Charlotte Bruun (Ed.)

Advances in Artificial Economics

The Economy as a Complex Dynamic System

With 93 Figures
and 30 Tables

Springer

Editor

Professor Charlotte Bruun
Department of Economics, Politics
and Public Administration
Aalborg University
9100 Aalborg, Denmark
cbruun@socsci.aau.dk

ISBN-10 3-540-37247-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-37247-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Typesetting: Camera ready by author
Cover: Erich Kirchner, Heidelberg
Production: LE-TEX, Jelonek, Schmidt & Vöckler GbR, Leipzig

SPIN 11812852 Printed on acid-free paper – 88/3100 – 5 4 3 2 1 0

Preface

The symposium "Artificial Economics 2006" is the second in a planned line of symposia on artificial economics, following a symposium held in Lille, France in 2005, organized by Phillipe Mathieu, Bruno Beaufils and Olivier Brandouy [1]. The organizing theme of these symposia, is the computational study of economies perceived as complex dynamic systems.

With the latter being a non-existing phenomenon, the defining distinction is not between *artificial* and *natural* economics, but rather between aiming to understand economic processes by constructively simulating them, as opposed to reductionistically analyzing economic systems. With this distinction the game is set, and doors are open for new understandings of economic systems.

Artificial economics is a methodological approach rather than a paradigmatic approach. Neoclassicals, Keynesians, Marxists etc. may all benefit from the methods of artificial economics. Surely some New Classicals have felt the straight jacket of eg. having to assume homogeneous or representative agents, and certainly many Keynesians have dreamt of unifying microeconomics and macroeconomics without totally giving up on their macromodel. Artificial economics provide a toolbox fit for turning towards such fundamental problems anew, without adopting a predetermined idea of what the answers are going to be.

What artificial economics does embrace is an encouragement to economics and economic subdisciplines, to take off the blinkers, and learn about other disciplines. Artificial economics encompasses implementation of ideas and modeltypes from other sciences into economics, integration of different economic submodels, as well as the export of economic conceptions to other sciences. The three invited speakers of Artificial Economics 2006, Akira Nametame, Thomas Lux and Kumaraswamy "Vela" Velupillai, together with a number of contributors, all prove that much may be gained by moving between disciplines.

18

Confronting Agent-Based Models with Data: Methodological Issues and Open Problems

Giorgio Fagiolo[1], Alessio Moneta[2], and Paul Windrum[3]

[1] Faculty of Economics, University of Verona (Italy), and Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa (Italy) giorgio.fagiolo@sssup.it
[2] Max Planck Institute of Economics, Evolutionary Economics Group, Jena (Germany) moneta@econ.mpg.de
[3] Manchester Metropolitan University Business School, Manchester (UK) and MERIT, University of Maastricht (The Netherlands) p.windrum@mmu.ac.uk

Summary. This paper addresses the problem of finding the appropriate method for conducting empirical validation in AB models. We identify a first set of issues that are common to both AB and neoclassical modellers and a second set of issues which are specific to AB modellers. Then, we critically appraise the extent to which alternative approaches deal with these issues. In particular, we examine three important approaches to validation that have been developed in AB economics: indirect calibration, the Werker-Brenner approach, and the history-friendly approach. Finally, we discuss a set of open questions within empirical validation.

18.1 Introduction

Agent-based (AB) researchers in economics have enjoyed significant success over the last twenty years. The models that have been developed indicate the viability and vitality of an alternative to mainstream neoclassical economics. Indeed, deep philosophical differences exist between neoclassical and AB modellers regarding the world faced by real-world agents and, hence, the type of models that it is useful for economists to construct. AB modellers reject the aprioristic commitment of new classical models to individual hyper-rationality, continuous equilibrium, and representative agents. Everything in the neoclassical world can, in principle, be known and understood. It is often assumed that the entire set of objects in the world (e.g. techniques of production, or products) is known at the outset. The opposite is the case in the AB world. Here the set is unknown, and agents must engage in an open-ended search for new objects. Associated with this distinction are important differences with regards to the types of innovative learning and adaptation that are considered, definitions of bounded rationality, the treatment of heterogeneity amongst individual agents and the interaction between these individuals, and whether the economic system is characterized as being in equilibrium or far from equilibrium. Mainstream economists have often recognized the significance of the AB *Weltanschauung*, and have reacted by extending their own modelling framework to incorporate (certain) aspects of heterogeneity, bounded rationality, learning, increasing returns, and technological change. Another sign of the vitality of the AB community has been the development of its own specialist international journals and annual conferences, and the diffusion of its ideas to other areas such as management science, political science and to policy circles.

Nevertheless, there is a perceived lack of robustness in AB modelling, due to the problematic relationship between AB models and empirical data. There is a lack of standard techniques not only for constructing and analyzing AB models, but also to conduct empirical validation. Key areas of debate include: is a 'realist' methodology appropriate? Why should empirical validation be the primary basis for accepting or rejecting a model? Do other tests of model validation exist than the reproduction of stylised facts? If we do proceed down the path of empirical validation, then how should one relate and calibrate the construction of parameters, initial conditions, and stochastic variability in AB models to the existing empirical data? Which classes of empirically observed objects do we actually want to replicate? How dependable are the micro and macro stylised facts to be replicated? To what extent can we truly consider output traces to be stylised facts or, alternatively, counterfactuals? What are the consequences, for the explanative power of a model, if the stylised facts are actually 'unconditional objects' that only indicate properties of stationary distributions and, hence, do not provide information on the dynamics of the stochastic processes that generated them?

The aim of this paper is to provide a critical overview of how AB modellers have been tackling the issue of empirical validation. A strongly heterogeneous set of approaches can be found in the AB literature. An important (and novel) contribution of the paper is a taxonomy that maps the different dimensions of the empirical validation approaches found in AB models. In the next section we shall draw attention to some crucial issues of empirical validation, faced by both AB and neoclassical modellers.

18.2 Core Issues of Empirical Validation

Any model isolates some features of an actual phenomenon. It is usually assumed, in economics as in any other science, that some causal mechanism (deterministic or non-deterministic) has produced the data. We call this causal mechanism "real-world data generating process" (rwDGP). A model approximates portions of the rwDGP by means of a "model data generating process" (mDGP). The mDGP must be simpler than the rwDGP and generates a set of simulated outputs. The extent to which the mDGP is a good representation of the rwDGP is evaluated by comparing the simulated outputs of the mDGP with the real-world observations of the rwDGP. We identify a set of key methodological issues associated with this process of backward induction. These issues are generic in empirical validation, and so apply to neoclassical and AB economists alike.

The first issue is how to deal with the trade-off between concretisation and isolation. Faced with the essential complexity of the world, scientific (not only economic) models proceed by simplifying and focusing on the relationships between a very limited number of variables. Is it possible to model all the different elements of the rwDGP? How can we possibly know all the different elements of the rwDGP? Leading economists (for example, J.S. Mill and J. M. Keynes) have in the past expressed serious doubts about whether we can expect to have models that are fully concretised. In a highly complex world, a fully concretised model would be a one-to-one mapping of the world itself! Thus, economists usually agree that models should isolate some causal mechanisms, by abstracting from certain entities that may have an impact on the phenomenon under examination [13]. A series of open questions remains. How can we assess that the mechanisms isolated by the model resemble the mechanisms operating in the world? In order to isolate the mechanisms, can we make assumptions 'contrary to fact,' that is, assumptions that contradict the knowledge we have of the situation under discussion? This also related to the trade-off between analytical tractability and descriptive accuracy that is faced by all theoreticians seeking to model markets and other economic systems. Indeed, the more accurate and consistent is our knowledge about reality with respect to assumptions, and the more numerous the number of parameters in a model, the higher is the risk of failing to analytically solve the model. By contrast, the more abstract and simplified the model, the more analytically tractable it is. The neoclassical paradigm comes down strongly on the side of analytical tractability.

This brings us to the second core issue of empirical validation: instrumentalism versus realism. Realism, roughly speaking claims that theoretical entities 'exist in the reality,' independent of the act of inquiry, representation or measurement [14]. On the contrary, instrumentalism maintains that theoretical entities are solely instruments for predictions and not true descriptions of the world. A radical instrumentalist is not much concerned with issues of empirical validation, in the sense that (s)he is not much interested in making the model resemble mechanisms operating in the world. His/her sole goal is prediction. Indeed, a (consistent) instrumentalist is usually more willing than a realist to 'play' with the assumptions and parameters of the model in order to get better predictions. While the neoclassical paradigm has sometimes endorsed instrumentalist statements à la Friedman [7], it has never allowed a vast range of assumption adjustments in order to get better predictions. In this sense it fails to be consistent with its instrumentalist background.

The third issue is related to the choice of a pluralist or apriorist methodology. Methodological pluralism claims that the complexity of the subject studied by economics and the boundedness of our scientific representations implies the possibility of different levels of analysis, different kinds of assumptions to be used in model-building, and legitimacy of different methodological positions. Apriorism is a commitment to a set of a priori assumptions. A certain degree of commitment to a set of priori assumptions is normal in science. Often these assumptions correspond to what Lakatos [9] called the 'hard core' assumptions of a research program. But strong apriorism is the commitment to a set of a priori (possibly contrary to the facts) assumptions that are never exposed to empirical validation (e.g. general equilibrium and perfect rationality). Theory is considered prior to data and it is denied the possibility of interpreting data without theoretical presuppositions. Typically, strong apriorist positions do not allow a model to be changed in the face of anomalies, and encourages the researcher to produce ad hoc excuses whenever a refutation is encountered. Lakatos [9] dubbed the research programs involved with such positions as 'degenerating.'

The fourth issue regards the under-determination or identification problem. This is the problem that different models can be consistent with the data that is used for empirical validation. The issue is known in the philosophy of science as the 'under-determination of theory by data.' In econometrics, the same idea has been formalised and labelled 'the problem of identification.' As Haavelmo [8] noted, it is impossible for statistical inference to decide between hypotheses that are observationally equivalent. He suggested specifying an econometric model in such a way that (thanks to restrictions derived from economic theory) the problem of identification does not arise. The under-determination problem is also strictly connected to the so-called Duhem-Quine thesis: it is not possible to test and falsify a single hypothesis in isolation. This is because any hypothesis is inevitably tied to some auxiliary hypotheses. Auxiliary hypotheses typically include background knowledge, rules of inference, and experimental design that cannot be disentangled from the hypothesis we want to test. Thus, if a particular hypothesis is found to be in conflict with the evidence, we cannot reject the hypothesis with certainty, since we do not know if it is the hypothesis under test or one of the auxiliary hypotheses which is at odds with the evidence. As shown by Sawyer et al. [16], hypothesis testing in economics is further complicated by the approximate nature of theoretical hypotheses. The error in approximation, as well as the less systematic causes disturbing the causal mechanism object of modelling, constitutes an auxiliary hypothesis of typically unknown dimension. For example, in time-series econometric models a distinction is made between 'signal' (which captures the causal mechanisms object of interest) and noise (accounted by the error terms). But it may be the case, as pointed out by Valente [17], that noises are stronger than signals, and that the mechanisms involved undergo several or even continuous structural changes. Econometricians have adopted sophisticated tests which are robust to variations in the auxiliary hypotheses (see, for example, [10]). Nonetheless, the Duhem-Quine thesis still undermines strong apriorist methodologies that do not check the robustness of the empirical results under variations of background assumptions.

18.3 A Taxonomy of the Existing Approaches

A discrete set of approaches for empirical validation, not only different with each other but different to those developed within neoclassical economics, have been developed by the AB community. We suggest that there are two reasons for this heterogeneity. First, AB modellers are interested in phenomena such non-linearities, stochastic dynamics, non-trivial interactions among agents, and feedbacks between the micro and the macro level. These are not amenable to traditional equilibrium modelling approaches and tools. One of the consequences is that AB modellers face an additional set of issues that are not faced by neoclassical modellers. Second, and relatedly, the highly diverse structural content of AB models means they need to be analyzed in very different ways. We propose a taxonomy that maps out the key areas in which AB researchers differ.

The first dimension is the *nature of the objects under study*. This determines the stylised facts (empirically observed facts) that the model is seeking to explain. Significant differences exist with respect to the nature of the objects being studied in AB models. Where neoclassical modellers are interested in quantitative change, AB modellers are equally interested in qualitative change of economic systems themselves. For instance, there are AB models that investigate how R&D spending affects the qualitative nature of macroeconomic growth. Other AB models investigate its quantitative impact, or else seek to explain some statistically observed quantitative property of aggregate growth (e.g. its autocorrelation patterns). Another important distinction is between AB models that seek to investigate a single phenomenon, and those that jointly investigate multiple phenomena. For instance, a model may consider the properties of productivity and investment time-series, in addition to the properties of aggregate growth. Transient versus long-run impact is a further distinction. For example, there are AB models that examine the effect of R&D spending on growth along the diffusion path (the transient) of a newly introduced technology. Other AB models are only concerned with the magnitude of a technologys long-run impact (when the economic system has stabilised somewhat). Finally, an important distinction exists between AB models that investigate micro distributions and macro aggregates. The former are concerned with the dynamics of industry-level distributions, such as a cross-section of firm productivity distributions, for a particular sector, in a particular year. The latter are concerned with longer time-series data for nation states, or the world economy, over a number of years.

A second dimension in which AB models differ is in the *goal of the analysis*. AB models tend to deal with in-sample data. In-sample data is relevant when one is interested in describing or replicating observed phenomena. Out-of-sample exercises, although they are less frequently carried out by AB economists, are essential for the sake of policy evaluation.

A third dimension concerns the nature of the most important *modelling assumptions*. Some models contain many degrees of freedom, others do not. For example, agents in AB models may be characterised by many variables and parameters. Their decision rules may, in turn, be highly-parameterised. Alternatively, agents and decision rules may be described in a very stylised way. Individual decision rule sets and interaction structures may be exogenously fixed. They may change over time. Change may be driven by exogenous, stochastic factors. Alternatively, change may be driven by agents endogenously selecting new decision rules and interaction structures according to some meta-criteria (as it happens in endogenous network formation models, see [6]).

The fourth and final dimension is the *methodology of analysis*. In order to thoroughly assess the properties of an AB model, the researcher needs to perform a detailed sensitivity analysis. This sensitivity analysis should, at the very least, explore how the results depend on (i) micro-macro parameters, (ii) initial conditions, and (iii) across-run variability induced by stochastic elements (e.g. random initial conditions, and random individual decision rules).

There are three important approaches to empirical validation within AB economics: indirect calibration [5], [4], the Werker-Brenner approach to empirical calibration [18], and the history-friendly approach [12], [11].

The *indirect calibration approach* is based on a four-step procedure. In the first step, the modeller identifies a set of stylised facts that (s)he is interested in reproducing and/or explaining with a model. Stylised facts typically concern the macro-level (e.g. the relationship between unemployment rates and GDP growth) but can also relate to cross-sectional regularities (e.g. the shape of the distributions on firm size). In the second step, along with the prescriptions of the empirical calibration procedure, the researcher builds the model in a way that keeps the microeconomic description as close as possible to empirical and experimental evidence about microeconomic behaviour and interactions. This step entails gathering all possible evidence about the underlying principles that inform real-world behaviours (e.g. of firms, consumers, and industries) so that the microeconomic level is modelled in a not-too-unrealistic fashion. In the third step, the empirical evidence on stylised facts is used to restrict the space of parameters, and the initial conditions if the model turns out to be non-ergodic. In the fourth and final step, the researcher should deepen his/her understanding of the causal mechanisms that underlie the stylised facts being studied and/or explore the emergence of 'fresh' stylised facts (i.e. statistical regularities that are different to from the stylised facts of interest), against which the model can be validated ex post). This might be done by further investigating the subspace of parameters that resist to the third step, i.e. those consistent with the stylised facts of interest.

A stream of recent AB contributors to the field of industry and market dynamics has been strongly rooted in the four-step empirical validation procedure just presented. For example, Fagiolo and Dosi [4] study an evolutionary growth model that is able to reproduce many stylised facts about output dynamics, such as I(1) patterns of GNP growth, growth-rates autocorrelation structure, absence of size-effects, etc., while explaining the emergence of self-sustaining growth as the solution of the trade-off between exploitation of existing resources and exploration of new ones. Similarly, Fagiolo et al. [5] present a model of labour and output market dynamics that is not only able to jointly reproduce the Beveridge curve, the Okun curve and the wage curve, but also relates average growth rates of the system to the institutional set-up of the labour market.

The *Werker-Brenner approach* is a three-step procedure for calibrating AB models. The first two steps are consistent with all calibration exercises. The third step is novel. Step 1 uses existing empirical knowledge to calibrate initial conditions and the ranges of model parameters. As mentioned above, AB models contain many dimensions, including the set of assumptions about agents behaviour, their actions, interactions, causal relationships, and the simplifying assumptions of the model. Werker-Brenner propose that, where sensible data are not available, the model should be left as general as possible, i.e. wide ranges should be specified for parameters on which there is little or no reliable data.

Step 2 involves empirical validation of the outputs for each of the model specifications derived from step 1. Through empirical validation, the plausible set of dimensions within the initial dimension space is further reduced. It is possible to run the model specification and generate a Monte Carlo set of micro and macro time-series data for that particular combination of empirically-plausible parameter values. The resulting time-series data — one for each parameter combination — can be thought of as a particular 'theoretical realisation' of the model that is being tested. Of course, any two time-series may overlap to a large extent. This is to be expected since the combinations of parameter values that are being tested are likely to be similar in some dimensions, while different in others. Having generated a set of theoretical realisations for each model specification, one is able to compare these outputs with real-world data. The real-world data that we observe are an 'empirical realisation' that is generated by the rwDGP that we are trying to model. The Werker-Brenner approach advocates the use of Bayesian inference procedures in order to conduct this output validation. Each model specification is assigned a likelihood of being accepted based on the percentage of 'theoretical realisations' that are compatible with each 'empirical realisation.' In this way, empirically observed realisations are used to further restrict the initial set of model specifications (parameter values) that are to be considered. The modeller only retains those parameter values (i.e. model specifications) that are associated to the highest likelihood by the current known facts (i.e. empirical realisations). Model specifications that conflict with current data are discounted.

From a methodological perspective, it is step 3 of the Werker-Brenner approach that is of particular interest. The aim is to find an explanation to the phenomena being studied by exploring the remaining set of model specifications. This is achieved through methodological 'abduction.' Abduction is a process that seeks to describe and explain empirical facts in terms of their underlying structures [18]. In practice, this involves a further validation exercise for all empirical realisations that can be collected. Here, however, the modeller focuses on the shared properties and the characteristics shared by all surviving model specifications in order to identify the invariant properties of the underlying structural model. The authors argue that "these [shared] characteristics can be expected to hold also for the real systems (given the development of the model has not included any crucial and false premises)" [18]. If the characteristics within a group of model specifications differ, then this also offers important insights. "It can be examined which factors in the model are responsible for the differences. Hence, although we will not know the characteristics of the real systems in this case, we will obtain knowledge about which factors cause different characteristics" [18].

While the Weker and Brenner's calibration approach addresses the over-paramete-risation problem by reducing the space of possible 'worlds' that are explored in an AB model, the *history-friendly approach* offers an alternative solution to this problem. Like the calibration approaches discussed above, it seeks to bring modelling more closely 'in line with the empirical evidence' and thereby constrains the analysis to reduce the dimensionality of a model. The key difference is that this approach uses the specific historical case studies of an industry to model parameters, agent interactions, and agent decision rules. In effect, it is a calibration approach which uses particular historical traces in order to calibrate a model.

In part, the history-friendly approach represents an attempt to deal with criticisms levelled at early neo-Schumpeterian AB models of technological change. Two of the key protagonists of history-friendly modelling, R. Nelson and S. Winter, were founding fathers of neo-Schumpeterian AB modelling. While the early models were much more micro-founded and empirically-driven than contemporary neoclassical models, empirical validation was weak. There was a lack of thorough sensitivity and validation checks and empirical validation, when carried out, tended to consist of little more than a cursory comparison of outputs generated by a just a handful of simulation runs with some very general stylised facts. Further, the early models contained many dimensions and so it was rather easy to generate a few outputs that matched some very general observations (the over-parameterisation problem).

In terms of our taxonomy, the history-friendly approach is strongly quantitative and mainly focuses on microeconomic transients (industrial paths of development). In this approach a good model is one that can generate multiple stylised facts observed in an industry. The approach has been developed in a series of papers. Key amongst these are [12] and [11]. In [12], Malerba, Nelson, Orsenigo and Winter outlined the approach and then applied it to a discussion of the transition in the computer industry from mainframes to desktop PCs. In [11], the approach was applied to the pharmaceutical industry and the role of biotech firms therein. Through the construction of industry-based AB models, detailed empirical data on an industry informs the AB researcher in model building, analysis and validation. Models are to be built upon a range of available data, from detailed empirical studies to anecdotal evidence

Table 18.1. Taxonomy of dimensions of heterogeneity in empirical validations of AB models

Approach	Domain of Application	Which kind of data should one employ?	How to employ data?	What to do first?
Indirect Calibration	-Micro (industries, markets) -Macro (countries, world economy)	-Empirical data	-Assisting in model building -Validating simulated output	-First validate, then indirectly calibrate
Werker-Brenner	-Micro (industries, markets) -Macro (countries, world economy)	-Empirical data -Historical knowledge	-Assisting in model building -Calibrate initial conditions and parameters -Validating simulated output	-First calibrate, then validate
History-Friendly	-Micro (industries, markets)	-Empirical data -Casual, historical and anecdotic knowledge	-Assisting in model building -Calibrate initial conditions and parameters -Validating simulated output	-First calibrate, then validate

to histories written about the industry under study. This range of data is used to assist model building and validation. It should guide the specification of agents (their behaviour, decision rules, and interactions), and the environment in which they operate. The data should also assist the identification of initial conditions and parameters on key variables likely to generate the observed history. Finally, the data are to be used to empirically validate the model by comparing its output (the simulated trace history) with the actual history of the industry. It is the latter that truly distinguishes the history-friendly approach from other approaches. Previous researchers have used historical case studies to guide the specification of agents and environment, and to identify possible key parameters. The authors of the history-friendly approach suggest that, through a process of backward induction one can arrive at the correct set of structural assumptions, parameter settings, and initial conditions. Having identified the correct set of 'history-replicating parameters,' one can carry on and conduct sensitivity analysis to establish whether (in the authors' words) 'history divergent' results are possible.

Table 18.1 summarizes the main characteristics of the three different approaches. The first dimension, in which these approaches differ, is the domain of application. The direct and indirect calibration approaches can, in principle, be applied to micro and macro AB models (e.g. to describe the dynamics of firms, industries, and countries). By contrast, the history-friendly approach only addresses micro dynamics. A second dimension of heterogeneity is the type of data that are used for empirical validation. In addition to empirical datasets, the Werker-Brenner approach advocates the use of historical knowledge. The history-friendly approach allows one to employ casual and anecdotic knowledge as well. The third dimension is the way in which data is actually used. All three approaches use data to assist model building, as well as validating the validation of the simulated outputs of models. Unlike the other two approaches, indirect calibration does not directly employ data to calibrate initial conditions and parameters. The fourth dimension is the order in which validation and calibration is performed. Both the Werker-Brenner and the history-friendly approaches first perform calibration and then validation. By contrast, the indirect calibration approach first performs validation, and then indirectly calibrates the model by focusing on the parameters that are consistent with output validation.

18.4 Open-ended Issues and Conclusions

There is a set of core issues that affect all the approaches and which (so far) remain unresolved. In this concluding section we shed some light on that.

1. *Alternative strategies for constructing empirically-based models.* There is intense debate about the best way to actually construct empirically-based models, and to select between alternative models. What happens, for instance, if there are alternative assumptions and existing empirical data does not assist in choosing between them? A number of different strategies exist for selecting assumptions in the early stages of model building [3]. One strategy is to start with the simplest possible model, and then proceed to complicate the model step-by-step. This is the KISS strategy: 'Keep it simple, stupid!' A very different strategy is the KIDS strategy: 'Keep it descriptive, stupid!' Here one begins with the most descriptive model one can imagine, and then simplify it as much as possible. The third strategy, common amongst neoclassical economists, is TAPAS: 'Take A Previous model and Add Something.' Here one takes an existing model and successively explores the assumption space through incremental additions and/or the relaxation of initial assumptions.

2. *Problems that arise as a consequence of over-parameterisation in AB models.* Whatever the strategy employed, the AB modeller often faces an over-parameter-isation problem. AB models with realistic assumptions and agent descriptions invariably contain many degrees of freedom. There are two aspects to the over-parameterisation problem. Firstly, the dimensions of the model may be so numerous that it can generate any result. If this is the case, then the explanative potential of the model is little better than a random walk. Secondly, the causal relations between assumptions and results become increasingly difficult to study. A possible strategy is to use empirical evidence to restrict the degrees of freedom, by directly calibrating initial conditions and/or parameters. Then, one can indirectly calibrate the model by focussing on the subspace of parameters and initial conditions under which the model is able to replicate a set of stylised facts. Unfortunately, this procedure still tends to leave the modeller with multiple possible 'worlds.'

3. *The usefulness and implications of counterfactuals for policy analysis* How does one interpret the counterfactual outputs generated by a model? It is tempting to suggest that outputs which do not accord with empirical observations are counterfactuals, and that the study of these counterfactuals are useful for policy analysis. Cowan and Foray [2] suggest that it is exceedingly difficult, in practice, to construct counterfactual histories because economic systems are stochastic, non-ergodic, and structurally evolve over time. As AB models typically include all these elements in their structure, Cowan and Foray argue that using (evolutionary) AB models to address counterfactual-like questions may well be misleading. More generally, comparing the outputs generated by AB models with real-world observations involves a set of very intricate issues. For example, Windrum [19] observes that the uniqueness of historical events sets up a whole series of problems. In order to move beyond the study of individual traces, we need to know if the distribution of output traces generated by the model mDGP approximates the actual historical traces generated by the rwDGP under investigation. A way to circumvent the uniqueness problem is to employ a strong invariance assumption on the rwDGP, thereby pooling data that should otherwise be considered a set of unique observations. For example, one typically supposes that cross-country aggregate output growth rates come from the same DGP. Similarly, it is supposed that the process that driving firm growth does not change across industries or time (up to some mean or variance scaling). This allows one to build cross-section and time-series panel data. Unfortunately we cannot know if the suppositions are valid. But this is often not possible in practice. Consider the following example. Suppose the rwDGP in a particular industry does not change over time (i.e. it is ergodic). Even if this is the case, we do not typically observe the entire distribution of all observations but rather a very limited set of observations — possibly only one, unique roll of the dice. The actual history of the industry we observe is only one of a set of possible worlds. So how do we know that the actual historical trace is in any sense typical (statistically speaking) of the potential distribution? If we do not know this, then we have nothing against which to compare the distributions generated by our model. We cannot determine what is typical, and what is atypical.

4. *Definition of sufficiently strong empirical tests.* The fundamental difficulties in defining strong tests for model outputs is highlighted by Brock's [1] discussion of 'unconditional objects' in economics. Empirical regularities need to be handled with care because we only have information on the properties of stationary distributions. The data that we observe does not provide information on the dynamics of the stochastic processes that actually generated them. Therefore, replication does not necessary imply explanation. For example, many evolutionary growth models can generate similar outputs on differential growth-rates between countries, technology leadership and catch-up, even though they differ significantly with respect to the behaviour and learning procedures of agents, and in their causal mechanisms [19]. Similarly, the Nelson and Winter [15] model replicates highly aggregated data on time paths for output (GDP), capital and labour inputs, and wages (labour share in output), but these outputs can also be replicated by conventional neoclassi-

cal growth models. In the same vein, there might be many different stochastic processes (and therefore industry dynamic models) that are able to generate, as a stationary state, a power-law distribution for the cross-section firm size distribution. Although one may be unable to narrow down a single model, we may be able to learn about the general forces at work, and to restrict the number of models that can generate a set of statistical regularities [1]. Therefore, as long as the set of stylised facts to be jointly replicated is sufficiently large, any 'indirect' validation could be sufficiently informative, because it can effectively help in restricting the set of all stochastic processes that could have generated the data displaying those stylised facts. Another way out the conditional objects critique would be to not only validate the macro-economic output of the model, but also its micro-economics structure, e.g. agents behavioural rules. This requires one to only include in the model individual decision rules (e.g. learning) that have been validated by empirical evidence. Of course, this would require highly detailed and reliable data about micro-economic variables, possibly derived from extensive laboratory experiments.

5. *Availability, quality and bias of datasets.* Empirically-based modelling depends on high quality datasets. Unfortunately, the datasets that exist are invariably pre-selected. Not all potential records are retained; some are fortuitously bequeathed by the past but others are not captured. Datasets are constructed according to criteria that reflect certain choices and, as a consequence, are biased. As econometricians know only too well, it may simply be the case that data that would have assisted in a particular discussion has simply not been collected. A further and often neglected problem is that standard econometric methods are influenced by prevailing theoretical orthodoxy.

References

[1] Brock W (1999) Scaling in economics: a reader's guide, Industrial and Corporate Change, 8: 409-446

[2] Cowan R., Foray D (2002) Evolutionary economics and the counterfactual threat: on the nature and role of counterfactual history as an empirical tool in economics, Journal of Evolutionary Economics, 1 2: 539-562

[3] Edmonds B, Moss S (2005) From KISS to KIDS - an 'anti-simplistic' modelling approach. In Davidsson P, Logan B, Takadama K (eds.): Multi Agent Based Simulation 2004. Springer, Lecture Notes in Artificial Intelligence, 3415:130-144

[4] Fagiolo G, Dosi G (2003) Exploitation, Exploration and Innovation in a Model of Endogenous Growth with Locally Interacting Agents, Structural Change and Economic Dynamics, 14: 237-273

[5] Fagiolo G, Dosi G, Gabriele R (2004a) Matching, Bargaining, and Wage Setting in an Evolutionary Model of Labor Market and Output Dynamics, Advances in Complex Systems, 14: 237-273

[6] Fagiolo G, Marengo L, Valente M (2004b) Endogenous Networks in Random Population Games, Mathematical Population Studies, 11: 121-147

[7] Friedman M (1953) The Methodology of Positive Economics, in Essays in Positive Economics, Chicago: University of Chicago Press

[8] Haavelmo T (1944) The Probability Approach in Econometrics, Econometrica, 12: 1-115

[9] Lakatos I (1970) Falsification and the Methodology of Scientific Research Programmes, in I Lakatos and A. Musgrave, Criticism and the Growth of Knowledge, Cambridge: Cambridge University Press, pp. 91-196

[10] Leamer EE (1978) Specification Searches, Ad Hoc Inference with Non-experimental Data, New York: John Wiley

[11] Malerba F, Orsenigo L (2001) Innovation and market structure in the dynamics of the pharmaceutical industry and biotechnology: towards a history friendly model, Conference in Honour of Richard Nelson and Sydney Winter, Aalborg, 12th - 15th June 2001

[12] Malerba F, Nelson RR, Orsenigo L, Winter SG (1999) History friendly models of industry evolution: the computer industry, Industrial and Corporate Change, 8: 3-41

[13] Mäki U (1992) On the Method of Isolation in Economics, Poznan Studies in the Philosophy of the Sciences and the Humanities, 26:19-54

[14] Mäki U (1998) Realism, in JB Davis, D Wade Hands U Mäki (eds.), The Handbook of Economic Methodology, Cheltenham, UK: Edward Elgar, pp. 404-409

[15] Nelson R, Winter S (1982) An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge

[16] Sawyer K R, Beed C, Sankey H (1997) Underdetermination in Economics. The Duhem-Quine Thesis, Economics and Philosophy, 13: 1-23

[17] Valente M (2005) Qualitative Simulation Modelling, Faculty of Economics, University of L'Aquila, L'Aquila, Italy, mimeo

[18] Werker C, Brenner T (2004) Empirical Calibration of Simulated Models, Papers on Economics and Evolution 0410, Max Planck Institute of Economics, Jena

[19] Windrum P (2004) Neo-Schumpeterian simulation models, forthcoming in H Hanusch, A Pyka (eds.) The Elgar Companion to Neo-Schumpeterian Economics, Edward Elgar: Cheltenham