LEM
Working Paper Series

Empirical Validation of Agent-Based Models:
A Critical Survey

Giorgio FAGIOLO*
Paul WINDRUM**
Alessio MONETA***

*Faculty of Economics, University of Verona,
and LEM, Sant'Anna School of Advanced Studies, Pisa (Italy)
** Manchester Metropolitan University Business School (UK),
and MERIT, University of Maastricht (the Netherlands)
*** Max Planck Institute of Economics, Evolutionary Economics Group, Jena (Germany)

2006/14   **May 2006**

Empirical Validation of Agent-Based Models:
A Critical Survey

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

Abstract

This paper addresses the problem of finding the appropriate method for conducting empirical validation in agent-based (AB) models, which is often regarded as the Achilles' heel of the AB approach to economic modelling. The paper has two objectives. First, to identify key issues facing AB economists engaged in empirical validation. Second, to critically appraise the extent to which alternative approaches deal with these issues. 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. This second set of issues is captured in a novel taxonomy, which takes into consideration the nature of the object under study, the goal of the analysis, the nature of the modelling assumptions, and the methodology of the analysis. Having identified the nature and causes of heterogeneity in empirical validation, 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. We also discuss a set of open questions within empirical validation. These include the trade-off between empirical support and tractability of findings, the issue of over-parameterisation, unconditional objects, counterfactuals, and the non-neutrality of data.

Keywords: empirical validation, agent-based models, calibration, history-friendly modelling.

[1] Faculty of Economics, University of Verona, and Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa, Italy, e-mail: giorgio.fagiolo@sssup.it

[2] Paul Windrum, Manchester Metropolitan University Business School, Manchester (UK) and MERIT, University of Maastricht (The Netherlands), e-mail: p.windrum@mmu.ac.uk

[3] Max Planck Institute of Economics, Evolutionary Economics Group, Jena, Germany, email: moneta@econ.mpg.de

1 Introduction

The aim of this paper is to provide a critical guide to the alternative approaches to empirical validation developed in agent-based (AB) economics in recent years, and contrasts these with neoclassical approaches to empirical validation. Particular attention is given to key issues that face all those engaged in empirical validation, and to how different methodologies deal with these key issues. The general themes considered are: (1) the relationship between theory and empirical research, (2) the relationship between models and the real-world system that is being modelled, and (3) the way in which alternative approaches to validation deal with (1) and (2).

The last two decades have seen a rapid growth in agent-based (AB) modelling in economics. 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. In fact, the interest in AB modelling[4] was stimulated by the rise of the new classical school that, in the late 1970s / early 1980s, became the dominant way of representing a dynamic, decentralised economy. AB models reject the aprioristic commitment of new classical models to individual hyper-rationality, continuous equilibrium, and representative agents[5]. 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 characterised as being in equilibrium or far-from-equilibrium.

AB models tend to contain the following ten features:

1. *A bottom-up perspective*. A satisfactory account of a decentralised economy is to be addressed using a bottom-up perspective because aggregate properties are the outcome of micro-dynamics involving basic entities (agents) (Tesfatsion, 2002). This contrasts with the top-down nature of traditional neoclassical models, where the bottom level typically comprises a representative individual and is constrained by strong consistency requirements associated with equilibrium and hyper-rationality.

---

[4] We do not aim to provide a complete survey of AB models in economics, nor to discuss the (often subtle) differences which characterise different research schools that have been employing AB models to study market and industry dynamics (e.g. evolutionary economics, agent-based computational economics, neo-Schumpeterian, and history-friendly models). The interested reader is referred to Lane (1993a,b), Dosi and Nelson (1994), Nelson (1995), Silverberg and Verspagen (1995), Tesfatsion (1997, 2001b, 2002), Windrum (2004), Dawid (2005), and Pyka and Fagiolo (2005). Also see Gilbert and Troitzsch (1999), and Wooldridge and Jennings (1995) for a discussion of AB techniques in other social sciences.

[5] An alternative view (though one which we doubt would be shared by AB economists themselves) is that the AB approach is *complementary* to neoclassical economics. Departures from standard neoclassical assumptions, found in AB models, can be interpreted as 'what if', instrumentalist explorations of the space of initial assumptions. For example, what happens if we do not suppose hyper-rationality on the part of individuals? What if agents decide on the basis of bounded rationality? and so on.

2. *Heterogeneity*. Agents are (or might be) heterogeneous in almost all their characteristics. This can range from initial endowments and other agents' properties, all the way through to behavioural rules, competencies, rationality, and computational skills.

3. *The evolving complex system (ECS) approach*. Agents live in complex systems that evolve through time (Kirman, 1997a, 1997b). Therefore, aggregate properties are thought to emerge out of repeated interactions among simple entities, rather than from the consistency requirements of rationality and equilibrium imposed by the modeller. This focus on interactions goes hand-in-hand with a bottom-up approach to systems modelling. Higher-level economic phenomena and structures, it is argued, are the product of subtle differences within a heterogeneous agent population and the local interactions that occur between individual agents. The higher level phenomena that emerge will differ as a consequence of the interactions that occur between the individual members, and subtle differences that exist within the heterogeneous population. This has important implications for methodological practice. One cannot, it is argued, deduce macro behaviour from the behaviour of an 'average' or 'representative' individual (Kirman, 1989, 1992; Forni and Lippi, 1997). For example, one cannot understand the collective rationality and operation of firms and consumers through the study of an individual firm or consumer. Equally, one is unable to derive collective rationality in economic phenomena, such as the formation of strategic alliances, through the construction of abstract representative agents. Rather, macro phenomena are a product of the interactions of heterogeneous firms and consumers. Complex dynamics of the macro structure arise, even if each individual follows a very simple behavioural rule. Hence, AB models highlight the heterogeneity of firms and consumers, the rules that govern their individual behaviour, and the rules governing their interactions (Lane, 1993a, 1993b).

4. *Non-linearity*. The interactions that occur in AB models are inherently non-linear (Silverberg *et al.*, 1988). Local interactions between individual agents are non-linear. Additionally, non-linear feedback loops exist between micro and macro levels. A model may contain a number of such feedback loops, some of which are positive and some negative. The focus is on the self-organizing properties that emerge through these feedback loops. As Silverberg *et al.* (1988) observe, in economics we see "complex interdependent dynamical systems unfolding in *historical*, i.e. *irreversible*, time, economic agents, who make decision today the correctness of which will only be revealed considerably later, are confronted with irreducible uncertainty and holistic interactions between each other and with aggregate variables" (Silverberg *et al.*, 1988, p.1036, italics in original). Thus, the emergence property is linked to the presence of multi-scale effects in AB models. Small-scale interactions produce a 'field' at the macro level that in turn influences and modifies activity at the local, micro level.

5. *Direct (endogenous) interactions*. Agents interact directly. The decisions undertaken today by an agent directly depends, through adaptive expectations, on the past choices made by other agents in the population (Fagiolo, 1998; Windrum and Birchenhall, 1998). These may contain structures, such as subgroups of agents or local networks. In such structures, members of the population are in some sense closer to certain individuals in the socio-economic space than others. These interaction structures may themselves endogenously change over time, since agents can strategically decide who to interact with according to the expected payoffs. When combined with heterogeneity and bounded rationality, it is likely that aggregation processes will be non-trivial, non-linear and, sometimes, will generate the emergence of structurally new objects (Lane, 1993a, 1993b).

6. *Bounded rationality*. The environment in which real-world economic agents live is too complex for hyper-rationality to be a viable simplifying assumption (Dosi *et al.*, 2005). It is suggested that one can, at most, impute to agents some local and partial (both in time and space) principles of

rationality, e.g. myopic optimisation rules. More generally, agents are assumed to behave as boundedly rational entities with adaptive expectations.

7. *The nature of learning*. Neoclassical models are interested in learning that leads to improvements in allocative efficiency. Two types of learning are investigated: inferential learning based on a Bayesian updating of decision rules where there is asymmetric or imperfect information, and action/strategy learning (notably in evolutionary games). In each case, learning is conducted within an equilibrium framework, the focus of the analysis is inter-temporal coordination and, where the problem arises, ways of dealing with multiple equilibria. Risk is probabilistic in these models. AB models engage in the open-ended search of dynamically changing environments. This is due to two factors. The first is the ongoing introduction of novelty and the generation of new patterns of behaviour, which are themselves a force for learning and adaptation. Agents are not initially endowed with an understanding of the underlying structure of the environment in which they operate but must develop a representation of the underlying structure. The introduction of radical innovation makes the task more difficult since the introduction of new objects alters this underlying structure and, hence, the payoffs associated with alternative actions. The second factor underpinning open-ended search is the complexity of the interactions between heterogeneous agents (5 above).

8. *'True' dynamics*. Partly as a consequence of adaptive expectations (i.e. agents observe the past and form expectations about the future on the basis of the past), AB models are characterised by true, non-reversible, dynamics: the state of the system evolves in a path-dependent manner (Marengo and Willinger, 1997).

9. *Endogenous and persistent novelty*. Socio-economic systems are inherently non-stationary. There is the ongoing introduction of novelty in economic systems and the generation of new patterns of behaviour, which are themselves a force for learning and adaptation. Hence, agents face 'true (Knightian) uncertainty' (see Knight, 1921) and are only able to partially form expectations on, for instance, technological outcomes. At the same time, firms deliberately introduce new objects (innovative products and processes) into their competitive environment in order to alter the underlying structure, i.e. to alter the payoffs associated with alternative actions, and even to change the rules of the game in their favour. Hence, new technologies are introduced into open-ended technological spaces, and payoffs to R&D are non-static and cannot be known *ex ante* (Nelson and Winter, 1982; Dosi, 1988). This not simply because agents face problems with respect to collecting and processing information, but the sheer algorithmic complexity of the problem they face, and their ability to define preferences over expected actions, events, and outcomes. As a consequence, agents face the extremely difficult task of learning and adapting themselves in turbulent and endogenously changing environments. On this basis, AB researchers have argued that assumptions of individual hyper-rationality coupled with rational expectations are inappropriate starting points for modelling.

10. *Selection-based market mechanisms*. Agents are typically selected against (Nelson and Winter, 1982). Most obviously, the goods and services produced by competing firms are selected by consumers. The selection criteria that are used may themselves be complex and span a number of dimensions. Turbulence in industry dynamics can be created through successive rounds of firm entry and exit (Saviotti, and Pyka, 2004; Windrum and Birchenhall, 2005; Windrum, 2005).

AB researchers have enjoyed significant success over the last 20 years. The models that have been developed indicate the viability and vitality of an alternative to mainstream neoclassical economics.

Indeed, mainstream economists have recognised the significance of the AB critique, 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.

While orthodox economists have taken on board, and responded to, aspects of the AB critique, they have not been moved to join the AB camp. There are many possible explanations for this, but an important aspect, recognised by AB modellers themselves, concerns the perceived lack of robustness in AB modelling. This threatens the AB research enterprise as a whole. Four key problem areas were identified in a recent conference and special workshop attended by the authors[6]. We discuss each of these in turn.

First, the neoclassical community has consistently developed a core set of theoretical models and applied these to a range of research areas. The AB community has not done this. Indeed, the sheer diversity of alternative AB models put forward over the last 20 years is striking. There is little or no understanding of the connection amongst the set of highly heterogeneous models that have been developed. Take, for example, the many types of models that have been put forward to explain technological diffusion. What are the deep, fundamental relationships between sand pile models, Polya urn models, dynamic learning models such as genetic algorithms, evolutionary games, and network models (e.g. small worlds models)? The simple answer is 'nobody knows'!

A second, related set of issues concerns a lack of comparability between the models that have been developed. Not only do the models have different theoretical content but they seek to explain strikingly different phenomena. Where they do seek to explain similar phenomena, little or no in-depth research has been undertaken to compare and evaluate their relative explanatory performance. Rather, models are viewed in isolation of one another, and validation involves examining the extent to which the output traces generated by a particular model approximates one or more 'stylised facts' drawn from empirical research. The problem is compounded by the high degrees of freedom in AB models. Not only do AB models contain highly non-linear, recursive interactions and feedbacks, but they tend to have many dimensions and, hence, degrees of freedom. With many degrees of freedom, almost any simulation output can be generated by an AB model. The problem is further compounded when the empirical phenomena that are specified as stylised facts are themselves rather general in nature. Under these conditions, comparisons with stylised facts not only represent a weak test for the validity of individual models, they fail to provide a strong methodological basis for comparing competing models.

This leads us to a third set of issues. These concern the lack of standard techniques for constructing and analysing AB models. It has been argued that developing a set of commonly accepted protocols for AB model building would benefit the profession (Leombruni, 2002; Richiardi, 2003). This would address, for instance, issues such as how and when sensitivity analysis (over the space of initial conditions and parameters) should be conducted, how one should deal with non-ergodicity in underlying stochastic processes, and how one should interpret, in terms of real-world time, the timing and lag structures that modellers typically build into their AB models.

---

[6] At a special session on 'Methodological Issues in Empirically-based Simulation Modelling', hosted by Windrum and Fagiolo at the 4th EMAEE conference, Utrecht, May 2005, and at the ACEPOL 2005 International Workshop on 'Agent-Based Models for Economic Policy Design', Bielefeld, July 2005.

The fourth, and final, set of issues concerns the problematic relationship between AB models and empirical data. As well as there being diversity with respect to the process of constructing AB models, fundamental differences exist in the ways researchers conduct empirical validation, and the methodological of AB researchers. 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?

Before moving on, we note that we have implicitly assumed that a high degree of heterogeneity is problematic. Heterogeneity makes it difficult to compare alternative models that seek to explain the same phenomenon, makes it difficult to advance a new, alternative paradigm, and to contrast it with the existing neoclassical paradigm. Having a small set of core models, developed by researchers over time using a (few) commonly accepted protocol(s) for model building and empirical validation would, it is suggested, be better for the community of AB researchers. It could, however, be argued that a high degree of heterogeneity is, initially at least, commensurate with Kuhn's (1962) discussion of the formation of new paradigms. Heterogeneity and flexibility are a consequence of a high degree of scientific debate and dissent on how best to proceed with the development of a new paradigm, and so is a necessary prerequisite for the emergence of a more codified AB paradigm. Further, it should be remembered that a degree of heterogeneity and dissent exists within the neoclassical community, and plays a necessary role in maintaining the vitality of that research paradigm. Still, while a degree of heterogeneity is beneficial, this must be balanced against the benefits of establishing a core set of models and modelling techniques that define a research paradigm, and are the focus of a practicing community that adheres to, and develops, a coherent research paradigm.

As previously stated, the focus of this paper focuses on the relationship between AB models and empirical data. The aim is to provide a critical overview of how AB modellers have been tackling the issue of empirical validation. As noted, 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. We shall draw attention to the generic nature of the empirical validation issues.[7] That is to say, neoclassical modellers face these issues just as much as AB modellers do. How they are dealt with may, of course, differ between the two communities but the issues are nevertheless the same.

The paper is structured as follows. Section 2 discusses the methodological basis of empirical validation, i.e. the comparison of discrete-time models with empirical data. The section begins with a discussion of the core issues of empirical validation faced by *all* modellers. Having identified these core issues, we consider how these were dealt with by the Haavelmo-Cowles approach. The discussion clearly identifies the rationale for empirical validation, and simultaneously highlights the methodological difficulties involved in using empirical validation to develop and asses the quality

---

[7] The very meaning of empirical validation is the subject of lively debate, see Bailey (1988).

of discrete-time models. This paves the way for a discussion of how, having broken the consensus of the Haavelmo-Cowles approach, a number of competing approaches have been developed by neoclassical economists.

Section 3 starts the discussion of how empirical validation is carried out in AB models. This is also conducted in very diverse ways. The heterogeneity of approaches, we suggest, is due to two factors. First, AB modellers face a whole class of problems associated with their modelling non-linearities, stochastic dynamics, non-trivial interactions among agents, feedbacks from the micro to the macro level, and so on. The second factor is the diverse structural content of AB models, and the very different ways in which AB models are analysed. We present a novel taxonomy of AB models which maps out the key areas in which AB researchers differ. These are the nature of the object under study (i.e. of the stylised fact(s) under analysis), the goal of the analysis (in-sample vs. out-of-sample), the nature of the main modelling assumptions (e.g. parameters, decision rules, interaction, and timing), and the methodology of the analysis (i.e. sensitivity analyses). Following this, section 4 surveys the three key approaches to empirical validation in AB economics. These are the indirect calibration approach (section 4.1), the Werker-Brenner approach to empirical calibration (section 4.2), and the history-friendly approach (section 4.3)[8]. Having highlighted the pros and cons of each approach, we conclude by discussing some outstanding issues for empirically-oriented AB modellers (section 5).

2   The methodological basis of empirical validation: comparing discrete-time models with empirical data

In the first part of this section (2.1), we discuss a set of core empirical validation issues that are common to all modellers (neoclassical and AB alike). As we shall see, despite neoclassical economists having a more codified methodology than AB researchers, a lack of consensus exists on how best to deal with the problems that arise from these issues. Interestingly, until the 1970s, there was a large consensus in econometrics on how to build and test an empirical model. This consensus was built around the Haavelmo-Cowles approach. We examine, in section 2.2, this approach, the Lucas and Sims critiques of the approach, and four alternative approaches to empirical validation developed by mainstream neoclassical economists as a consequence of the critiques. Having identified the core issues and analysed the responses of neoclassical economics, we are able to situate the responses of AB modellers (section 4).

2.1   Core issues of empirical validation

Let us consider the typical situation faced by any empirically-grounded economist attempting to replicate and/or explain a set of stylised facts. The point of departure is almost always a set of empirically observed data (e.g. panel data) whose generic form is:

---

[8] We will not express judgments about the performance of the different approaches, and will not enter into deep epistemological discussions about, for example, the meaning of 'cause and effect', 'reality', or 'models'. The interested reader is directed to Aruka (2001), Blaug (1980), Brenner and Murmann (2003), Cowan and Rizzo (1996), Downward *et al.* (2002), Feyerabend (1975), Hollis (1994), Mäki (1994), and McCloskey (1995) for introductory discussions of crucial methodological problems in social sciences.

$$(\underline{z})_i = \{ \, \underline{z}_{i,t}, \, t = t_0, \, \dots, \, t_1 \}, \, i \in I.$$

Here the set *I* refers to a population of agents (e.g. firms and households) whose behaviour has been observed across the finite set of time-periods $\{t_0, \dots, t_1\}$ and refers to a list of, say, *K* variables contained in the vector $\underline{z}$. Whenever agent-level observations are not available, the modeller has access to the *K*-vector of aggregate time-series:

$$\underline{Z} = \{ \, \underline{Z}_t, \, t = t_0, \, \dots, \, t_1 \},$$

which can be obtained by summing the *K* micro-economic variables $\underline{z}_{i,t}$ over $i \in I$. In both cases, the observed dataset(s) generate(s) a number of 'stylised facts' or statistical properties that the modeller is seeking to explain.

The datasets $(\underline{z})_i$ and/or $\underline{Z}$ are the unique outcome of an unknown, real-world data-generation process (*rwDGP*). Due to the impossibility of knowing the 'true' model of the world, we can think of the *rwDGP* as a very complicated, multi-parameter, stochastic process that governs the generation of a unique realisation which we can actually observe. The goal of the modeller is then to provide a sufficiently good 'approximation' of the *rwDGP* through a model. The model will contain a simplified data-generation process, the model-DGP (*mDGP*). This *mDGP* should provide a meaningful explanation of the causal mechanisms underlying the observed stylised facts, as well as a good representation of the observed data. Therefore, the empirical validation of a model is, in essence, a process of backward induction in which the modeller tries to evaluate the extent to which their *mDGP* is a good representation of the *rwDGP*[9].

As mentioned above, there are methodological issues associated with empirical validation that concern all the natural and social sciences, and which remain the focus of intense debate in the philosophy of science. We have, however, identified a set of core issues that relate to, and assist in explaining, the different approaches to empirical validation found in discrete-time-based economic modelling. Neither neoclassical economists nor AB modellers have precise and definite strategies to address these issues. But it is possible to assess important differences in the methodological strategies carried out by the two approaches. We will deal with this task in the subsequent sections. First we need to identify and elucidate the core issues.

1. *Concretisation vs. 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 have in the past expressed serious doubts about whether we can expect to have models that are fully concretised. [10] That is, models that contain all the possible variables that have an influence on the object under examination. Another way to look at the problem is via complexity theory. Consider the extreme case of randomness in the behaviour of some elements under investigation. Randomness, as defined by algorithmic complexity theory (associated with Kolmogorov and Chaitin), refers to the degree of

---

[9] One may notice that this account (or meta-model) of empirical validation reveals a realist bias since we assume there is a set of causal mechanisms 'out there' in the world which need to be explained, and explanation can be achieved through the construction of models. Note that an instrumentalist may also agree that there exists a set of causal mechanisms that generate the data we observe. What (s)he denies is that models are supposed to represent or describe such causal mechanisms.

[10] John Stuart Mill and John Maynard Keynes are the most cited examples. The reader is referred to Janssen (1994) for a clear discussion.

complexity of a string of binary elements in some well-defined computational procedure. Roughly speaking, the longer the computer program (or ideal Turing machine) needed to reproduce a given string, the more complex is that program. If the string is so complex that the length of the computational description approaches the length of the string, the string is deemed 'random' (see Mirowski, 1998). This demonstrates that, 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 (see Gibbard and Varian, 1978; Mäki, 1992 and 2005; Janssen, 1994). A series of open questions remain. 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 brings us to the second core issue of empirical validation.

2. *As-if assumptions.* As Mäki (2003) suggests, two kinds of *as if* assumptions can be found in economic modelling. One says 'phenomena behave *as if* forces that are isolated in a model are real', the other says 'phenomena behave *as if* certain ideal conditions were met: conditions under which only those real forces that are isolated in a model are active' (see Mäki 2005, p. 501). While the first position is instrumentalist, the second is consistent with realism. Realism, roughly speaking claims that theoretical entities 'exist in the reality', independent of the act of inquiry, representation or measurement.[11] 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.[12] 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 (1953), 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 .

3. *Methodological pluralism* vs. *strong apriorism*. 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. A great variety of 'pluralist' positions is conceivable. They range from the extreme cases of epistemological anarchism (Feyerabend, 1975) to scientific pluralist positions that allow the possibility of different methods of investigation but which have to be eventually integrated (Mitchell, 2003). Scientific pluralism opposes the strong apriorism that can be traced in many neoclassical approaches to empirical validation (as we will show in section 2.2). Apriorism is a commitment to a set of *a priori* assumptions. A certain degree of commitment to a set of priori assumptions are normal in science, and they correspond to what Lakatos (1970) 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

---

[11] Indeed several possible qualifications of realism are possible (see Mäki, 1998).
[12] The reader is referred to Moneta (2005) for an account of realist and anti-realist positions about causality in econometrics.

whenever a refutation is encountered. Lakatos (1970) dubbed the research programs involved with such positions as 'degenerating'.

4 . *Analytical solvability vs. descriptive accuracy*. We said in (1) that fully concretised models are impossible to build in a complex world. They are not useful either. Indeed, what is the point of modelling? Models serve two purposes in modern science: representing and intervening (Morrison and Morgan, 1999). In other words, models are useful, either for increasing our understanding of the world or of a theory, or because they provide the kind of information that allows us to intervene in the world, or both. As Nigel Gilbert (2004) points out, a model is like a map, and "a large scale map that shows every house is not much use for crossing a city" (Gilbert 2004, p. 403). There exists a basic 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.

5. *The identification / under-determination problem*. In the philosophy of science, there are various views on how one can give empirical support to a hypothesis or theoretical statement.[13] Regardless of which theory of confirmation one endorses, one must face 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 with respect to data.' In econometrics, the same observation has been formalised and labelled 'the identification problem.' As Haavelmo (1944) 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 identification problem does not arise. In section 2.2 we will scrutinise how the problem of identification has been addressed in the econometric methodology associated with the neoclassical paradigm.

6. *The Duhem-Quine thesis*. A second source of indeterminacy besets the relationship between theoretical statements and data. The Duhem-Quine thesis (developed in philosophy of science) observes that it is not possible to test and falsify a single hypothesis in isolation. This is because a hypothesis is inevitably tied to other hypotheses – the *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. As shown by Sawyer *et al*. (1997), 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 (2005), 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, Leamer 1978). Nonetheless, the Duhem-Quine thesis still undermines

---

[13] The main theories of confirmation can be divided in probabilistic theories of confirmation, in which evidence in favour of an hypothesis is evidence that increases its probability, and nonprobabilistic theories of confirmation, associated with the names of Popper, Fisher, Neyman and Pearson. The reader is referred to Howson (2000).

strong apriorist methodologies that do not check the robustness of the empirical results under variations of background assumptions.

2.2 Empirical Validation in Neoclassical Economics

Neoclassical economics has always been concerned with the relationship between theoretical models and data. Hence, issues (1), (2), (5), and (6), in particular, have been hotly debated by neoclassical economists. Particularly famous are the 'measurement-without-theory' debate between Koopmans and Vining (see Hendry and Morgan, 1995) and the discussion of as-if assumptions by Friedman (1953). Notwithstanding differences between individual neoclassical economists, a consensus on modelling and empirical validation existed from the 1930s to the 1970s. The consensus was built on the work of the Cowles Commission in the 1930s and on the work of Haavelmo (1944). The consensus broke down following the Lucas critique and the Sims critique. In 2.2.1 we explore the basis of the Haavelmo-Cowles approach and the Lucas and Sims critiques. In 2.2.2 we discuss the alternative methodological approaches that have been developed in post Haavelmo-Cowles.

2.2.1 The Haavelmo-Cowles Approach

The Haavelmo-Cowles approach suggests that the real-world DGP is constituted by causal relationships, which have three characteristics. First, they are stochastic, in the sense that each variable is determined by a composite of steadily acting (exact) plus random influences. The *rwDGP* can thus be represented by a model DGP, comprising a system of equations, in which the systematic factors are the dependent variables (regressors), and the disturbing elements are accounted for by the error terms. Such terms follow a joint probability distribution, which is well-defined if the system is properly specified. Second, causal relationships are autonomous, in the sense that the systematic relations will not break down when external conditions change. Third, causal relationships are simultaneous. Therefore, the economy as a whole can be characterised as a Walrasian general equilibrium system in which "everything depends upon everything else" (Haavelmo 1944, p. 22)[14].

A key issue addressed by Haavelmo- Cowles is the identification problem. The question faced by the Haavelmo-Cowles programme can be rephrased as following: how do we identify (among the possibly infinite theoretical models) the theoretical model whose DGP corresponds to the DGP of the real-world economic system? Haavelmo (1944) formalised the identification problem, showing the number of algebraic restrictions that one has to impose to an econometric model in order to identify it. The solution pursued by the Haavelmo-Cowles program was that such restrictions had to be derived from economic theory. The theory in consideration was Keynesian theory, filtered by a neoclassical synthesis that introduced the Walrasian notion of general equilibrium[15]. The object of the Haavelmo-Cowles program was more general, however. It did not explicitly specify which theory one had to obtain the restrictions. The crucial issue was that restrictions had to be derived from economic theory.

---

[14] Since simultaneity is simply the effect of aggregation, one-way causal relationships within the various sectors of the economy are not ruled out (Haavelmo 1944, p. 22).
[15] The neoclassical synthesis was formalised by Hicks (1937) and Modigliani (1944), and empirically tested within large-scale macro-econometric models developed by Klein (1964).

Once the model was identified, it could be estimated using rigorous statistical methods, and tested against empirical data. Thus, model validation and selection between competing theoretical models was not the starting point. Instead, it assumed that a generally accepted body of economic theory already existed. Given this, data was to be used to measure the strength of the causal given by that body of established theory. Still, Hoover (1994, 2006) points out that this methodology is open to alternative interpretations. Koopmans, in his debate with Vining about the possibility (denied by Koopmans) of 'measurement without theory', demonstrated what he calls a *strong apriorist* view (Hoover (1994). With strong apriorism one considers theory to exist prior to data, denies the possibility of interpreting data without theoretical presuppositions. Econometric models have to be built imposing restrictions derived from a well-articulated theory, which is accepted a priori. Hence, the purspose of econometrics is measurement, not validation or discovery. This view of theory-laden data is strictly connected to the Duhem-Quine thesis. Hoover (2006) observes that this strong aprioristic position "places the empiricist in a vicious circle: how do we obtain empirically justified theory if empirical observation can only take place on the supposition of a true background theory?".

The position of Haavelmo himself was very different to Koopmans. Although Haavelmo also maintained that empirical investigations should be founded on a priori theoretical restrictions, he favoured the statistical testing of theories in order to change and improve our a priori beliefs. Thus he endorsed a view of econometrics that Hoover (1994) calls *weak apriorism*. This recognises the need for an interplay of theoretical models and empirical data, and the possibility of empirically validating the apriori assumptions used in a model.

The Haavelmo-Cowles approach was subjected to two major critiques; the first by Lucas, the second by Sims. Let us consider each in turn. Lucas (1976) claims that the structural parameters, identified using the Haavelmo-Cowles approach, are not stable under a change of economic policy regime. In Haavelmo's terminology, the causal relationships measured by large-scale econometric models are not autonomous. This is because Keynesian models did not take into account the fact that people hold rational expectations, which lead them to change their behaviour as soon as the intervention takes place, in order to take advantage of the new policy regime. Consequently, large-scale econometric models using restrictions derived from Keynesian macroeconomics cannot be employed for policy evaluation. If the estimated models are used to implement economic policy actions in a systematic way, individual agents will adapt their behaviour to get the maximum advantage from the new policy. The estimated equations, originally used to predict the effects of the new policy, no longer hold (as a consequence of individuals changing their behaviour), and the policy will not have the intended effect.

It is important to recognise that the focus of the Lucas Critique was not the Haavelmo-Cowles understanding of the relationship between theory and modelling, but the Keynesian foundations that had been used to construct large-scale econometric models. Lucas called for the construction of models that used *micro-founded behavioural principles*. He believed that this would ensure structural relations that are invariant to changes in policy could be identified. These principles were the utility functions of representative consumers, the production functions of representative firms, and hyper-rational behaviours (and expectations). Further, these should be set within an equilibrium framework in which there is continuous short- and long-run market-clearing. This is the basis of the 'new classical macroeconomics' (NCM).

The critique put forward by Sims (1980) is more radical and is directed at the method of identification pursued by Haavelmo-Cowles. His critique is two-fold. First, the Haavelmo-Cowles

approach encourages the imposition of an extremely large number of restrictions. This is not done on the basis of sound economic theory, institutional or factual knowledge. It is a consequence of the need to have enough restrictions to secure identification. Second, the method for theoretical restrictions for identification is itself not well-grounded: economic relations cannot be identified in principle. The interdependencies are so numerous that each variable should appear in each equation. According to Sims, we should 'let the data speak' to see what can be learned from macroeconomic data without imposing theoretical restrictions (at least as far as the estimation step is concerned).

2.2.2 Beyond the Haavelmo-Cowles Approach

The impact of the Lucas and Sims critiques on neoclassical economics has been dramatic. They broke the consensus built around the Haavelmo-Cowles approach (Favero, 2001). In the fall-out, four competing econometric methodologies have been developed: (i) the structural modelling approach, (ii) vector autoregressive models (VARs), (iii) calibration, and the (iv) London School of Economics (LSE) approach. We shall now briefly discuss how each of these competing approaches addresses the key issues of empirical validation.

The NCM school was deeply concerned – at least in the beginning – with the relationship between theoretical data and models. Indeed, its origins lay in a search for an explanation to the empirical phenomena of stagflation experienced by western economies in the 1970s. NCM sought a 'sound theoretical and econometric' alternative to Keynesian macroeconometric models (Lucas and Sargent, 1979). The founders of NCM initially supplemented economic theory with the rational expectation hypothesis. Subsequently, Hansen and Sargent (1980) developed the 'structural modelling approach'. The structural model to be identified was invariably a representative agent model in which a single agent, characterised by a utility function (usually Cobb-Douglas), consumes one type of good and lives for an infinite time. All products and markets are perfectly competitive, and the agent has rational expectations. Using this theoretical framework, a set of identifiable equations is derived. These are estimated using the econometrics of simultaneous systems developed by Haavelmo-Cowles. This is particularly important. Methodologically, the Hansen-Sargent approach to empirical validation does not deviate from Haavelmo-Cowles - Hansen and Sargent put theory before data and are strongly aprioristic. The important practical (not methodological) difference is that NCM theory has replaced Keynesian theory as the theoretical basis of econometrics.

Sims's (1980) critique of Haavelmo-Cowles led him to develop the VAR approach as an alternative to structural modelling. This approach seeks to examine macroeconomic data without imposing restrictions. For this reason, the VAR approach solely uses unrestricted, reduced-form equations. Each variable is assumed to be endogenous, and is regressed on lagged values of itself, and on lagged values of all other variables. Sims (1982) argues that the VAR approach is immune to the Lucas critique because policy action involves the implementation of a fixed or slowly changing rule, with agents forming probability distributions over a range of possible policy stances. Exogenous shocks embody all surprises and innovations to the information set of economic agents. VAR models have been shown to be extremely efficient instruments for summarising the statistical properties of economic time series data. However, they cannot be used for policy evaluation. An estimated VAR is a reduced form model and, hence, cannot say anything about causal relations.

Methodologically, the VAR approach appears to be pluralist, even a-theoretical, with its desire to 'let the data speak for itself'. Certainly, the suggested use of *a priori* restrictions, independent of

theory, has been heavily criticised as arbitrary and advocating a program of a-theoretical empirical macroeconomics (see, for example, Cooley and Le Roy, 1985). However, as our discussion of the Duhem-Quine thesis highlighted (point 5 in section 2.1), an apparently 'theory free' model will nevertheless contain a set of related, auxiliary hypotheses that represent a set of interconnected and mutually supporting theories. Notably, VARs invariably contain basic assumptions of general equilibrium, perfect rationality, and perfect aggregation. So, while it may be that a set of theoretical restrictions are not imposed at the point of estimation, the collection of stylised facts through VARs is not an independent exercise uninformed by neoclassical theory.[16]

A third approach, the 'calibration approach', can be considered an attempt to solve the trade-off between *a priori* and a-theoretical approaches. It was originally developed as a method to empirically assess real business cycle (RBC) models (Kydland and Prescott, 1982). The approach appeals to the principle that a theory is better supported when validated on information that is not used in its formulation. According to this approach, a theoretical model does not need to fit the data according to criteria dictated by statistical theory,. Indeed, the model could easily be rejected since it is built on a very idealised set of assumptions. Consequently, the model does not take into account all possible contingencies related to the deep structure of the *rwDGP*. Knowledge of these contingencies is essential for policy. The presence of such disturbing factors, unaccounted for in the model but present in reality, would deform parameter estimates. Hence, the model needs to be calibrated. A model is calibrated when its parameters are not estimated in the context of its own model, but are picked from unrelated empirical micro-econometric investigations, and/or are chosen to guarantee that a simulated model matches some particular, unrelated features of historical data that are drawn from macroeconomic data (e.g. national accounts).

Once calibrated, the model is validated through simulation. A model is validated if its simulated outputs match the moments of the data. In real business cycles (RBC) models, the test of a model's ability to capture business-cycle behaviour is not its fit to some historical time series (which is only one of many possible realisations), but its ability to characterise the *distribution* from which a particular real-world realisation is drawn (Hoover 1995b). The model is further validated by looking at the stylised facts obtained by a VAR analysis of the macroeconomic time series. The model is assessed positively if it manages to reproduce, at least qualitatively, the impulse-response functions of the VAR (see Christiano *et al*., 1999).

Calibration appeals to the principle that a theory is better supported when it is validated using information that is not used in its formulation. This is a sound principle that offers a way of tackling problems associated with the Duhem-Quine thesis. But the acceptance of this principle, in the context of RBC, is not clear for for two reasons. First, as previously discussed, the collection of stylised facts through VARs is not an independent exercise. It is necessary to identify VARs that are capable of being given an economic interpretation. In order to identify a VAR, as the literature on the so-called *structural* VARs has demonstrated, one has to make appeal to some background knowledge, unless one uses arbitrary identification schemes like the Choleski-recursive scheme. The problem is that the background knowledge typically used to identify a VAR in the context of RBC is hardly independent of the basic assumptions of general equilibrium, perfect rationality, perfect aggregation and so on, which also constitute the tenets of the theoretical model to be calibrated. Second, all NCMs are based on the simplification of the representative agent. Thus, when NCMs are calibrated using parameters derived from microeconomic investigations, it is

---

[16] We will return to the relationship between stylised facts and theory in section 4.

tacitly assumed that aggregation does not fundamentally alter the structure of the aggregate model. This assumption is hard to defend, as Kirman (1992), and Forni and Lippi (1997) have shown.

A key characteristic of NCM calibration is the strong commitment general equilibrium, rational expectations, and perfect aggregation. These are taken for granted *a priori*. This form of *apriorism* is even stronger than Koopman's because it rules out likelihood-based statistical estimates of model parameters, which are standard in any version of the Haavelmo-Cowles methodology. This raises the question of how we are to judge competing calibrated models. More importantly, is there any possibility of a growth in knowledge if the *hard core* of NCM theory is immune to revision, thanks to the protection of the calibration methodology?[17] The issue is illustrated by the discrepancy that exists between NCM models designed for forecasting, and NCM models designed for policy evaluation. There is the widespread use of a-theoretical models such as VARs for economic forecasting, on the one hand, and structural (or theoretical) calibrated models for policy evaluation, on the other. The separation of economic forecasting from economic policy analysis "can be justified theoretically by realising that the forecasts arising from a structural model are just a function of current and past data; if this function can be estimated consistently without reference to the underlying theoretical models, than the resulting forecasts will have the same forecast error variance, to first-order asymptotically, as if this function were known" (Stock, 2001, p.31). The problem that arises from this discrepancy is that theoretical models built for policy analysis avoid forecasting and, hence, are not open to empirical falsification. Hoover observes that no guidance can be given on "how to proceed nor leeway to adjust assumptions, if the data are widely at variance with the model's predictions. The new classical organizing principles are uncontradictable" (Hoover, 1994, pp.71-72).

We also note that the lack of a robust empirical validation of the theoretical apparatus, coupled with the desire to generate sharp analytical implications (see point 4 in section 2.2), often makes it extremely difficult to construct models that reproduce and explain more than one stylised fact at the time. It has been suggested that the number of over-simplifying, ad-hoc assumptions required to derive an analytically solvable implication (that is in line with the observed phenomena) increases enormously according to the number of phenomena one would like to explain simultaneously.[18] From labour market dynamics, to growth and development, to consumption and demand, there are simple neoclassical models that address just one phenomenon (or stylised fact) in isolation. We do not see robust models that explain a number of related phenomena (Fagiolo *et. al*, 2004a). This is true, not only in terms of the macro-dynamics, but (more dramatically) at the micro level. Consider, for instance, the lack of micro-macro models that jointly replicate micro stylised facts such as firm size, growth and productivity distributions across sectors, together with macro stylised facts concerning statistical properties of aggregated growth time-series.

The fourth approach developed since the demise of Haavelmo-Cowles is the LSE econometric methodology. It is strongly at odds with the calibration methodology. As noted, that approach concurs with the new classical line that the key problem of Haavelmo-Cowles was that it was based on Keynesian theory. By replacing Keynesian theory with new classical theory, one now has as basis for secure theoretical knowledge. The LSE methodology rests on a very different view. It explains the failure of Haavelmo-Cowles as a lack of rigorous statistical testing. The modelling strategy relies on a sequential reduction-procedure: the so-called 'general-to-specific approach' (see Cook and Hendry, 1994). The objective of this approach is the identification of a characterisation of

---

[17] See point 3 among the core issues of empirical validation discussed in section 2.1.
[18] See Fagiolo *et al*., 2004a for a discussion and for further reading.

the data that is simpler than the *rwDGP*, but which still retains all the important information on the phenomena under study.

The concept of exogeneity, refined and distinguished into different categories by Engle, Hendry and Richard (1983), plays also a crucial role in the LSE methodology. Indeed, the dimensionality of the model is reduced by omitting equations for those variables for which the null hypothesis of exogeneity is not rejected. Furthermore, the LSE methodology stresses the importance of testing alternative specifications. The theory of encompassing (Hendry, 1988) serves this scope. Given any two competing specifications, which can be nested in a joint model, "[i]f one of the specifications is a valid restriction of the joint model and the other not, then the one encompasses the other" (Hoover 2006). In the sequential reduction, at each step encompassing of a parsimonious specification is tested against a set of alternative specifications.

The LSE methodology endorses a relationship between models and data that is far removed from the apriorism of Koopmans' interpretation of Haavelmo-Cowles and, in particular, the calibration approach. Theory still maintains its important role of defining the range of variables that are to be included in the *mDGP*, and of suggesting the restrictions that are to be rigorously tested. In some ways, the LSE methodology seeks a middle ground between the structural modelling of Haavelmo-Cowles and the a-theoretical VAR approach (Hoover 1995a).

We learn four lessons from this brief excursion. First, there is no consensus within neoclassical economics on the most appropriate methodology for empirical validation, and the debate is still very much open. The Haavelmo-Cowles methodology once provided a basis for consensus and offered a practical means of dealing with aspects of the identification and under-determination problems. However, following the Lucas and Sims critiques, consensus broke down and a number of competing approaches to empirical validation now exist. Second, each of the alternative approaches currently pursued in by neoclassical economists contains a very different methodological position. Notably, the strongly *aprioristic* calibration approach stands at one end of the methodological spectrum, while the pluralistic LSE approach stands at the other end. Of all four approaches, only the LSE approach is methodologically pluralist, allowing for the comparison of competing theoretical models. The other approaches are, to varying degrees, aprioristic and clearly take neoclassical economics as the sole theoretical basis for economic modelling. Third, the Duhem-Quine thesis undermines econometric methodologies that do not check the robustness of the empirical results under variations of background assumptions. Fourth, the basic tenets of NCMs – continuous market clearing, representative agent, strong rationality – are never subjected to empirical validation. This is particularly noticeable in the calibration approach.

3 Empirical Validation and Heterogeneity of AB Models

As in neoclassical economics, empirical validation is carried out in very diverse ways by AB economists. In this and in the next section we examine this heterogeneity. We suggest that this heterogeneity partly depends on the fact that AB models frequently contain non-linearities, stochastic dynamics, non-trivial interaction structures among economic agents, and micro-macro feedbacks (section 1 above). The heterogeneity is also linked to the diverse structural content of the models, and the ways they are analysed. This is another key difference to neoclassical modelling, where variety is intimately associated with the diverse roles played by statistical inference. In order to capture the diversity of structural content (and methods of analysis) within AB economics, we present a taxonomy that contains four dimensions: the nature of the object(s) under study, the goal

of the analysis, and the modelling assumptions. In the remainder of this section we discuss a set of empirical validation issues that are particularly pressing for AB modellers. These are the alternative strategies for constructing empirically-based models, over-parameterisation, defining sufficiently strong tests for empirical validation (e.g. stylised facts, and unconditional objects), issues arising from the availability, quality and bias of available datasets, and the usefulness and implications of counterfactuals for policy guidance.

Let us go return to the meta-model *rwDGP* vs. *mDGP* introduced at the beginning of section 2.1. The extent to which the *mGDP* accurately represents the *rwDGP* depends on many preliminary, model-related factors (see Fig. 1). These range from the quality of micro and macro parameters that are specified, to initial conditions at the micro and macro level that are taken to proxy real-world initial conditions. The problem of developing a good representation is compounded when discrete-time models contain (as invariably AB models do):

    (i)   non-linearities and randomness in individual behaviours and interaction networks;

    (ii)   micro and macro variables that are governed by complicated stochastic processes that can hardly be analysed analytically (hence the need for computer simulation);

    (iii)   feedbacks between the micro and macro levels.

Using Figure 1, let us consider one possible procedure for studying the output of an AB model. Suppose the modeller knows (from a preliminary simulation study, or from some *ex ante* knowledge coming from the particular structure of the AB model under study) that the real-world system is ergodic, and that the *rwDGP* displays a sufficiently stationary behaviour for a time period after T^* for (almost all) points of the parameter space and initial conditions. For a particular set of initial conditions, micro and macro parameters (i.e. θ, $\underline{\Theta}$, x_0, and $\underline{X}_0$), we assume the *rwDGP* runs until it reaches some form of stable behaviour (for at least $T>T^*$ time steps).

Now suppose we are interested in a set of statistics $S = \{s_1, s_2, \dots\}$ that are to be computed on the simulated data generated by the *mDGP* $\{x_t, t=1,\dots, T\}$ and $\{\underline{X}_t, t=1,\dots, T\}$[19]. For any given run ($m=1,2,\dots,M$), the simulation will output a value for s_j. Given the stochastic nature of the process, each run − and thus each value of s_j − will be different from the others. Therefore, after having produced M independent runs, one has generated a distribution for s_j containing M observations. This distribution can be summarised by computing, for example, its mean $E(s_j)$, its variance $V(s_j)$, and so on. Recall, however, that the moments will depend on the initial choices that were made for θ, $\underline{\Theta}$, x_0, and $\underline{X}_0$. By exploring a sufficiently large number of points in the space of initial conditions and parameter values, and by computing $E(s_j)$, $V(s_j)$, etc. at each point, one can gain a deep understanding of the behaviour of the *mDGP* of the model system[20].

---

[19] For example, one of the micro variables might be an individual firm's output and the corresponding macro variable may be GNP. In this case, we may be interested in aggregate statistics s_j such as the average rate of growth of the economy over T time-steps (e.g. quarterly years).

[20] Consider the example of footnote 13 once again. One may plot $E(s_j)$, which is the Monte-Carlo mean of an economy's average growth rates, against key macro parameters such as the aggregate propensity to invest in R&D. This may allow one to understand whether the overall performance of the economy increases in the model with that propensity. Moreover, non-parametric statistical tests can be conducted to see if $E(s_j)$ differs significantly in two extreme cases, such as high vs. low propensity to invest in R&D.
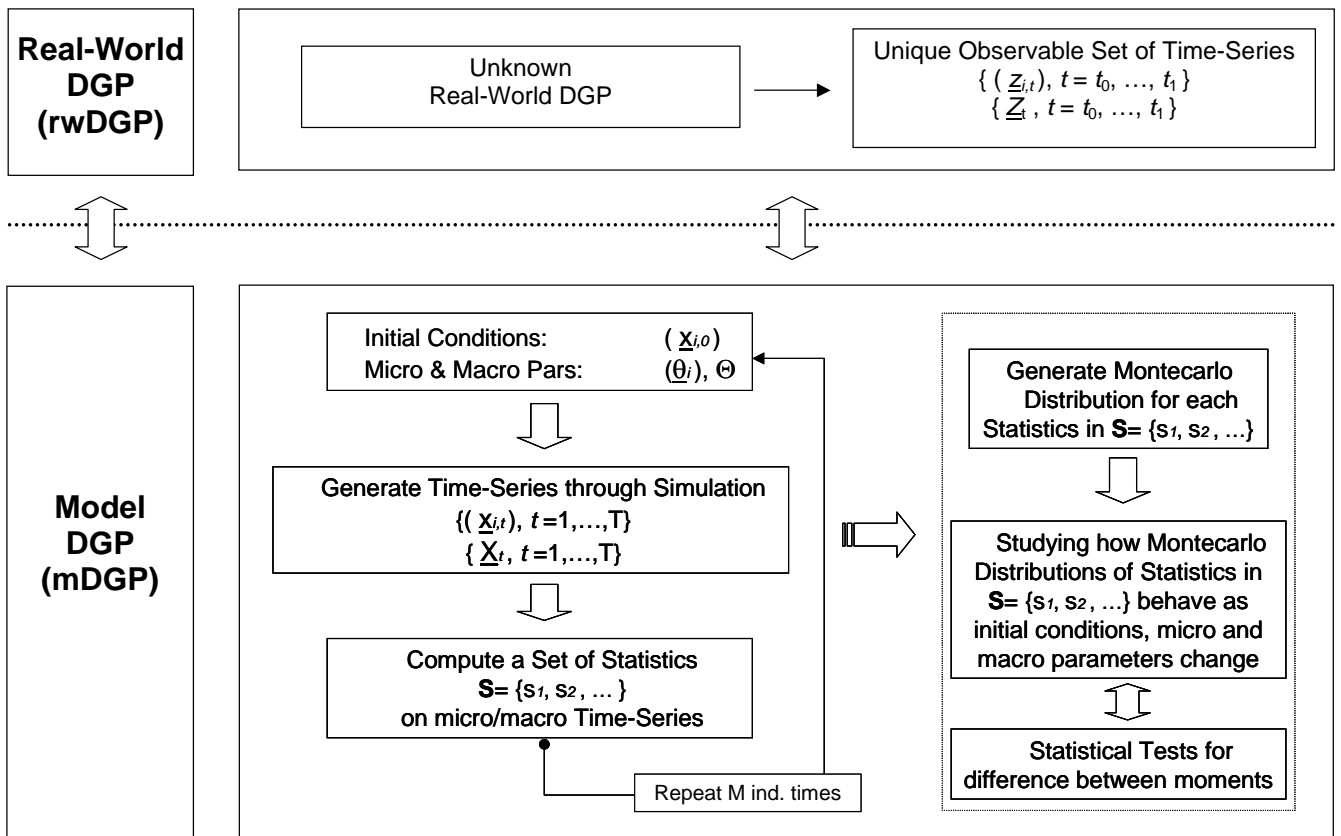
Real-World DGP (rwDGP)

| Unknown Real-World DGP | → | Unique Observable Set of Time-Series $\{ (\underline{z}_{i,t}),\ t=t_0, ..., t_1 \}$ $\{ \underline{Z}_t,\ t=t_0, ..., t_1 \}$ |

Model DGP (mDGP)

Initial Conditions: $(\underline{x}_{i,0})$
Micro & Macro Pars: $(\underline{\theta}_i), \Theta$

⇩

Generate Time-Series through Simulation
$\{ (\underline{x}_{i,t}),\ t=1,...,T \}$
$\{ \underline{X}_t,\ t=1,...,T \}$

⇩

Compute a Set of Statistics
$\mathbf{S}= \{s_1, s_2, ... \}$
on micro/macro Time-Series

Repeat M ind. times

⇒

Generate Montecarlo Distribution for each Statistics in $\mathbf{S}= \{s_1, s_2, ...\}$

⇩

Studying how Montecarlo Distributions of Statistics in $\mathbf{S}= \{s_1, s_2, ...\}$ behave as initial conditions, micro and macro parameters change

⇕

Statistical Tests for difference between moments

Figure 1. A procedure for studying the output of an AB model

In any case, it is easy to see that the comparison of *mDGP* and *rwDGP* in AB models opens up a whole range of new methodological problems (we will consider these in the next sections). The bottom line is that AB modellers have in practice performed empirical validation in (sometimes very) diverse ways. This, we suggest, is also due to the fact that the ways in which AB models have been traditionally analysed (i.e. the statistical and simulation procedures that were employed) are themselves very different. In order to show this, Table 1 presents a taxonomy of AB models along four key dimensions[21].

The *first* dimension is the *nature of the object(s) under study* (first column of Table 1). These are the stylised facts (empirically observed facts) that the model is seeking to explain. Significant differences exist with respect to the nature of the object 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

---

[21] Space constraints prevent us from discussing how different classes of AB models (e.g. evolutionary industry and growth models, history-friendly models, and ACE models) fit each single field of the entries in Table 1. See Windrum (2004), and Dawid (2005) for detailed discussions of this topic.

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 technology's 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* (second column of Table 1). AB models tend to deal with in-sample data. In-sample data are relevant when one is interested in describing or replicating observed phenomena. Out-of-sample exercises are less frequently carried out by AB economists. For instance, there are no examples of AB models, dealing with technological change and business cycles, that attempt to provide predictions of the out-of-sample behaviour of GDPs over given periods of time (e.g., with the goal of answering control-related issues). Only recently have AB models have been employed to generate policy implications and to address issues related to market-design (see Marks, 2005; Koesrindartoto *et al.*, 2005)[22].

A *third* dimension concerns the nature of the most important *modelling assumptions* (Table 1, third column[23]). 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 $\mathfrak{R}_{i,t}$, and interaction structures (i.e. the graph G_t) may be exogenously fixed (e.g. $\mathfrak{R}_{i,t}=\mathfrak{R}_i$ and/or $G_t=G$). 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 happens in endogenous network formation models, see Fagiolo *et al.*, 2004b).

The *fourth* and final dimension is the *methodology of analysis* (Table 1, fourth column). In order to thoroughly assess the properties of an AB model, the researcher needs to perform a detailed sensitivity analysis, along the lines sketched in Figure 1. 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). To illustrate, let us suppose the researcher is interested in the growth rate of an economy. Once the initial conditions (e.g. initial capital levels of firms) and the micro-macro parameters are fixed (e.g. the individual propensity to engage in R&D and the macroeconomic availability of technological opportunity), a run of the AB model will generate a single instance of the growth rate. Now, suppose one wants to compare the growth rate of an economy when there are high and low technological opportunities. It is not enough to generate a single simulation run, or to check the results for a single choice of micro parameters (i.e. the individual propensity to engage in R&D) and initial conditions (i.e. firm capital levels). On the contrary, a thorough exploration needs to be made of how the average macro growth rate depends on each of the elements (i)-(iii). More generally, sensitivity analysis entails a careful investigation of how the outputs of a model vary when one alters its inputs (Law and Kelton, 1991; Leombruni *et al.*, 2006). Therefore, apart from

---

[22] Also see papers presented at the recent conference 'Agent-Based Models for Economic Policy Design' (ACEPOL05), Bielefeld, June 30, 2005 - July, 2, 2005 (http://www.wiwi.uni-bielefeld.de/~dawid/acepol/).
[23] The pros and cons of this heterogeneity in modelling assumptions for AB economists were discussed in section 1 of this paper. Also see Richiardi (2003), Pyka and Fagiolo (2005), and Leombruni *et al.* (2005).

sampling the space of parameters and initial conditions, researchers need to check the robustness of the results against changes in (i) the distribution of random variables generating noise in the system, (ii) timing and updating mechanisms, and (iii) level of aggregation of microeconomic variables.

Nature of Object under Study	Goal of Analysis	Modelling Assumptions	Methodology of Analysis
• Qualitative-Quantitative Analysis • Single – Multiple Variables • Transients vs. Long-run Impacts • Micro – Macro	• In-Sample (descriptive) • Out-of-Sample ○ forecasting ○ prediction / control ○ policy implications	• Size of the space of ○ micro/macro parameters ○ micro/macro variables ○ decision rules • Treatment of time/updating ○ discrete / discontinuous ○ parallel / asynchronous • Type of Decision rules ○ adaptive, myopic vs. optimising, best-reply ○ deterministic vs. stochastic • Type of Interaction Structure ○ local vs. global ○ deterministic vs. stochastic • Dynamics of Decision Rules and Interaction Structures ○ exogenously given/changing ○ endogenously selected	• Sensitivity of results ○ to micro/macro parameters ○ to initial conditions ○ ergodicity ○ to across-run variability

Table 1. Taxonomy of dimensions of heterogeneity in AB models

These four key dimensions of heterogeneity strongly inform the choice of empirical validation procedure that is used. The focus on qualitative or quantitative phenomena, on micro or macro phenomena, and on transients or long-run impacts, determine the type of data that is required for empirical validation, the statistical procedures to be followed, and the ability to generate empirically testable implications. Additionally, the extent to which sensitivity analysis is performed prior to empirical validation has important implications for the universality of the simulation results that are obtained. Whether the analysis is a descriptive (in-sample) exercise, or seeks to generate (out-of-sample) predictions, also necessitates different approaches to data collection and analysis. Out-of-

sample analysis requires the researcher to calibrate parameters and initial conditions. As discussed, this should be governed by empirical evidence (where it is available).

4 Alternative Approaches to Empirical Validation in AB Models

The previous section outlined the key dimensions along which one finds heterogeneity in AB models. In this section we explain, in greater detail, the nature of this heterogeneity. We review three of the most influential approaches to empirical validation that have been developed in the AB literature. These are indirect calibration approach, the Werker-Brenner approach, and the history-friendly approach. These three approaches are strongly shaping the debate on validation in AB economics. Our goal is to provide a general overview of each approach, and to consider how each of them tackles the methodological issues discussed in sections 2 and 3. Each, in its own way, seeks to reduce the parameters of the model, although each does this in a different way. All seek to reduce the space of possible 'worlds' that can be explored by tying AB models to an observed empirical reality. The history-friendly approach constrains parameters, interactions, and decision rules in the model in line with the specific, empirically-observable history of a particular industry. It can be interpreted as a calibration exercise with respect to unique historical traces. The other two approaches do not impose a preliminary set of restrictions on parameters but, rather, indirectly employ empirical evidence to identify sub-regions in the potential parameter space. Within these sub-regions, a model is expected to replicate some relevant statistical regularities or stylised facts.

Prior to this discussion, we need to briefly consider qualitative AB models and their contribution to the validation debate. There exist a significant number of AB models that engage in purely qualitative theorising, and which are not validated in any meaningful sense. In economics there is a long tradition, stretching back to the earliest classical economists, of using models as a means to engage in abstract gedankenexperiments. Many AB models do this. In accordance with our thesis on the relationship between model content and empirical validation, a significant number of AB models seek to explain qualitative phenomena that are intrinsically closed to quantitative analysis. There is no rationale for testing such models against existing empirical data sets. Notable examples are evolutionary game-theoretic models (Vega-Redondo, 1996), and Polya urn models (Arthur, 1988, 1994). There is a weak relationship between the micro-macro variables/parameters of these models and their empirically observed counterparts. The focus of such models is the emergence of qualitative aggregate patterns, such as the emergence of coordination and cooperation. Forecasting exercises are possible but they typically generate unpredictability results. For example, one knows with certainty that users will lock in to one of the competing technologies in Arthur's (1994) Polya urn model but it is impossible to know *ex ante* which of the competing technologies will be selected.

Some AB economists, engaged in qualitative modelling, are critical of the suggestion that meaningful empirical validation is possible. They suggest there are inherent difficulties in trying to develop an empirically-based social science that is akin to the natural sciences. Socio-economic systems, it is argued, are inherently open-ended, interdependent and subject to structural change. How can one then hope to effectively isolate a specific 'sphere of reality', specify of all relations of phenomena within the sphere with the external environment, and build a model describing all important phenomena observed within the sphere (together with all essential influences of the external environment)? In the face of such difficulties, some AB modellers do not believe it is possible to represent the social context as vectors of quantitative variables with stable dimensions (Valente, 2005).

One possible reaction is to use the computer as an artificial laboratory in which basic, causal relationships can be tested in order to gain some knowledge on the underlying (much more intricate and convoluted) real-world causal structure. The danger of this strategy is that one ends up building auto-referential formalisations that have no link to reality (Edmonds and Moss, 2005). Certainly there are those in other social science disciplines who have taken the step of accepting they are constructing and analysing synthetic artificial worlds which may or may not have a link with the world we observe (Doran, 1997). Those taking this position open themselves to the proposition that a model should be judged by the criteria that are used in mathematics: i.e. precision, importance, soundness and generality. This is hardly the case with AB models! The majority of AB modellers do not go down this particular path. Instead, they employ methodological approaches that seek to deal with the difficult issues and problems discussed in sections 2 and 3.

4.1 The Indirect Calibration Approach

Advocates of qualitative simulation warn us about the problems that arise from the inherent structural non-stationarity and the interdependence of socio-economic systems. The points echo Kaldor's discussion of stylised facts (Kaldor, 1961; Kwásnicki, 1998). In his view, the study of socio-economic phenomena is strongly hampered by difficulties associated with collecting large, detailed datasets that are sufficiently reliable for us to empirical validate our models. He suggested that empirical validation should be conducted on a more aggregate level, by focusing on stylised facts or statistical regularities. By stressing the reproduction (explanation) or prediction of a set of stylised facts, one hopes to circumvent problems of data availability and reliability.

Drawing upon a combination of stylised facts and empirical datasets, many AB modellers have been developing a pragmatic four-step approach to empirical validation. 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.

The foregoing procedure is an exercise in 'indirect calibration'. Suppose, for example, that the Beveridge curve is one of the statistical regularities being investigated. The model must be able to replicate a relationship in which unemployment rates decrease with vacancy rates in the labour market (cf. Fagiolo *et al*., 2004a). The researcher should restrict the further analysis to all (and only) parameter combinations under which the model does not reject that hypothesis (at some confidence level). This step is the most sensible because it involves a fine sampling of the parameter space. It is also computationally demanding and requires the use of Monte Carlo techniques. For any given point in the parameter space, one must generate a distribution for the statistics summarising the stylised facts of interest (e.g. the slope of the relationship between unemployment and vacancy

rate), and test the null hypothesis that the empirically observed valued can be generated by our model under that particular parameter combination (see Figure 1).

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 the stylised facts of interest) which the model can validate *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. For example, one might study how the absolute value of the Monte Carlo average of the slope of the unemployment-vacancy rate relation varies with some macro-parameter (if any) that governs wage setting and/or union power in the model. This can shed light on the causal mechanism underlying the emergence of a Beveridge curve. Similarly, one can ask whether business cycle properties (e.g. average and volatility of growth rates) change with the slope of the Beveridge relation. If this is the case, a fresh implication generated by the model (under empirically plausible parameters) can be taken to the data - and further provide support for the AB model under scrutiny.

A stream of more recent AB contributions to the fields of industry- and market-dynamics has been strongly rooted in the four-step empirical validation procedure just presented. For example, Fagiolo and Dosi (2003) 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*. (2004a) 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.

Although appealing, the indirect calibration approach is open to criticism in at least two important respects. First, notice that no attempt is made to calibrate micro and macro parameters using their empirical counterparts. There are two reasons for this. On the one hand, the models address in-sample exercises. On the other hand, due to the difficulties of matching theoretical and empirical observations, one is bounded to be as agnostic as possible as to whether the details of a model (variables, parameters) can be really compared with empirically-observable ones. However, in order for this indirect, weak, calibration procedure to be effective, the empirical phenomena of interest should not be very general. Otherwise, they might not necessarily represent a difficult test for the model. If this is the case, the model might pass the validation procedure without providing any effective explanation of the phenomena of interest (e.g. no restrictions on the parameter space would be made). This parallels Brock's discussion of 'unconditional objects'. Here the fundamental issue of discriminating between the 'descriptions' and 'explanations' of reality pops up once more.

The second problem is far subtler, and has to do with the interpretation of the points belonging to the sub-region of the parameter space (and initial conditions) that resist the sort of 'exercise in plausibility' that one performs in the third step of the procedure. After a suitable sub-region of the parameter space (and initial conditions) has been singled out - according to the capability of the model to replicate the set of stylised facts of interests in that sub-region - how should one interpret all comparative exercises that aim at understanding what happens when one tunes the parameters within that sub-region? For instance, suppose that one has found the range of values for the parameter 'strength of union power' which is consistent with the emergence of a Beveridge curve in the labour market, and is interested in asking the question: How does the average growth rate of the economy change when the strength of union power moves within those bounds? In all these cases, an interpretation problem arises: What does it really add - to our understanding of reality -

providing an answer to this type of questions? How can one interpret alternative parameter values in an evolutionary world where history, indeterminacy, and non-linear feedbacks between the micro and macro levels may strongly affect the outcomes?

4.2 The Werker-Brenner Approach to Empirical Calibration

An interesting application of empirical calibration to AB models has been proposed by Brenner and co-authors. The approach is discussed in Werker and Brenner (2004), and has been applied in Brenner and Murmann (2003), and Brenner (2004). The origins of calibration lie in Kaldor's work in the 1960s. In a nutshell, the calibration approach proposes that when (as is usually the case), theory does not enable us to restrict the dimension space of initial conditions and micro/macro parameters, we should use empirical knowledge to first calibrate the dimension space, and thereafter use empirical knowledge to validate the resulting model specification. In this way, calibration uses empirical evidence to identify sub-regions of the parameter space where the model is able better to replicate empirically observed, statistical regularities. Where high quality datasets do not exist, Kaldor advocated the use of stylised facts to pin down the key parameter values or ranges of values for key parameters.

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 we have seen, 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. As discussed in section 2, we can 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 (Werker and Brenner, 2004). In practice, this involves a further validation exercise for all empirical realisations that can be collected. Here, though, 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)" (Werker and Brenner, 2004, p.13). 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" (Werker and Brenner, 2004, p.13).

The Werker-Brenner approach is attractive in a number of respects. First, it addresses head-on many of the issues of model evaluation: it offers a means of reducing the degree of freedom in models, it advocates testing procedures for sensitivity analysis on large numbers of simulations, and it proposes the application of well-established Bayesian inference procedures for establishing rigorous empirical tests for both model parameters and outputs. It also avoids a number of the potential pitfalls associated with developing models based on single histories (like in the history-friendly approach). Second, it appears to offer a powerful methodology for developing rigorous, empirically-grounded simulation models; models that explicitly take into account competing theories and assumptions.

As with all approaches, there are strengths and weaknesses to empirical calibration. Let us consider some important methodological and operational issues associated with calibration. These, we hasten add, are not specific to Werker-Brenner, but are generic to all calibration approaches. To start with, there is a strong tendency for calibration to influence the types of models that we develop. There is a tendency to build models with ingredients that are readily calibrated (Chattoe, 2002). This is because empirically calibrated models force the researcher to focus on variables and parameters that can be measured in reality. Yet, there exist many. There is tendency to focus on dimensions that are readily calibrated and, at the same time, to exclude dimensions that are not easily calibrated. This has important implications for the micro-parameters of a model, particularly the behavioural aspects of individual agents and interactions between agents. For instance, the mental models that agents use to make their decisions are an important component in many AB economics models, but the mental models used by real world agents tend to be unobservable in practice, and so not open to empirical calibration. The net result is that calibration tends to induce the modeller to either abstract from the micro features of the economy, or to force calibration of those parameters employing unreliable or inconsistent data. Calibration also affects the types of model outputs that one focuses upon. Again, there is a temptation to focus on outputs that are readily measured, and not to consider phenomena that cannot be measured or calibrated a priori. This might end up in models that are only able to replicate the reality, but unable to predict or give insight into the search for novel phenomena.

Another issue is the availability of quality of available empirical data. Effective calibration requires a wealth of high quality data. Indeed, the Werker-Brenner calibration approach is particularly demanding as it requires the modeller to engage in two rounds of empirical validation. Unfortunately, in economics (and in the other social sciences, for that matter) empirical data are always scarce while the capacity of economists to generate new theories is almost infinite. There are

three reasons for the lack of extensive data. The first reason is the cost of organising and collecting raw data, processing this raw data, and constructing organised datasets. The organisation and construction of national and international datasets depends on the existence of specialist statistical offices. Even then, high costs mean that national and international statistical offices are not able to collect data on all matters.

A second reason for limited datasets is the bias that is inherent to the collection process. People who collect data are informed by existing theory on the data to be collected. Established theory not only informs choices about which variables to measure (and not to measure), but also how to measure the selected variables - i.e. the key indicators to be used. Hence, there is an inherent tendency to maintain the status quo and to lock out new, alternative concepts, theories and models. For new ideas to succeed in economics, new datasets may be required in order to test new theories and models. Perhaps the best-known example of this is the Keynesian revolution. Theoretical developments went hand-in-hand with the collection of aggregate national data for the first time, notably on household consumption expenditure and firm investment expenditure. We shall return to issues of bias and incompleteness of existing datasets in a moment, for now we observe that this is highly problematic for calibration. Furthermore, the most common reason for under-determination in economics is the bias and incompleteness the available datasets. It is not always possible to exclude a particular model on the basis of existing empirical data because other types of data can potentially support the model, if they had been collected.

A third reason for incomplete data may lie in the nature of the phenomenon that is being observed. The phenomena may be rarely observed, or is a unique event that is non-reproducible. The issue is widely discussed in statistics texts. The practical upshot is that, while in principle we could generate as many theoretical observations as we like, in practice we may only have a few of such empirical realisations (possibly only one!). If we believe that the empirical observations come from an underlying DGP that could have been 'played twice' (i.e. could have generated alternative observations, other than the one we have) the problem of comparing simulated with empirical data becomes very complicated. We will return to this issue in section 4.3.

A further generic issue highlighted by calibration is the nature of the relationship between the model *mDGP* and the real-world *rwDGP*. First, there is the question of whether the *rwDGP* is ergodic or non-ergodic. If the underlying real-world *rwDGP* is thought to be non-ergodic (as well as the theoretical *mDGP* described in the AB model), then initial conditions matter. This raises a whole host of problems for the modeller. The modeller needs to identify the 'true' set of initial conditions in the empirical data, generated by the *rwDGP*, in order to correctly set the initial parameters of the model. Even if perfect data exists (which is unlikely), this is a very difficult task. How far in the past does one need to in order to identify the correct set initial values for the relevant micro and macro variables? There is a possibility of infinite regress. If this is the case, then one may need data stretching back a very long time, possibly before data started to be collected.

Even when the *mDGP* and *rwDGP* are thought to be (sufficiently) stationary processes, the problem of correctly setting t_0 remains. An important decision to make is about the particular sub-sample of simulated data (of length $\tau = t_n - t_0$) that is to be compared with the empirical data. The underlying *rwDGP* may generate a number of different regimes, e.g. the same macroeconomic structure may generate a diverse set of outcomes that include economic depression, full employment, inflation, hyper-inflation, and even stagflation. If this is the case, then one is faced with the problem of which sub-sample of simulated and observed time-series should be compared in order to carry out model validation. By incorrectly setting time t_0 in the model, one can generate a set of simulated output data that describes a different regime to that found in the empirical data. In addition to the issue of

correctly setting t_0 setting, one must identify the appropriate point at which to stop the simulation runs, i.e. to correctly set t_n. If t_n is set incorrectly then the simulated data may include multiple regimes that are covered by the empirical data. If the start or end points (or both) for the simulation runs are incorrectly set, there is the danger that one incorrectly rejects a 'true' model on the basis of its simulated outputs. We should also note that if, as is frequently the case, the modeller sets the simulation runs to end at a point where the model reaches a stationary or almost stationary behaviour, one is implicitly assuming that the empirical evidence comes from a stationary DGP. This may, or may not, be the case[24].

What if the observed micro and macro parameters are time dependent? One needs to be sure that the empirically estimated parameters that we assume are slow changing variables (and, hence, can reasonably treat as fixed within the timescale explored by the model) are not actually time dependent. If they are, then the researcher needs to go back and rethink the structural relationships between slow and fast variables, the timescale of the model[25], or both.

Finally, issues of prediction and counterfactuals are core to calibration. To what extent do the predictions of the models take into account data that lies outside the current regime? Here the Lucas critique comes to the fore once again. Real economic agents not only use statistical processes based on past experience (adaptive expectations) but use data to project into the future. In this way, agents are able to respond to exogenous economic shocks. As discussed in section 2.2.1, this was the basis for the rational expectations critique of Keynesian behavioural models. Exogenous economic shocks alter behaviours but leave the underlining structure unchanged. As a consequence Keynesian theories seriously mispredict the consequences of a shock, whereas a model of the micro fundamentals - individual production functions and utility functions - would not.

To sum up, calibration addresses the over-parameterisation problem by reducing the space of possible 'worlds' that are explored in an AB model (Kwásnicki, 1998). This is done through the use of empirical data, such that the model *mDGP* resembles as closely as possible the actual *rwDGP* that we observe. This can be a sensible goal whenever the analysis aims to forecast the future or generate policy implications. However, on the down side, calibration has a strongly conservative tendency. It supports the continuation of orthodox theories and models for which empirical data are readily available. It disadvantages new theories and new models for which empirical research has not yet caught up, and mitigates against the study of qualitative phenomena that are either difficult to measure or are inherently immeasurable by their very nature. We have also discussed the numerous practical problems involved in actually carrying out calibration exercises. A notable problem is the availability of high-quality data in sufficient quantities over the many parameters of a typical AB model. Even if this is achieved, the modeller faces a range of problems such as knowing *ex ante* whether the real-world system being modelled is ergodic or non-ergodic, and the ability to correctly set the initial conditions of the model, plus the beginning and end points of the simulation runs, to match that of the real-world system captured in the empirical dataset(s). Such requirements are far from trivial!

---

[24] See for example the calibration exercises performed by Bianchi *et al.* (2005) on the CATS model developed in a series of papers by Gallegati *et al.* (2003, 2005).

[25] An important issue related to time-scales in AB models, which we shall just mention here, concerns the choice made about the timing in the model. Whether we assume that the time-interval [t, $t+1$] describes a day, or a quarter, or a year (and whether one supposes that the 'updating scheme' is asynchronous or parallel), has non-trivial consequences for calibration and empirical validation.

4.3 The History-Friendly Approach

The history-friendly approach offers an alternative solution to the problem of over-parameterisation. 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, Richard Nelson and Sydney 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)[26].

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 is Malerba *et al*. (1999), and Malerba and Orsenigo (2001). In the first of these papers, the authors outlined the approach and then applied it to a discussion of the transition in the computer industry from mainframes to desktop PCs. In the second of these papers, the approach was applied to the pharmaceutical industry and the role of biotech firms therein. Here we shall keep the description of the approach succinct[27]. 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 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.

---

[26] See Windrum (1999) for a detailed discussion of early neo-Schumpeterian models.
[27] Interested readers are directed to Windrum (2004) for a detailed critique of history-friendly modelling.

There are many points here that deserve closer inspection. Let us begin with issues that concern the structure of the model and the object of analysis. First, the modelling activity that has been conducted is, in practice, informed by the history of a few, key companies rather than the history of an entire industry. For instance, Malerba *et al*. (1999) is calibrated to capture one particular computer company – IBM – rather than the entire industry. This severely restricts the universality of the model. As a consequence, the micro-economic description of the supply-side of the industry is highly stylised. The demand-side of the computer industry model is also highly stylised. Indeed, many of the behavioural assumptions made about the supply and demand sides do not appear to be driven by industry-specific empirical observations. Windrum (2004) suggests that this reflects practical difficulties in collecting sufficient amounts of high quality data at the industry level.

This leads us to an important question: to what extent can one hope to acquire all the relevant data needed to build an empirically sound industry-level model? If this is not possible, then a further question follows: what are we to do if the empirical evidence is incomplete, offers no guidance on a particular point, or else seems to contain alternative, competing viewpoints?

Finally, limited attention is given to sensitivity analysis in the history friendly models, as parameters and rules are supposed to be deduced from the industry under study. The lack of sensitivity analysis is particularly noticeable with regards to cross-run variability.

Aside from the issues relating to implementation, the history-friendly approach raises a set of fundamental methodological issues. First, the approach to empirical validation that is advocated involves comparing the output traces of a simulated model with detailed empirical studies of the actual trace history of an economic system. We are immediately confronted, once again, with problems associated with comparing individual output traces generated by the model *mDGP* with individual traces generated by the real-world *rwDGP*. This does not move us much further on from ascertaining whether a model is 'capable' of generating an output trace that resembles an empirically observed trace. It is not a very strong test. An individual simulated trace may, or may not, be typical of the model.

Second, another problem concerns the ability to backwardly induce the 'correct' set of structural assumptions, parameter settings, or initial conditions from a set of traces – even if we have a model that generates an appropriate distribution of output traces. Simply stated, there are, in principle, a great many combinations of alternative parameter settings that can produce an identical output trace. We cannot deduce which combination of parameter settings is correct, let alone the appropriate set of structural assumptions.

Third, the method implies that we are able to construct counterfactual histories (although the authors do not themselves engage in this in their papers). For example, we need to be able to construct a world in which IBM did not enter the PC market. This poses a very serious question. Could the PC market have developed in much the same way had IBM not invented the PC? Can we meaningfully construct a counterfactual history? As Cowan and Foray (2002) discuss, it is exceedingly difficult in practice to construct counterfactual histories because economic systems are stochastic, non-ergodic, and structurally evolve over time (more on that in Section 4.4).

Fourth, there is a key methodological question related to the meaning of history: To what extent can we actually rely on history to be the final arbiter of theoretical and modelling debates? To pose the question another way, can simulations, in principle, be guided by history? In practice, it is unlikely that we will be able to appeal to history, either to bear witness, or to act as a final arbiter in a

dispute. This is because history itself is neither simple nor uncontested, and any attempt to develop a historically based approach to modelling faces deep level methodological problems[28].

The development of high quality accounts, open to critical scrutiny, is essential to the history-friendly approach (and indeed any other historically based methodology). It is, after all, on the basis of these accounts that guidance is taken on particular modelling choices, on parameter testing, and output evaluation. In recognising the limitations of any historical account, we simultaneously identify the limitations of decisions based on that account. But this is a strength, not a weakness, of open academic discourse. How, then, are we to proceed? Let us suggest the following possibility. While a single 'typical' history may not exist, we may be able to draw some generalisations on the basis of a large collection of historical case studies. To use an analogy used by Jerry Silverberg, rather than seeking to develop a model that describes the fall of one particular leaf from a tree (the history friendly approach), we should seek to develop general models, such as the bromide diffusion model in physics, that can be used to explain the fall of many leaves from many trees (and other phenomena). To get to this point, what is needed is the construction of high quality datasets. A wealth of empirical studies within the neo-Schumpeterian tradition, written over the last twenty years, can be drawn upon. It is also likely that new databases will be needed to be constructed in order to develop greater understanding of micro, meso, and macro phenomena. We shall return to this issue later. For the moment, it is important to observe that modellers need to ensure they do not prematurely restrict their models, given the lack of high quality data that currently exists. If the AB models that are developed are not flexible enough to consider alterative scenarios, then we will be left with a set of models that are less, not more, compatible with one another.

5   Conclusions: Open-ended issues for empirically validating AB models

In this paper, we have critically examined some of the most prominent issues in the empirical validation of agent-based simulation models.

We began by defining the methodological basis of empirical validation (section 2.1). The modeller tries to approximate a real-world DGP (*rwDGP*) that generates a set of observed data. The model contains a theoretical DGP (*mDGP*), which is simpler that the *rwDGP*, that 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*. The paper has identified a set of key issues associated with that this process of backwards-induction. These are (1) the isolation of the right set of mechanisms, forces and causal relations going on in the *rwDGP*; (2) the balance between instrumentalist and realist approaches; (3) the choice of a pluralist or apriorist methodological view; (4) the solution of the trade-off between analytical tractability and descriptive accuracy; (5) the way the identification / under-determination problem is tackled; and, finally, (6) the possibility of testing and falsifying any single hypothesis in isolation, especially when the phenomena under study undergo several structural changes. These issues are generic in empirical validation, and so apply to neoclassical and AB economists alike.

Section 2.2 provided an extensive examination of how neoclassical economics has been dealing with these issues. The breakdown of the consensus that was based on the Cowles Commission

---

[28] A well-known example of the contestability of history is evidenced by the ongoing debate about whether inferior quality variants can win standards battles (Leibowitz and Margolis, 1990; Arthur, 1988). As Carr (1961) observed in his classic work, however, history can be contestable at more fundamental and unavoidable levels.

approach has led to the development of four alternative, and competing, methodologies: (1) the structural modelling approach, (2) vector autoregressive models (VARs), (3) calibration; and (4) the London School of Economics (LSE) approach. Each approach holds a (very) different methodological position with respect to the six key issues.

Against this background, section 3 began the discussion of how empirical validation is carried out within the AB community. A discrete set of approaches, not only different with each other but different to those developed within neoclassical economics, have been developed by the AB community. We suggested there are two reasons for this. First, AB modellers are interested in phenomena such non-linearities, stochastic dynamics, non-trivial interactions among agents, and feedbacks from the micro to 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/problems that are not faced by neoclassical modellers. Second, and relatedly, the highly diverse structural content of AB models means they need to be analysed in very different ways. We have presented a novel taxonomy that maps out the key areas in which AB researchers differ. These are the nature of the object under study (i.e. of the stylised fact(s) under analysis), the goal of the analysis (in-sample vs. out-of-sample), the nature of the main modelling assumptions (e.g. parameters, decision rules, interaction, and timing), and the methodology of the analysis (i.e. sensitivity exercises). Having identified the nature and causes of heterogeneity of empirical validation, section 4 summarised three important approaches within AB economics: indirect calibration, the Werker-Brenner approach to empirical calibration, and the history-friendly approach.

In this, the final section of the paper, we draw together these various elements of the discussion. We present a taxonomy that captures the dimensions along which the indirect calibration, Werker-Brenner, and history-friendly approaches differ. We end with a discussion of the unresolved issues that require future attention by the AB community.

The taxonomy (Table 3) identifies four dimensions in which the indirect calibration, Werker-Brenner, and history-friendly approaches differ. First, there 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 empirical observations (data) that are used for empirical validation. In addition to empirical datasets, the Brenner-Werker 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 are 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.

As a result of this heterogeneity, the outcomes of the three approaches may differ. Furthermore, our overview not only highlights problems associated with each approach, it also brings to light a set of core issues that affect all the approaches and which (so far) remain unresolved. These can be fruitfully classified under the following headings:

1. alternative strategies for constructing empirically-based models;
2. problems that arise as a consequence of over-parameterisation in AB models;
3. debates over the definition of sufficiently strong empirical tests (including issues arising from the replication of stylised facts, comparison with unique real-world observations, and the issue of unconditional objects);
4. issues arising from the availability, quality and bias of available datasets; and
5. the usefulness and implications of counterfactuals for policy analysis.

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
Brenner-Werker	- Micro (industries, markets) - Macro (countries, world economy)	- Empirical data - Historical knowledge	- Assisting in model building - Calibrating 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 - Calibrating initial conditions and parameters - Validating simulated output	- First calibrate, then validate

Table 2. Taxonomy of dimensions of heterogeneity in ABM empirical validation approaches

Let us draw the paper to a close by discussing these open issues. To begin with, there is the crucial issue of whether (and how) one should build an AB model that is based on empirical observations. Methodological realism, as discussed, embodies the principle that a model should (in addition to other criteria) be judged on the 'realism' of its assumptions. An assumption has a higher degree of realism when supported by robust empirical evidence (Kagel and Roth, 1995; Plott and Smith, 1998). In other words, the model *mDGP* should capture the key observed features of the real-world *rwDGP*. But 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? This is the under-determination problem in a new context.

A number of different strategies exist for selecting assumptions in the early stages of model building[29]. 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,

---

[29] See Mäki (1994), Edmonds and Moss (2005), Pyka and Fagiolo (2005).

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.

Whatever the strategy employed, the AB modeller often faces an over-parameterisation problem. AB models with realistic assumptions and agent descriptions invariably contain many degrees of freedom. First, the model may contain a large number of micro-macro parameters. Second, the modeller may explore different interaction setups and agent decision rules. While the latter can reasonably be considered a modelling choice (justified by the particular issue under study), the former is a dimension that is often non-reducible. This is true, even if one adopts a KISS approach to model building and model selection. 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 the more degrees of freedom there are in the model. The over-parameterisation of a model (in the spaces of micro-macro parameters, variables, and/or decision rules) can seriously impair any validation or calibration exercise because there are a number of different combinations of parameter settings that can produce the same output. Since the parameter space describes all possible regimes (behavioural, technological, institutional, etc.) in which an economic system can find itself, discriminating among the different combinations means choosing among possible realities. Which one should then be compared with the empirical evidence?

We have discussed a number of escape strategies for these problems. First, one can use empirical evidence to restrict the degrees of freedom, by directly calibrating initial conditions and/or parameters (i.e. the set of possible 'worlds' modelled by the *mDGP*). Second, 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'. The modeller needs to address the issue of how the remaining worlds should be interpreted. What does it mean when one is comparing the model outputs that are generated under alternative sets of initial conditions/parameters that resist direct/indirect calibration? In fact, each combination represents a different economy or world, as it defines a different institutional, technological, market/industry and behavioural setup.

The issue is particularly relevant for the indirect calibration approach because it is frequently the case that many combinations of parameters and initial conditions are consistent with the set of stylised facts of interest. In the direct calibration approach the modeller also finds her/himself in a situation where many parameters and/or initial conditions cannot be directly estimated. Contrasting the output of any two alternative setups − as is often done with the direct and indirect calibration approaches − means one must perform comparative dynamics exercises. Such exercises are informative from a theoretical point of view because they allow for a better understanding of the properties of the model. However the interpretation of the results is unclear as far as policy prediction is concerned.

The interpretation of alternative points in the space of parameters/initial conditions brings us to another question faced by AB modellers. 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 (2002) discuss the issue at length. They 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. The critique may also apply to the interpretation of empirically plausible parameters and initial conditions. In light of the Cowan-Foray critique, one must consider whether any comparative dynamics exercises be informative.

More generally, comparing the outputs generated by AB models with real-world observations involves a set of very intricate issues. For example, Windrum (2004) 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.

We see that defining a strong empirical test for an AB model can be a very thorny problem. A common criticism of early models of technological change (in the Nelson-Winter tradition) was that they were not given sufficiently strong tests. In effect, they were evaluated on the basis of whether they were capable of generating outputs that resembled very general, macro observations. The models were not subject to rigorous testing procedures, either for model variables or model outputs. Indeed, it was very common to find that authors had not engaged in any form of sensitivity analysis but rather provided illustrative outputs from just a handful of simulation runs.

The fundamental difficulties in defining strong tests for model outputs is highlighted by Brock's (1999) discussion of 'unconditional objects' in economics (another aspect of the under-determination problem). 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 (Windrum, 2004). Similarly, the Nelson and Winter (1982) 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 neoclassical 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 (Brock, 1999). 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[30]. This requires one to only include in the model individual decision rules (e.g. learning) that have been validated by experimental/empirical evidence. Of course, this would require highly detailed and reliable data about microeconomic variables, possibly derived from extensive laboratory experiments.

This points us to a final core issue for AB modellers; the availability, quality and bias of available 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. The datasets that do exist are invariably biased. Datasets are constructed according to criteria that reflect certain choices and, as a consequence, have inbuilt biases. 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. Such problems exist with data from the recent past, just as they do for data from the more distant past. Further, linking to a point raised in section 2, econometrics is influenced by prevailing theoretical orthodoxy. As a consequence, it is very difficult to test data on new, alternative theories because suitable data are not available. The most famous example of this is the Keynesian revolution. Theoretical developments following the publication of Keynes' *General Theory* could not be tested, or put into policy practice until government agencies started to collect aggregate national data, notably on household consumption expenditure and firm investment expenditure. The data that had been collected up to this point had been informed by pre-Keynesian economic theory.

One of the main reasons why AB models have been developed is to provide significant answers to empirically observed puzzles that are unable to be solved within the neoclassical paradigm (see e.g. Dosi *et al*., 1994). AB models have been extremely successful in this task. Moreover, they have been able to explain how some crucial macroeconomic phenomena can be generated by the evolving networks of interactions among boundedly-rational agents in economies where the fundamentals may endogenously evolve over time. Examples range from growth and development patterns, to industry and market dynamics, to technological innovation, to the evolution of consumption and demand. What is more, they did so by taking on board methodological pluralism and avoiding the apriorist view that characterizes neoclassical economics. However, the ways in which AB models deal with empirical data still suffer from too much heterogeneity. Some of this heterogeneity can be traced back to the same causes that also affect empirical validation approaches in neoclassical economics (and ultimately in all social sciences). Our analysis has identified two directions that can be fruitfully pursued. First, a commonly accepted, minimal protocol for the analysis of AB models must be developed and agreed upon (also see Leombruni (2002) and Leombruni *et al*. (2006). This would allow AB models to become more comparable and reach methodologically sounder conclusions. Second, much more work needs to be done to address the core issues of empirical validation. We believe that the recent trend, which seems to highlight an

---

[30] This point was made by John Duffy in his plenary talk at the 2005 International Workshop on "Agent-Based Models for Economic Policy Design" (ACEPOL05) in Bielefeld (Germany). Also see Gilbert (2004).

increasingly growing interest in methodological issues within the AB community, is an optimistic signal in this direction.

References

Arthur, W.B., 1988, Competing technologies: an overview., in G. Dosi, C. Freeman, R. Nelson, G. Silverberg and L. Soete (eds.), *Technical Change and Economic Theory*, London: Pinter, 590-607.

Arthur, W.B., 1994, *Increasing Returns and Path-dependency in Economics*, University of Michigan Press, Ann Arbor.

Aruka, Y., 2001, *Evolutionary Controversies in Economics: A New Transdisciplinary Approach*, Springer, Tokyo and Berlin.

Bailey, K.D., 1988, The conceptualization of Validity. Current perspectives, *Social Science Research*, 17:117-136.

Bianchi, C., P. Cirillo, M. Gallegati and P. Vagliasindi, 2005, Validation in ACE models. An investigation of the CATS model, Unpublished Manuscript.

Blaug, M. 1980, *The methodology of economics or how economists explain*, Cambridge: Cambridge University Press.

Brenner, T., 2004, Agent learning representation – advice in modelling economic learning, Papers on Economics and Evolution #0416, Max Planck Institute, Jena.

Brenner, T., and Murmann, J.P, 2003, The use of simulations in developing robust knowledge about causal processes: methodological considerations and an application to industrial evolution, Papers on Economics and Evolution #0303, Max Planck Institute, Jena.

Brock, W., 1999, Scaling in economics: a reader's guide, *Industrial and Corporate Change*, 8: 409-446.

Carr, E.H., 1961, *What is History?*, London: Macmillan.

Chattoe, E., 2002, Building Empirically Plausible Multi-Agent Systems: A Case Study of Innovation Diffusion, in Dautenhahn, K. (ed.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Dordrecht Kluwer.

Christiano, L.J., M. Eichenbaum, and C.L. Evans 1999, Monetary Policy Shocks: What Have We Learned and To What End?, in J. Taylor, and M. Woodford (eds.), *Handbook of Macroeconomics, Volume 1A*, Amsterdam: Elsevier Science Ltd., pp. 65-148.

Cook, S. and D.F. Hendry 1994, The Theory of Reduction in Econometrics, *Poznan Studies in the Philosophy of the Sciences and the Humanities*, Vol. 38, 71-100.

Cooley, T.F., and Le Roy, S., 1985, *Atheoretical Macroeconometrics. A Critique*, *Journal of Monetary Economics*, 16, 283-308.

Cowan, R. and D. Foray, 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.

Cowan, R. and M. Rizzo, 1996, Genetic causation and modern economic history, *Kyklos*, 49: 273-3 7.

Dawid, H., 2005, Agent-based models of innovation and technological change, in L. Tesfatsion and K. Judd, *Handbook of Computational Economics II: Agent-based Computational Economics*, forthcoming.

Doran,, J. 1997, From computer simulation to artificial societies, *SCS Transactions on Computer Simulation*, 14(2):69-78.

Dosi, G., 1988, Sources, Procedures and Microeconomic Effects of Innovation, *Journal of Economic Literature*, 26: 126-171.

Dosi, G., Freeman, C. and Fabiani, S. (1994), "The process of economic development: introducing some stylized facts and theories on technologies, firms and institutions", *Industrial and Corporate Change*, 3: 1-46.

Dosi, G. and Nelson, R.R., 1994, An Introduction to Evolutionary Theories in Economics, *Journal of Evolutionary Economics*, 4: 153-172.

Dosi, G. and L. Orsenigo, 1994, Macrodynamics and microfoundations: an evolutionary perspective, in Granstrand, O. (ed.), *The Economics of Technology*. North Holland, Amsterdam.

Dosi, G., Marengo, L. and Fagiolo, G., 2005, Learning in Evolutionary Environment, in Dopfer, K. (ed.), *Evolutionary Principles of Economics*, Cambridge, Cambridge University Press.

Downward, P., J. Finch and J. Ramsay, 2002, Critical realism, empirical methods and inference: a critical discussion, *Cambridge Journal of Economics*, 26: 48 -500.

Edmonds, B. and 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.

Engle, R. F., D.F. Hendry and J. F. Richard, 1983, Exogeneity, *Econometrica*, 51 (2), 277-304.

Fagiolo, G., 1998, Spatial interactions in dynamic decentralized economies: a review, in Cohendet, P., Llerena, P., Stahn, H. and Umbhauer, G. (eds.), *The Economics of Networks*. Interaction and Behaviours, Springer Verlag, Berlin - Heidelberg.

Fagiolo, G. and 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.

Fagiolo, G., Dosi, G. and 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.

Fagiolo, G., Marengo, L. and Valente, M., 2004b, "Endogenous Networks in Random Population Games", *Mathematical Population Studies*, 11: 121-147.

Favero, C., 2001, *Applied Macroeconometrics*, Oxford: Oxford University Press.

Feyerabend, P., 1975, *Against Method*. New Left Books, London.

Forni, M. and Lippi, M., 1997, *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Clarendon Press, Oxford.

Friedman, M., 1953, *The Methodology of Positive Economics, in Essays in Positive Economics*, Chicago: University of Chicago Press.

Gallegati, M., G. Giulioni, A. Palestrini and D. DelliGatti, 2003, Financial fragility, patterns of firms' entry and exit and aggregate dynamics, *Journal of Economic Behavior and Organization*, 51 : 79-97.

Gallegati, M., D. Delli Gatti, C. Di Guilmi, E. Gaeo, G. Giulioni and A. Palestrini, 2005, A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility, *Journal of Economic Behavior Organization*, 56: 489-512.

Gibbard, A. and H. Varian, 1978, Economic Models, *Journal of Philosophy*, 75, 664-677.

Gilbert, N. and K. Troitzsch, 1999, *Simulation for the Social Scientist*, Open University Press, Milton Keynes.

Gilbert, N., 2004, Open problems in using agent-based models in industrial and labor dynamics, in R. Leombruni and M. Richiardi (eds.), *Industry and Labor Dynamics: the agent-based computational approach*, 401-405, Singapore, World Scientific.

Haavelmo, T., 1944, The Probability Approach in Econometrics, *Econometrica*, 12: 1-115.

Hansen, L. and T. Sargent, 1980, Formulating and Estimating Dynamic Linear Rational Expectations Models, reprinted in R. Lucas and T. Sargent (eds.), *Rational Expectations and Econometric Practice*, 1981, London: Allen & Unwin, pp. 295-320.

Hendry, D.F., 1988, Encompassing, *National Institute Economic Review*, August, 88-92.

Hendry, D.F. and M. Morgan, 1995, *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Hicks, J. R., 1937, Mr. Keynes and the Classics, reprinted in *Critical Essays in Monetary Theory*, Oxford: Clarndon Press, 1967.

Hollis, M., 1994, *The Philosophy of Social Science: An Introduction*. Cambridge University Press, Cambridge.

Hoover, K., 1994, Econometrics as observation: the Lucas critique and the nature of econometric inference, *Journal of Economic Methodology*, 1, 1, 65-80.

Hoover, K. (ed.), 1995a, *Macroeconometrics, Developments, Tensions, and Prospects*, Boston-Dordrecht-London: Kluwer.

Hoover, K., 1995b, Facts and Artifacts: Calibration and the Empirical Assessment of Real-Business-Cycle Models, *Oxford Economic Papers*, 47, 1, 24-44.

Hoover, K., 2006, *The Methodology of Econometrics,* in T.C. Mills and K. Patterson (eds.) *Palgrave Handbooks of Econometrics*, vol 1, *Econometric Theory*, MacMillan.

Howson, C. 2000, Evidence and Confirmation, in W.H. Newton-Smith (ed.), *A Companion to the Philosophy of Science*, Malden, MA: Blackwell Publishers, pp. 108-116.

Janssen, M.C.W., 1994, Economic Models and Their Applications, *Poznan Studies in the Philosophy of the Science and the Humanities*, 38, 101-116.

Judd, K., 1998, *Numerical Methods in Economics*, MIT Press, Cambridge, Massachusetts.

Kagel, J.H. and Roth, A.E. (eds.), 1995, *The Handbook of Experimental Economics, Vol. 1*, Princeton University Press, Princeton, N.J.

Kaldor, N., 1961, *Capital Accumulation and Economic Growth*, in Lutz, F. A. and Hague, D. C. (ed.), *The Theory of Capital*, London: Macmillan, pp. 177-222.

Kirman, A.P., 1989, The intrinsic limits of modern economic theory: the emperor has no clothes, *Economic Journal*, 99: 126-139.

Kirman, A.P., 1992, Whom or what does the representative individual represent ?, *Journal of Economic Perspectives*, 6: 117-136.

Kirman, A.P., 1997a, The Economy as an Interactive System, in Arthur, W.B., Durlauf, S.N. and Lane, D. (eds.), *The Economy as an Evolving Complex System II*, Santa Fe Institute, Santa Fe and Reading, MA, Addison-Wesley.

Kirman, A.P., 1997b, The Economy as an Evolving Network, *Journal of Evolutionary Economics*, 7: 339-353.

Klein, L.R., 1964. Estimation of interdependent systems in macro-econometrics, *Econometrica*, 37, 171-192.

Knight, F.H., 1921, *Risk, Uncertainty, and Profits*, Chicago: Chicago University Press.

Koesrindartoto, D., Sun, J. and Tesfatsion, L., 2005, An Agent-Based Computational Laboratory for Testing the Economic Reliability of Wholesale Power Market Designs, IEEE Power Engineering Society Conference Proceedings.

Kuhn, T., 1962, *The Structure of Scientific Revolutions*, Chicago: Chicago University Press.

Kwásnicki, W., 1998, Simulation methodology in evolutionary economics, in Schweitzer, F. and G. Silverberg (eds.), *Evolution und Selbstorganisation in der Ökonomie*. Duncker and Humblot, Berlin.

Kydland, F.E., and E.C. Prescott, 1982, Time to Build and Aggregate Fluctuations, *Econometrica*, 50, 6, 1345-1370.

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.

Lane, D., 1993a, Artificial worlds and economics, part I, *Journal of Evolutionary Economics*, 3: 89-107.

Lane, D., 1993b, Artificial worlds and economics, part II, *Journal of Evolutionary Economics*, 3: 177-197.

Law, A. and Kelton, W.D., 1991, *Simulation Modeling and Analysis*, New York, McGraw-Hill.

Leamer, E.E., 1978, *Specification Searches, Ad Hoc Inference with Nonexperimental Data*, New York: John Wiley.

Liebowitz, S. J. and S. E. Margolis, 1990, "The Fable of the Keys," *Journal of Law and Economics*, 22, 1-26.

Leombruni, R., 2002, The Methodological Status of Agent-Based Simulations, Working Paper No. 19, LABORatorio R. Revelli, Centre for Employment Studies, Turin, Italy.

Leombruni, R., Richiardi, M. Saam, N. 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>.

Lucas, R., 1976, Econometric Policy Evaluation: A Critique, in K.Brunner, and A.H. Meltzer (eds.), *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, Spring, Amsterdam: North Holland, pp. 161-168.

Lucas, R., and T. Sargent, 1979, After Keynesian Macroeconomics, reprinted in R. Lucas and T. Sargent (eds.), *Rational Expectations and Econometric Practice*, 1981, London: Allen & Unwin, pp. 295-320.

Mäki, U., 1992, On the Method of Isolation in Economics, *Poznan Studies in the Philosophy of the Sciences and the Humanities* 26:19-54.

Mäki, U., 1994, Reorienting the assumptions issue, in R. Backhouse (ed.), *New Directions in Economic Methodology*, Routledge, London and New York.

Mäki, U., 1998, Realism, in J.B. Davis, D. Wade Hands, and U. Mäki (eds.), *The Handbook of Economic Methodology*, Cheltenham, UK: Edward Elgar, pp. 404-409.

Mäki, U., 2003, 'The methodology of positive economics' (1953) does not give us *the* methodology of positive economics, *Journal of Economic Methodology*, 10, 4, 495-505.

Mäki, U., 2005, Models are experiments, experiments are models, *Journal of Economic Methodology*, 12, 2, 303-315.

Malerba, F., and 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.

Malerba, F., Nelson, R.R., Orsenigo, L., and Winter, S.G., 1999, History friendly models of industry evolution: the computer industry, *Industrial and Corporate Change*, 8, 3-41.

Marengo, L. and Willinger, M., 1997, Alternative Methodologies for Modeling Evolutionary Dynamics: Introduction, *Journal of Evolutionary Economics*, 7: 331-338

Marks, B., 2005, Agent-Based Market Design, Australian Graduate School of Management, *mimeo*.

McCloskey, D.N., 1995, *The Rhetoric of Economics*, Brighton, Wheatsheaf.

Mirowski, P., 1998, Probability, in J.B. Davis, D. Wade Hands, and U. Mäki (eds.), *The Handbook of Economic Methodology*, Cheltenham, UK: Edward Elgar, pp. 391-393.

Mitchell, S., 2003, *Biological Complexity and Integrative Pluralism*, Cambridge University Press, Cambridge.

Modigliani, F., 1944, Liquidity Preference and The Theory of Interest of Money, *Econometrica*, 12, 44-88.

Moneta, A., 2005, Causality in Macroeconometrics: Some Considerations about Reductionism and Realism, *Journal of Economic Methodology*, 12(3), 433-453.

Morrison, M. and Morgan, M.S., 1999, Models as mediating instruments, in Morgan, M.S. and Morrison, M. (eds.), *Models as Mediators. Perspectives on Natural and Social Science*, pp. 10-37, Cambridge University Press, Cambridge.

Nelson, R. and S. Winter, 1982, *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge.

Nelson, R.R., 1995, Recent Evolutionary Theorizing About Economic Change, *Journal of Economic Literature*, 33: 48-90.

Plott, C.R. and Smith, V.L. (eds.), 1998, *Handbook of Experimental Economics Results*, North Holland, Amsterdam/New York.

Pyka, A. and G. Fagiolo, 2005, Agent-Based Modelling: A Methodology for Neo-Schumpeterian Economics, in Hanusch, H. and A. Pyka, eds., *The Elgar Companion to Neo-Schumpeterian Economics*. Edward Elgar, Cheltenham.

Richiardi, M., 2003, The Promises and Perils of Agent-Based Computational Economics, Working Paper No. 29, LABORatorio R. Revelli, Centre for Employment Studies, Turin, Italy.

Saviotti, P.P., and Pyka, A., 2004, Economic development, qualitative change and employment creation, *Structural Change and Economics Dynamics*, 15(3): 265-287.

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

Silverberg, G. and Verspagen, B. 1995, Evolutionary theorizing on economic growth, IIASA, Laxenburg, Austria, Working Paper, WP-95-78.

Silverberg, G., Dosi, G., and Orsenigo, L, 1988, Innovation, diversity and diffusion: a self-organisation model, *Economic Journal*, 98, 1032-1054.

Sims, C.A., 1980, Macroeconomics and reality, *Econometrica*, 48, 1, 1-47.

Sims, C.A., 1982, Policy analysis with econometric models, *Brookings Papers on Economic Activity*, 13, 1, 107-152.

Stock, J.H., 2001, Macro-Econometrics, *Journal of Econometrics*, 100, 29-32.

Tesfatsion, L., 1997, How Economists Can Get ALife, in Arthur, W., S. Durlauf and D. Lane (eds.), *The Economy as an Evolving Complex System II*. SantaFeInstitute, Santa Fe and Reading, MA, Addison-Wesley.

Tesfatsion, L., 2001a, Structure, behavior, and market power in an evolutionary labor market with adaptive search, *Journal of Economic Dynamics and Control*, 25: 419-457.

Tesfatsion, L., 2002, Agent-based Computational Economics: Growing Economies from the Bottom Up, Working Paper 1 , Iowa State University, Dept. of Economics.

Valente, M., 2005, Qualitative Simulation Modelling, Faculty of Economics, University of L'Aquila, L'Aquila, Italy, mimeo.

Vega-Redondo, F., 1996, *Evolution, Games, and Economic Behavior*, Oxford University Press, Oxford.

Werker, C. and Brenner, T., 2004, Empirical Calibration of Simulation Models, Papers on Economics and Evolution # 0410, Max Planck Institute for Research into Economic Systems, Jena.

Windrum, P., 1999, 'Simulation models of technological innovation: a review', *American Behavioral Scientist*, Vol. 42 (10), 1531-1550.

Windrum, P., 2004, 'Neo-Schumpeterian simulation models', *Merit Research Memoranda 2004-004*, MERIT, University of Maastricht. Forthcoming in Hanusch, H. and A. Pyka, eds., *The Elgar Companion to Neo-Schumpeterian Economics*. Edward Elgar, Cheltenham.

Windrum, P., 2005, 'Heterogeneous preferences and new innovation cycles in mature industries: the camera industry 1955-1974', *Industrial and Corporate Change* (in press).

Windrum, P. and Birchenhall, C., 1998, 'Is life cycle theory a special case?: dominant designs and the emergence of market niches through co-evolutionary learning', *Structural Change and Economic Dynamics*, Vol. 9, 109-134.

Windrum, P. and Birchenhall, C., 2005, 'Structural change in the presence of network externalities: a co-evolutionary model of technological successions', *Journal of Evolutionary Economics*, Vol. 15 (2),.123-148.

Wooldridge, M. and Jennings, N.R., 1995, Intelligent agents: theory and practice, *Knowledge Engineering Review*, 10: 115-152.

Further readings

Amman, H., Kendrick, D. and Rust, J., 1996, *Handbook of Computational Economics*, Elsevier-North Holland, Amsterdam, The Netherlands.

Boulding, K. E., 1991, What is evolutionary economics?, *Journal of Evolutionary Economics*, 1: 9-17.

Chiaromonte, F. and Dosi, G., 1993, Heterogeneity, competition, and macroeconomic dynamics, *Structural Change and Economic Dynamics*, 4, 39-63.

J.B. Davis, D. Wade Hands, and U. Mäki (eds.) 1998, *The Handbook of Economic Methodology*, Cheltenham, UK: Edward Elgar, pp. 391-393.

Dawid, H. and Riemann, M., 2004, Evaluating Market Attractiveness: Individual Incentives vs. Industrial Profitability, *Computational Economics*, 24, 321-355.

Dosi, G., Fabiani, S., Aversi, R., and Meacci, M., 1994, The dynamics of international differentiation: a multi-country evolutionary model, *Industrial and Corporate Change*, 2-3, 225-241.

Dosi, G., Freeman, C., Nelson, R.R., Silverberg, G. and Soete, L (eds.), 1988, *Technical Change and Economic Theory*, Francis Pinter, London.

Dosi, G., Marsili, O., Orsenigo, L., and Salvatore, R., 1995, Learning, market selection, and the evolution of industrial structures, Small Business Economics, 7, 411-436.

Epstein, J.M. and Axtell, R., 1996, Growing Artificial Societies: Social Science from the Bottom-Up, Washington D.C., MIT Press and Brooking Press.

Fagiolo, G., 2005, Endogenous Neighborhood Formation in a Local Coordination Model with Negative Network Externalities, *Journal of Economic Dynamics and Control*, 29: 297-319.

Fisher, F., 1983, *Disequilibrium Foundations of Equilibrium Economics*. Cambridge University Press, Cambridge.

Foley, D.K., 1994, A statistical equilibrium theory of markets, *Journal of Economic Theory*, 62: 321-345.

Frenken, K., 2005, History, state and prospects of evolutionary models of technical change: a review with special emphasis on complexity theory, in: J. Casti (ed.) *Complexity Focus,* Turin: Exystence.

Gourieroux, C. and Monfort, A., 1996, *Simulation-based Econometric Models*, Oxford University Press, Oxford.

Keuzenkamp, H.A., 2000, *Probability, Econometrics and Truth. The Methodology of Econometrics*, Cambridge: Cambridge University Press.

Kleijnen, J., 1998, Experimental design for sensitivity analysis, optimization, and validation of simulation models, in Banks, J. (ed.), *Handbook of Simulation*. Wiley, New Yor k.

Klejinen, J., 2000, Validation of Models: Statistical Techniques and Data Availability, *Proceedings of 1999 Winter Simulation Conference*.

Kwásnicki, W., and Kwásnicka H., 1992, Market, innovation and competition, *Journal of Economic Behaviour and Organisation*, 19, 343-368.

Sargent, R. P., 1998, Verification and validation of simulation models, Proceedings of 1998 Winter Simulation Conference.

Sargent, R., P. Glasow, J. Kleijnen, A. Law, I. McGregor and S. Youngblood, 2000, Strategic directions in verification, validation and accreditation research, Proceedings of 2000 Winter Simulation Conference.

Sargent, T.J., 1987, *Dynamic Macroeconomic Theory*, Harvard University Press, Cambridge.

Silverberg, G., 1987, Technical progress, capital accumulation and effective demand: a self-organisation model, in *Economic Evolution and Structural Adjustment*, D. Batten, J. Casti and B. Johansson (eds.), Berlin, Heidelberg, New York: Springer-Verlag.

Silverberg, G., and Lehnert, D., 1993, Long waves and evolutionary chaos in a simple schumpeterian model of embodies technical change, *Structural Change and Economic Dynamics*, 4, 9-47.

Silverberg, G. and Verspagen, B., 1994a, Collective learning, innovation and growth in a boundedly rational, evolutionary world, *Journal of Evolutionary Economics*, 4: 207-226.

Silverberg, G., and Verspagen, B., 1994b, Learning, innovation and economic growth: a long-run model of industrial dynamics, *Industrial and Corporate Change*, 3:1, 199-223.

Silverberg, G., and Verspagen, B., 2005, A percolation model of innovation in complex technology spaces, *Journal of Economic Dynamics and Control*, 29:1-2, 225-244.

Stanislaw, H, 1986, "Tests of computer simulation validity. What do they measure?", *Simulation and Games*, 17: 173-191.

Tesfatsion, L., 2001b, Agent-based modelling of evolutionary economic systems, *IEEE Transactions on Evolutionary Computation*, 5: 1-6.

Winter, S.G., Kaniovski, Y.M., and Dosi, G., 2000, Modelling industrial dynamics with innovative entrants, *Structural Change and Economic Dynamics*, 11, 255-293.

Yildizoglu, M., 2002, Competing R&D strategies in an evolutionary industry model, *Computational Economics*, 19: 5 -6.