reached, together with the observation that interactions inside subsystems are in general stronger and/or more frequent than interactions among subsystems (Courtois 1985), apart from being the basis of the definition of "nearly decomposable systems" (Simon and Ando 1961), provides support for the notion that, whatever micro behaviour (within limits), the macro behaviour that emerges is invariant, so that the complexity of the system at the macro scale is less than the

---

2. A model that uses a summary statistic instead of all possible states might be less complex.

3. In Marks et al. (2006), $a = 8$, $m = 1$, and $p = 4$, resulting in bit strings of length $\log_2(8) \times 8^4 = 12{,}288$, the log functional complexity of this model's bit-string agents, resulting in log functional complexity for the agent-based model of 98,304.

complexity of the aggregation of micro subsystems.

None of this is certain, however: some complex systems are not decomposable, or at least not decomposable under certain external conditions. The challenge is to identify the class of models that are less complex at the macro scale than at the micro, and that capture the properties of a system, that exhibit universality. Related to this issue is the problem of testability of representations through the validation of the mapping of the system to the representation (Bar-Yam 1997).

### *4.1 The Functional Complexity of Schelling's Segregation Model*

In order to put numbers on the measure of functional complexity (equation 1), we have chosen a specific implementation of Schelling's Segregation model (Schelling 1971, 1978): the version freely available for use in NetLogo (Wilensky 1998). Although not strictly an agent-based model (Gilbert and Troitzsch 2005), it is a model in which interactions at the micro scale (between each cell and its eight neighbours) lead at a higher scale to the emergence of segregated regions, under some conditions. By emergence, we mean macro behaviour not from superposition, but from interaction at the micro level. Schelling (2006) explains he used the model to demonstrate that segregated neighbourhoods could arise even when households possessed a degree of tolerance for different kinds of people living next door.

The ASCII code of Segregation (absent any comments) includes 1687 7-bit characters; after compression with the Unix program gzip, which reduces the size of the file using Lempel-Ziv coding (LZ77), the compressed size is 609 8-bit characters. In bits, the size of the code fell from 11,809 bits to 4,872 bits, after compression. The compressed size is a measure of algorithmic complexity, although it ignores the bulk of the NetLogo program, to which the code is only an input, or environmental variable. To obtain the full measure of algorithmic complexity of Segregation, we need the underlying NetLogo code as well.

With the definition of functional complexity, this need is obviated: we only need to have the measures of inputs (environmental variables) and outputs, and then use equation (1). For the NetLogo implementation of Segregation, these are the inputs (environmental variables):

1. The number of "turtles," or inhabitants. This number (call it $N$) ranges between 500 and 2500 in Segregation, which requires 11 bits to specify, strictly $\log_2(N-500)$.
2. The tolerance, or percentage of a similar colour desired as neighbours in the eight adjoining "patches". This number ranges from 0 to 100, which requires 7 bits to specify.
3. The initial randomisation. Each of the turtles is either red or green. Leaving aside how this random pattern is generated, each of the up to 2500 turtles requires a colour bit. Colour increases by 1 the number of bits ($\log_2(N-500)$) required to specify the initially occupied patches.

The NetLogo implementation of Segregation has as outputs two measures and the pattern:

1. The percentage of similar patches. This number ranges from 0.0 to 99.9, which requires 10 bits to specify.
2. The percentage of unhappy turtles. Ranging from 0.0 to 99.9, this too requires 10 bits to specify.
3. The pattern. The pattern appears on a square of $51 \times 51$ patches, where each patch can be red, green, or black (unoccupied). This requires $51 \times 51 \times 3$ = 7,803 possibilities, which requires 13 bits to specify.

If we ignore the two summary measures of the output, which anyway could be calculated from the final pattern, equation (1) indicates that the log functional complexity of this implementation of Segregation is:

$$C(f) = C(a)2^{C(e)} = 13 \times 2^{(\log_2(N-500)+7+1)}$$

The power of 2 in the equation above is limited to 19, so the maximum log functional complexity of this implementation of Segregation is $13 \times 2^{19}$ = 6,815,744, measured in bits. That is, there are at most $2^{6,815,744}$ possible mappings from sets of inputs to unique outputs for this model. Let us pray for universality!

Now, the Segregation model is a highly abstract model, and "it is not clear what data could be used to validate it directly" (Gilbert and Troitzsch 2005). But that's the point: if it were more specific, it would require even more inputs, which would double the functional complexity of the model for each additional bit required to specify the inputs, from equation (1). Although it might be difficult to know what data to use to validate a simpler, more abstract model, on the other hand a more realistic model would have a much greater complexity measure, as the possible mappings from inputs to possible outputs grow in number.

This suggests a trade-off: a less-complex model might in principle be easier to validate by virtue of its relative lack of complexity, while a more realistic model, which in theory could be fitted to historical data as part of its validation, is enormously functionally complex, although the required scale of accuracy must be specified — at some more aggregate level usually.

Given the evident complexities of agent-based models, and the difficulty of determining the class of models insensitive to many micro variables (or robust across Monte Carlo experiments with random micro variables), one is left asking whether any serious decisions could hang on the output of such models. Simulations of economic phenomena are used in most countries to enable policy makers to have some idea of how the national economy is performing. Simulations are also used in a narrower domain by economists considering possible market consequences of proposed mergers between companies, so-called merger simulations, the results of which must be robust enough to convince the court, in many instances, that the competition regulator's decision is the right one. Convincing lawyers and judges of the validity of the results of one's economic simulation model is not easy.

Whether a standard check-list for the validation of simulation models can be, first, developed by the simulation community, and, second, adopted by, say, journal editors as a necessary (but not, of course, sufficient) condition for consideration for publication of papers presenting simulation results remains to be seen. Not just Leombruni et al. (2006), and Midgley et al. (2007) but also Fagiolo et al. (2006) propose agendas for concentrated methodological research by the agent-based modelling community, as a first step.

## 5. Conclusion

As Shannon (1948) taught us, the notion of description is linked to the idea of the space of possibilities. We are drawn to the use of computer models of economic phenomena for at least two reasons: first, to escape the restrictions on the model necessary to obtain closed-form solutions, and, second, to explore the space of possibilities. But the very attractiveness of exploring this vast space creates problems when we want others to have confidence in our results. Because of the degrees of freedom inherent in computer models compared to closed-form solutions — even if the closed-form restrictions enable solutions to uninteresting problems only — the skeptic seeks validation of our results, a validation which is more difficult precisely because of the complexities implicit in the large space of possibilities. This paper has attempted to demonstrate one measure of the complexity of these models, and argued that we must first specify a scale of discussion before meaningful validation can be attempted, even if at a larger scale the model's macro behaviour might be insensitive to some micro variables' specific values. The challenge is to identify the class of models whose macro behaviour is robust to changes in these variables' values.

## 6. References

[1] Barreteau, O., and others (2003), Our companion modelling approach, *Journal of Artificial Societies and Social Simulation,* 6(1). http://jasss.soc.surrey.ac.uk/6/2/1.html

[2] Bar-Yam, Y. (1997), *Dynamics of Complex Systems,* New York: Perseus Press. http://www.necsi.org/publications/dcs

[3] Bar-Yam, Y. (2003), Unifying principles in complex systems, in *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science,* M.C. Roco and W.S. Bainbridge, eds, New York: Springer, pp. 380–409. http://www.necsi.org/projects/yaneer/ComplexSystems.pdf

[4] Brenner, T. and Werker, C. (2006), A practical guide to inference in simulation models, Max Planck Institute of Economics, Evolutionary Economics Group, Jena, #0602. https://papers.econ.mpg.de/evo/discussionpapers/2006-02.pdf (accessed 2006/06/23)

[5] Bunn, D.W. and F.S. Oliveira (2003), Evaluating individual market power in electricity markets via agent-based simulation, *Annals of Operations Research*, 121: 57–77.

[6] Courtois P.-J. (1985), On time and space decomposition of complex structures, *Communications of the ACM*, 28(6): 590–603.

[7] Durlauf S. (2005), Complexity and empirical economics, *The Economic Journal,* 115 (June), F225–F243.

[8] Fagiolo, G., P. Windrum, and A. Moneta, (2006), Empirical validation of agent-based models: a critical survey, LEM Working Paper 2006/14, Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa, Italy, May.

[9] Friedman M. (1953), *Essays in Positive Economics*, University of Chicago Press.

[10] Gilbert N. and K.G. Troitzsch (2005), *Simulation for the Social Scientist,* Buckingham: Open University Press, 2nd edition.

[11] Haefner J.W. (2005), *Modeling Biological Systems: Principles and Applications,* New York: Springer, 2nd ed.

[12] Leombruni, R., Richiardi, M., Saam, N.J., and Sonnessa, M. (2006), A common protocol for agent-based social simulation, *Journal of Artificial Societies and Social Simulation*, 9(1). http://jasss.soc.surrey.ac.uk/9/1/15.html (accessed 22/6/06)

[13] Manson S.M. (2002), Validation and verification of multi-agent systems, in *Complexity and Ecosystem Management,* ed. by M.A. Janssen, Cheltenham: Edward Elgar.

[14] Marks, R.E. (1992) Breeding optimal strategies: optimal behaviour for oligopolists, *Journal of Evolutionary Economics*, 2: 17–38.

[15] Marks R.E., Midgley D.F., Cooper L.G. (2006), Co-evolving better strategies in oligopolistic price wars, in *Handbook of Research on Nature-Inspired Computing for Economy and Management,* edited by Jean-Philippe Rennard, (Hershey, PA: Idea Group Inc., Chapter 52, forthcoming).

[16] Midgley, D.F., R.E. Marks and L.G. Cooper (1997), Breeding competitive strategies, *Management Science*, 43(3): 257–275.

[17] Midgley D.F., Marks R.E., and Kunchamwar D. (2007) The building and assurance of agent-based models: an example and challenge to the field, *Journal of Business Research*, forthcoming.

[18] Resnick, M. (1997), *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*, MIT Press.

[19] Schelling, T.C. (1971), Dynamic models of segregation, *Journal of Mathematical Psychology,* 1: 143–186.

[20] Schelling, T.C. (1978), *Micromotives and Macrobehavior*. New York: Norton.

[21] Schelling, T.C. (2006), Some fun, thirty-five years ago, in Tesfatsion and Judd (2006), 1639–1644.

[22] Shannon, C.E. (1948), A mathematical theory of communication, *Bell Systems Technical Journal*, 27: 379–423 July, 623 August.

[23] Sharpe, W.F., (2006), APSIM: An Asset Price and Portfolio Choice Simulator, in *Investors and Markets: Portfolio Choices, Asset Prices and Investment Advice,* Princeton University Press, forthcoming.

[24] Simon, H.A., and Ando A. (1961), Aggregation of variables in dynamic systems, *Econometrica*, 29: 111–138.

[25] Tesfatsion, L., and Judd, K.L., (eds.) (2006), *Handbook of Computational Economics, Volume 2: Agent-Based Modeling*, in the series *Handbooks in Economics*, (general editors K. Arrow and M. D. Intriligator), (Amsterdam: Elsevier Science).

[26] Vriend, N., (2000), An illustration of the essential difference between individual and social learning and its consequences for computational analyses, *Journal of Economic Dynamics and Control,* 24: 1–19.

[27] Wilensky, U. (1998), NetLogo Segregation model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
ccl.northwestern.edu/netlogo/models/Segregation.