INSTITUTE OF ECONOMICS

Scuola Superiore Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Validation of Agent-Based Models in Economics and Finance*

Giorgio Fagiolo °
Mattia Guerini §°
Francesco Lamperti °¶
Alessio Moneta °
Andrea Roventini °§

* *Forthcoming in Beisbart, Claus and Nicole J. Saam (Eds.)"Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, Philosophical Perspectives". Cham: Springer International Publishing*

° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
§ OFCE-Sciences Po, France
¶ FEEM, Milan, Italy

2017/23 **September 2017**
ISSN(ONLINE) 2284-0400

Validation of Agent-Based Models in Economics and Finance*

Giorgio Fagiolo[†1], Mattia Guerini[‡2,1], Francesco Lamperti[§1,3], Alessio Moneta[¶1], and Andrea Roventini[‖1,2]

[1]Istituto di Economia, Scuola Superiore Sant'Anna, Pisa, Italy

[2]OFCE - Sciences Po, Paris, France

[3]FEEM, Milano, Italy

September 20, 2017

Abstract

Since the influential survey by Windrum et al. (2007), research on empirical validation of agent-based models in economics has made substantial advances, thanks to a constant flow of high-quality contributions. This Chapter attempts to take stock of such recent literature to offer an updated critical review of existing validation techniques. We sketch a simple theoretical framework that conceptualizes existing validation approaches, which we discuss along three different dimensions: (i) comparison between artificial and real-world data; (ii) calibration and estimation of model parameters; and (iii) parameter space exploration.

JEL codes: C15, C52, C63

Keywords: agent based models, validation, calibration, sensitivity analysis, parameter space exploration

*We gratefully acknowledge the support by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 649186 - ISIGrowth. Further, we express our gratitude to Francesca Chiaromonte, Giovanni Dosi, Mauro Napoletano, Marcelo Pereira, Amir Sani and Maria Enrica Virgillito for helpful comments and discussions on the issues surveyed in this chapter. This work will appear as a chapter of: Beisbart, Claus and Nicole J. Saam. Eds. (forthcoming). Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, Philosophical Perspectives. Cham: Springer International Publishing.

†Correspondent author. Address: Piazza Martiri della Libertà 33 - 56127 Pisa (Italy). Email: giorgio.fagiolo@santannapisa.it

‡mattia.guerini@sciencespo.fr

§francesco.lamperti@santannapisa.it

¶alessio.moneta@santannapisa.it

‖andrea.roventini@santannapisa.it

1 Introduction

Modeling economies as complex systems using agent-based models (ABMs) is a relatively recent approach in economics (LeBaron and Tesfatsion, 2008; Farmer and Foley, 2009; Battiston et al., 2016; Turrell, 2016). Nevertheless, since the eighties of the last century, it has increasingly been attracting many scholars belonging to several sub-fields, becoming both a complement and a substitute for more traditional economic-modeling methodologies (Schelling, 1969, 1971; Epstein and Axtell, 1996; Axelrod, 1997). For example, ABMs are nowadays considered as a valid and effective competitor of standard Dynamic Stochastic General Equilibrium (DSGE) models in macroeconomics (see e.g. Dosi et al., 2010, Assenza et al., 2015, Hommes, 2013, and the survey Fagiolo and Roventini, 2017). Likewise, ABMs of financial markets are often considered better than traditional models based on the efficient-market hypothesis in explaining the statistical properties of buy-and-sell high-frequency dynamics (cf. e.g. Franke and Westerhoff, 2012; Leal et al., 2016).

Existing literature agrees that ABMs in economics provide two main added values, as compared to their orthodox counterparts. First, ABMs allow for more descriptive richness, as they describe ecologies of agents, locally interacting through non-obvious network structures, learning using incomplete information, and competing within imperfect markets. Second, the modeler developing an ABM has typically more flexibility in both input and output validation of its model.

This second feature of ABMs has always attracted a lot of attention and has generated, especially in the last years, a booming number of contributions. Back in 2007, the influential article by Windrum et al. (2007) attempted to survey ABMs validation methods, concluding that a lot of work would have been needed in order to fully develop a satisfactory set of techniques that consistently take ABMs to the data. In fact, many developments have occurred in the last ten years, which this Chapter tries to review. We go along such developments distinguishing between three different dimensions: (i) calibration and estimation of model parameters; (ii) comparison between artificial and real-world data; and (iii) parameter space exploration.

The chapter is organized as follows. First, we offer an introduction to the most-diffused practices in building and running agent-based models in economics (Section 2). In Section 3 we also sketch a simple theoretical framework that conceptualizes existing validation approaches. Section 4 provides a critical review of the literature, whereas in Section 5 we describe the most recent trends as to validation techniques in ABMs. Finally, Section 6 concludes with some critical considerations on future work.

2 Agent-Based Computational Economics: Common Practices

Notwithstanding the existence of different types of agent-based models, which have been developed by various sub-fields within economics, such as macroeconomics, industrial dynamics, finance, asset pricing, etc., one can identify some general patterns and common practices in the building process, under a common umbrella that we refer to as Agent-Based Computational Economics (ACE).

2.1 The Development of a Typical Agent-Based Model

Researchers typically do not know the *true* data generating process of phenomena under study, which we refer to as the real-world DGP ($rwDGP$). This can be seen as a very complicated, multi-parameter, stochastic process that governs the generation of a unique realisation of some time series and stylized fact that we can empirically observe and estimate. The goal of the modeler is therefore to provide a sufficiently good approximation of the $rwDGP$ by using an ABM. Naturally, the model releases a simplified DGP, which we refer to as the model-DGP ($mDGP$) and which should provide a meaningful explanation of the causal mechanisms generating the set of observed stylized facts, and, more generally, a good representation of the data. The empirical validation of an ABM is then the process by which one evaluates the extent to which the $mDGP$ is a good representation of the $rwDGP$.

The most adopted procedure for the development of an ABM is the *indirect calibration approach* (see Windrum et al., 2007).[1] This procedure is composed of four separate steps. The first consists in the identification of some real-world stylized facts of interest that the modeler wants to explain. In the second, one specify the model, the time-line of the events, the micro level dynamic equations which embody the individual agents' behavior, the set of parameters, and the set of random disturbances. Validation and the hypothesis testing are performed in the third step in order to compare model's output with the observations obtained from real world datasets. Finally, there could be a fourth step, where the ABM is employed for policy analysis exercises, implemented by changing some of the behavioral equations (e.g. capital requirements for macroprudential policy, as in Popoyan et al., 2017) or some of the parameters (e.g. tax rates for fiscal policies, as in Dosi et al., 2010). In what follows, we will explore these four steps in more details.

Stylized facts identification. The starting point of most ABMs is the identification of a set of micro and macro stylized facts and empirical regularities (e.g. static or dynamic correlations, empirically observed distributions, etc.). For the sake of generality, let us define as a stylized fact any possible type

[1] However, also other viable strategies are available: see for example the calibration approach proposed by Werker and Brenner (2004); Brenner and Werker (2007) and the history friendly models developed by Malerba et al. (1999).

of measurable unconditional object that can be investigated by means of some econometric exercises or more generally by statistical techniques. In such unconditional objects (see Brock, 1999), the causal generating mechanism, or data generating process (DGP), is unclear or too complex to be explained by a simple, low-dimensional system of dynamic equations. Examples of micro and macro stylized facts that have been empirically identified and replicated by means of ABMs in different fields encompass fat-tailed distributions of returns, endogenous emergence of flash crashes, long-run coexistence of heterogeneous investing rules in finance; fat-tailed distributions of firm growth rates, Zipf distribution of firm size, negative correlations between prices and market shares in monopolistic competitive markets in industrial dynamics; investment lumpiness, Okun and Beveridge curve, cyclical co-movements of variables in macroeconomics.

Model specification. After having singled out a set of possibly-interlinked stylized facts, one can try to find an explanation of the underlying causal forces, i.e. learning and describing the exact form of the real-world DGP, or at least a sufficiently accurate approximation of it. This is the ultimate objective of any ABM. The great advantage of ABMs vis-á-vis traditional ones commonly employed in economics and finance derives from its generative bottom-up approach genuinely rooted in evolutionary, complex-system theories (more on that in Lane, 1993; Tesfatsion, 2006; Fagiolo and Roventini, 2012, 2017). This indeed allows the researcher to keep into account the complex dynamics of a system that is populated by heterogeneous and boundedly-rational agents possessing a partial and possibly biased amount of information about the global system in which they live. However, agents are adaptive and learn in order to survive and prosper in such an uncertain framework following some forms of "Simonesque" (see Simon, 1991) satisfying principle.[2] Obviously, also when ABMs are developed to approximate the $rwDGP$, the number of degrees of freedom is high and different researchers can follow alternative routes according to their different expertise, backgrounds and theoretical hypothesis about the underlying generating process.[3]

Output validation. After the modeler has specified the behavioral equations of the actors populating the system, the ABM takes the form of a high dimensional, discrete-time stochastic process. Indeed, a part of the ACE community (especially in financial and asset pricing models) has strongly relied on Markov processes theory and on statistical physics tools in order to reduce the dimensionality of the model and eventually – under specific circumstances – to analytically solve the simplest model. But in general, as their complexity is high, ABMs are usually simulated by means of extensive Monte Carlo (MC) exercises in

[2]In that there is a major departure with respect to neoclassical models, where the (representative) agent has axiomatic preferences and maximize some smooth objective function with an easily computable bliss point.

[3]This is also one of the critique that is usually addressed to ACE. Since ABMs do not stick to some generally accepted axiomatic rule of behavior, they introduces discretionary choices that the modeler shall take. We will see how practitioners have coped with this issue in section 4.2.1.

which the random seed is modified along the MC dimension.[4] Once such MC exercises are performed and the synthetic data collected, the researcher can verify whether the model is able to generate unconditional objects which are not statistically significantly different from the ones previously observed in real world datasets.[5] Naturally, all these unconditional objects can be related to micro and macro variables.

Policy analysis. Once the model has been validated and proved to be able to account for the micro and macro empirical regularities under study, it can then be employed as a policy laboratory. Indeed, the impact of different economic policies in alternative scenarios can be studies by (i) varying some parameters, in particular those related to policy-maker interventions or to some broad institutional setting (e.g. tax rates); (ii) modifying initial conditions related to agents' state variables (e.g. income distribution, firms' technology); (iii) changing some agents' behavioral rules and interaction patterns (accounting e.g. for different market set-ups); (iv) introducing macro and/or micro heterogenous shocks (e.g. innovation or climate-damages shocks). These can be interpreted as *exogenous* policy changes, which allow a researcher to evaluate their effects in a fully controllable environment, where treatment effects can be easily identified, and endogeneity issues are almost absent.

In what follows, we will focus mostly on validation, discussing more in depth what is the relationship between an ABM, its inputs and outputs. In particular, the interpretation of the ABM as a process that transforms a set of inputs into outputs, poses two relevant questions: (i) how a *ceteribus paribus* variation of one input affects the output (a detailed discussion will be presented in Section 4.2.1), and (ii) to which extent the generated output is a good approximation of the real world phenomenon that the modeler aims to explain (discussion in Section 4.2).

2.2 Inputs of Agent-Based Models

In ABMs we can characterise two broad categories of inputs: initial conditions and parameters.

Initial conditions. They determine the values of macro and agents' state variables at the beginning of the simulation. In small scale ABMs, which are typically characterized by a deterministic skeleton and may posses at least one computable deterministic fixed point, initial conditions can be set at the equilibrium or in some contour of it (see Brock and Hommes, 1997; Westerhoff and Dieci, 2006; Guerini, 2013; Guerini et al., 2017) and then the ABM can be used to locally study the dynamics of the system.[6] However, in complex stochastic models, characterized by high levels of dimensionality, fixed points or

[4] As stated in Turrell (2016), the first agent-based model was developed in the thirties by the physicist Enrico Fermi in order to study the transport of neutrons through matter. Fermi's agent-based techniques was lately called Monte Carlo method (Metropolis and Ulam, 1949).

[5] In section 4.2 we will discuss the tools available for the verification and validation of ABMs.

[6] One can also study the basins of attraction of the dynamical system to study the robustness with respect to initial conditions.

statistical equilibria may not exist or not being know to the modeler. In such a framework, the selection of initial conditions can become a not-trivial issue, affecting the ergodicity and dynamics of the system, its output and more generally the very validity of the model. Different solutions are proposed in the ACE literature. The first one initializes the model in a homogeneity situation, where all the state variables of the agents are set equal to some economically meaningful values (see Dosi et al., 2010, 2013, 2015). The second solution instead draws initial conditions of a category of agents from a specific distribution, possibly grounded on some empirical regularity (e.g. fat-tailed firm size distribution as in e.g. Bianchi et al., 2007, 2008a). Finally, if rich enough datasets are available, one can employ them to directly impute initial conditions values (Hassan et al., 2008).

Parameters. They can fix some macro conditions, determine the size of agents' reactions to events, or they characterize the distributions from which stochastic decisions are taken by agents or shocks are drawn. In many economic and finance ABMs, parameters are of particular interest because they might drive the dynamic of the system to different statistical equilibria, they characterize some specific policy relations or some particular institutional arrangement that the modeler want to investigate. Parameters are usually calibrated or they can be estimated if appropriate data are available (see Section 4.1). Moreover, several methods allow to perform *sensitivity analysis* exercises in order to map the model responses to parameter variations (see Lee et al., 2015; Dosi et al., 2017c). These techniques will be discussed in more detail in section 5.

2.3 Outputs of Agent-Based Models

ABMs can generate both micro-level and aggregate outputs.

Micro-level output. The output of an ABM is composed of MC (the number of Monte-Carlo simulations) panel datasets containing different micro variables for a set I of agents over a specified time window T. Therefore the data can be collected in the form:

$$Z_{m,k} \in \mathbb{R}^{K \times MC}, \quad Z_{m,k} = \{z_{m,k,i,t};\ i = 1, \ldots, I;\ t = t_0, ..., T\}, \quad \forall k \in K, \tag{1}$$

where m denotes a specific Monte Carlo run, k indicates a micro-variable, i represents the agent cross-section dimension, and t capture the time dimension. As an example, in a macroeconomic ABMs these variables can represent households income or consumption levels, firm prices, capital, profits, etc.

Aggregate output The output of each Monte-Carlo simulation m contains also some time-series variables which emerge from the aggregation along the agent cross-section dimension. These aggregate series (denoted by an upper bar) take the form:

$$\bar{Z}_{m,h} \in \mathbb{R}^{H \times MC}, \quad \bar{Z}_{m,h} = \{\bar{z}_{m,h,t};\ t = t_0, ..., T\}, \quad \forall h \in H, \tag{2}$$

where h denotes the macro variable observed at different time steps t. For example, in an macroeconomic ABM, one can aggregate the micro variables concerning households, firms to compute GDP, price indexes, or the unemployment level.

2.4 Relation between Input and Output

For micro and for aggregate variables, the simulated synthetic datasets can generate a number of stylised facts or statistical properties that the modeler can compare with those obtained from the empirical analysis of the corresponding real-world dataset. This is the core of the indirect calibration approach presented above and the first validation test that an ABM must satisfy. The similarity between model-generated and real-world datas constitute the essence of the validation problem for ABMs, and it will be extensively discussed in sections 4 and 5. For the moment, let us only anticipates that in the last decade, different strategies have emerged tackling a set of related, but slightly different issues.

For any validation method, one should consider that in ABMs, the set of generated micro and macro variables $\{Z_{m,k}, \bar{Z}_{m,h}\}$ are not intrinsic features of the model itself, but are emerging properties coming from the complex interaction between model institutional arrangements and model inputs. Therefore, the statistical properties of the output might exist only conditional on the selected initial conditions, parametrization, the chosen random seed, the selected institutional arrangement. This means that a stylized fact that has been obtained under a specific set of inputs, might not necessarily hold true under different arrangements, and robustness analyses must be performed before using ABMs for policy analysis exercises.

3 Agent-Based Model Validation: Theoretical Framework

Validation of computer simulation models encompasses a variety of inter-related issues and concepts. Manson (2002) distinguishes between *output validation* and *structural validation*. The latter asks how well the simulation model represents the (prior) conceptual model of the real-world system, while the former assess how successfully the simulations' output exhibits the historical behaviors of the real-world target system. Further, output validation can be directly related to what Leombruni et al. (2006) define as *empirical validity* of a model, i.e. validity of the empirically occurring true value relative to its indicator.

Following Rosen (1985), let us consider two parallel unfolding: the evolution of the system (an economy, a market, etc.) and the evolution of the model of the system. If the model is correct, properly calibrated and initial conditions have been fixed according to the initial status of the real system, the simulation should mirror the historical evolution of the real-world system with respect to the variables of interest. This is exactly the assessment of the relationship between simulated and empirical data that constitutes the focus of this chapter. However, there are many other validity issues that we do not explicitly address. For example, Leombruni et al. (2006) discuss *theoretical validity* (the validity of the theory relative to the simulation), *model validity* (the validity of the model relative to the theory), *program validity* (the validity of the simulating program relative to the model), and *operational validity* (the validity of the theoretical concept to its indicator or measurement).

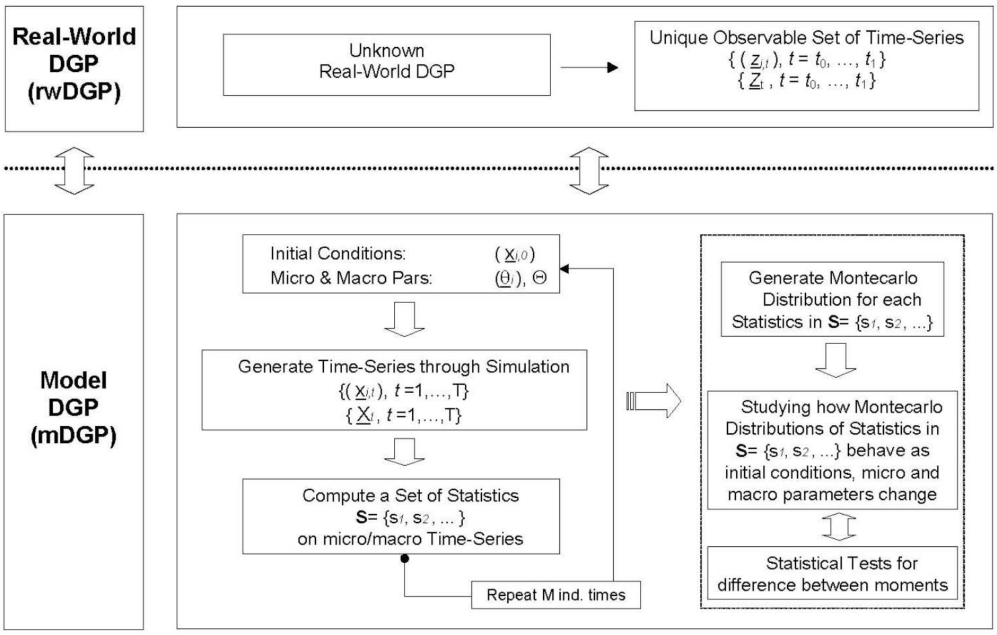
Using figure 1, let us consider the procedure for studying the output of an agent-based model already introduced in Windrum et al. (2007). Assume that the modeler knows (from a preliminary simulation study, or from some ex ante knowledge about the model under scrutiny) that the real-world system is ergodic, and that the $rwDGP$ displays a sufficiently stationary behavior for a time period after T^*. Further, let us assume that for a particular set of initial conditions, micro and macro parameters, the $mDGP$ runs until it reaches some form of stable behavior, which can be further summarized by a set of statistics $S = \{s_1, s_2, ...\}$. If the model is stochastic, as it is typically the case in ACE, each run will produce different values of the summary statistics s_j. Then, one must perform a sufficiently high number of independent Monte Carlo runs to estimate the distributions of those statistics, from which moments can be computed.[7] Such moments will depend on the initial choices that were made in terms of parameters and initial conditions. However, by exploring a sufficiently large number of points in the space of initial conditions and parameter values, and by computing, at least, the first two moments ($E(s_j)$ and $V(s_j)$) at each point, one could gain a deep understanding of the behavior of the model, test the robustness of the results, and identify the set(s) of parameters providing the most relevant dynamics. Finally, modelers and practitioners can make use of the uniquely observable real-world micro and macro time series and, under the assumptions outlined above, compute their longitudinal moments, which can be then compared to the simulated counterparts. The last step of such a procedure is nothing but an empirical validation exercise.

Starting from such a relatively general framework, a variety of approaches might be undertaken on the basis of how

- parameters and initial conditions are chosen;
- summary statistics are selected and combined to characterize model's behavior;

[7] See also Secchi and Seri (2017) on the issue of selecting the number of times a computational model should be run.

Figure 1: A procedure to study the output of ABMs.



- the fit between $rwDGP$ and $mDGP$ is measured;
- the space of initial conditions and parameter values is explored.

In what follows we provide a critical review of the most recent literature addressing the empirical validation of ABMs. In doing so, we highlight how the different contributions tackle one or more of the four issues above and we focus, in particular, on a set of approaches that are under development in the Institute of Economics at Scuola Superiore Sant'Anna.

4 Agent-Based Model Validation: Literature Review

The most general classification scheme for agent-based models (ABM) according to their level of empirical validity has been proposed by Axtell and Epstein (1994) and Barde and van der Hoog (2017) and consists of four levels:

- **Level 0**: the model is a caricature of reality, as established through the use of simple graphical devices (e.g., allowing visualization of agent motions).
- **Level 1**: the model is in qualitative agreement with empirical macro structures, as established by plotting e.g. the distributional properties of agent population. This is easiest way to matching stylized facts.

- **Level 2**: the model produces quantitative agreement with empirical macro-structures, as established through on-board statistical estimation routines.

- **Level 3**: the model exhibits quantitative agreement with empirical micro-structures, as determined from cross-sectional and longitudinal analysis of the agent population.

Publishing standards for economic and finance ABMs require at least that satisfaction of Level 1.[8] Under the Level 1 approach, an agent-based model gets validated through a statistical comparison of unconditional objects: stylized fact observed in real world data and emergent properties derived from the simulation environment. This amounts therefore at replicating a large number of possible micro and macro stylized facts characterizing the phenomena of interest.[9]

Current developments in empirical validation of ABMs shows a progression from Level-1 to to Level-2 models, as the mere replication of empirical regularities and other unconditional objects is increasingly replaced or supplemented by quantitative estimation. Such a fresh stream of research requires the models to generate series that exhibit the same dynamics (Marks, 2007; Lamperti, 2017), the same conditional probabilistic structure (Barde, 2016b), or the same causal structures (Guerini and Moneta, 2017) as those observed in the real world data. Furthermore, new methods to estimate and calibrate the parameters of ABMs have been developed with the aim of minimizing the distance between some distributional properties of the real simulation outcomes.

We claim that such new contributions will bring agent-based models on the same ground of advancement of the DSGE literature. Indeed the emerging literature on validation and estimation of ABMs represents the ACE counterpart of the progresses occurred in the estimation of DSGE models and well represented by the works of Del Negro and Schorfheide (2006); Canova and Sala (2009); Paccagnini (2010); Fernández-Villaverde et al. (2016).

Notwithstanding the possible overlaps between calibration, estimation and validation strategies, in what follows we propose a classification based on the central aim of each procedure. We therefore present in section 4.1 calibration and estimation procedures, which are essentially exercises for tuning model parameters or understanding the likelihood that a parameter is responsible for simulation results. We then discuss in section 4.2 the validation procedures, which evaluate how the inputs or outputs of simulated models resemble some well defined real world statistical properties.

[8]Level 0 models can be somehow accepted if their aim is merely exploratory rather than descriptive.

[9]See for examples Dosi et al. (2010, 2013, 2015, 2016a) for replication of business cycle and growth stylized facts; Dosi et al. (2017a) for accounting of labour-market micro and macro regularities; Popoyan et al. (2017) for the reproduction of many credit and interbank markets properties; Lamperti et al. (2017a) for capturing co-evolution of economic fundamentals with energy and emission quantities; Pellizzari and Dal Forno (2007); Leal et al. (2016) for simulating financial markets booms and busts.

4.1 Calibration and Estimation

Notwithstanding the fuzzy difference between calibration and esitmation, in what follow we discuss the two approaches as both aim at solving the same class of problems (in line with Hansen and Heckman, 1996). Calibration and estimation exercises have peculiar difficulties in Agent-Based modelling: the complex microeconomic interactions and the presence of ubiquitous nonlinearities (even in the simplest models) do not allow one to obtain a closed-form solution of the likelihood function and of the moments conditions. Therefore, one must resort to indirect inference or other simulation methods.

4.1.1 Indirect Inference

Indirect inference (Gourieroux et al., 1993) is the standard approach that has been developed for the estimation of small-scale agent-based models, characterized by relatively few parameters and short computational time.[10] Indirect inference allows one to estimate or to make inferences about the parameters of a model by means of simulation methods. It has been considered the preferred estimation choice since the very firsts ABM estimation attempts (see e.g. Winker and Gilli, 2001, 2004).[11] Also in the ABM framework, one could try to employ the Generalized Method of Moments approach (GMM), as in the very stylized models by Alfarano et al. (2005, 2006). However, in most of financial and economic Agent-Based Models, the moment function is completely unknown and one has to approximate it via Monte Carlo simulation exercises. In such a framework, the consistency and efficiency of the parameters estimates strongly depend on how well approximated is the moment generating function.

Following Chen et al. (2012), the procedure of the Method of Simulated Moments (MSM) can be summarized as follows:

> We first choose a vector of parameter values to generate the simulated time series by running the agent-based model with this chosen set of parameter values. We then compare some statistics (moments) of this simulated time series, the simulated moments, with those using real data, the sample moments. The difference between the two is used to form a distance function (the objective function). The MSM is purported to minimize the distance by searching over the entire parameter space.

Formally, one must estimate the vector of parameters θ^* that solves the following minimization problem:

$$\operatorname{argmin} \mathcal{L}(X^{RW}, X^{AB}; \theta) \tag{3}$$

[10] Benchmark models are for example the Brock and Hommes (1998) asset pricing model and the Kirman (1991) speculative bubbles model.

[11] See also Boswijk et al. (2007); Bianchi et al. (2008b); Goldbaum and Mizrach (2008); Franke (2009); de Jong et al. (2010); Franke and Westerhoff (2012); Chiarella et al. (2014); Platt and Gebbie (2016).

where X^{RW} and X^{AB} represent respectively the set of chosen moments observed in the real-world data and their counterpart derived from the ABM simulation.

This procedure is sufficiently general and in principle it is applicable to any type of ABM, but three drawbacks make it unfeasible in practice when the model complexity increases and the simulation time becomes a relevant constrain. First, an analytical solution for the problem of minimization of the approximated distance function is rarely available, forcing one to rely on numerical approximations. Second, moment selection is arbitrary and different choices may lead to differing estimation results. Third, the procedure is computationally intensive as one needs to run a sufficient number of Monte Carlo simulations of the model for each instance of the parameter space, and then evaluate the distance between the generated moments and those observed in sampled real-world data.

Very close alternatives to the MSM for estimating an agent-based model are the Simulated Minimum Distance (SMD) approach, which have been adopted by Fabretti (2013) and by Grazzini and Richiardi (2015) and the Simulated Maximum Likelihood (SML) by Kukacka and Barunik (2016).

4.1.2 Bayesian Approaches

As documented in the previous section, most of estimation and calibration works have been following a frequentist approach. However, after the popularization of Bayesian methods for the estimation of DSGE models (see Fernández-Villaverde and Rubio-Ramírez, 2007; Fernández-Villaverde et al., 2016), Bayesian inference techniques for estimating ABMs have been introduced in Grazzini et al. (2017). In general, the adoption of Bayesian strategies should reduce the discretionary choices involved in the somehow ad-hoc selection of the moments to be taken into consideration, the auxiliary model, or in any other metric that allows to evaluate the distance between the real and the simulated time series. Moreover, Bayesian approach could be more asymptotically efficient as it exploits the information provided by the whole distribution of the data and not only those of some specific moments.

However Bayesian methods are not exempt of issues. First, as documented by Canova and Sala (2009) and Fagiolo and Roventini (2012, 2017), the selection of the prior distribution can possibly generate an artificial curvature to the posterior distribution, when the likelihood tend to be flat, thus ending up in a interval calibration exercise. Second, the computational cost of Bayesian techniques is especially high when they are applied to ABMs for estimating the likelihood function. Such computational costs can be reduced by adopting efficient sampling schemes or likelihood function approximations, whose appropriateness should be evaluated on a case-by-case basis. However, as ABMs do not typically have a closed form solutions, a large number of Monte Carlo instances still need to be simulated (see Lamperti et al., 2017b).

4.2 Validation

4.2.1 Input Validation

The main focus of input validation has been (i) testing some of the behavioral assumptions typically included in Agent-Based models; (ii) selecting the initial conditions of the model under investigation; and (iii) exploring the parameter space. Let us now consider each of them.

Selection of behavioral rules. Well in line with behavioral economics (see e.g. the seminal contribution of Kahneman and Tversky, 1979), the very first input validation exercises of ABMs has resorted to laboratory experiments, which allow the researcher to directly verify how an individual behaves in a controlled environment. Typically these experiments have been used to test specific assumptions on agents' behavior embedded in small scale ABMs (see Hommes, 2011, 2013; Anufriev et al., 2016). Later, controlled laboratory experiments have been employed to estimate heuristic switching models (as in Anufriev and Hommes, 2012; Assenza et al., 2013).

Instead, in more complex ABMs, specific behavioral assumptions cannot be directly tested and other approaches have been adopted. We present here three of them that have allowed researchers to reduce the problem known in the literature as the "wilderness of bounded rationality".[QUOTE SIMS HERE] In the *adjustment heuristics* approach (Gaffeo et al., 2008; Assenza et al., 2015; Guerini et al., 2017), economic agents follow very basic economic principles in order to set some of their state variables. For instance, in these models, prices are fixed by the principle of excess demand.In the *management science* approach (Dawid et al., 2016, see), the decisions of agents are modeled starting from the researches carried out in the management literature. More specifically, consumers and firm behaviors are modelled following respectively the indications provided by the marketing and firm strategy literatures. Finally, the *empirical microeconomics* approach attempts to model the behavior of agents relying on microeconomic empirical evidence. This is the case, for example, in the "Schumpeter meeting Keynes" stream of models (Dosi et al., 2010, 2013, 2015).

Selection of initial conditions. Input validation can concern the selection of initial conditions of the model. Even simple and deterministic ABMs can display chaotic dynamics wherein small deviations between two configurations may generate extremely different time series (see Brock and Hommes, 1997, 1998; Hommes, 2013). However, if the model is ergodic, it explores the whole state space and reaches a stationary distribution. The problem of sensitive dependence on initial conditions can be tackled in small scale models, which are typically analytically solvable and where boundaries conditions and basins of attraction can be easily studied. On the contrary, it is still an open issue in more complex models, where the large support from which initial-condition values can be drawn impliess huge computational costs.

Exploration of the parameter space. Apart from parameter estimation and calibration, which have been thoroughly discussed in the previous section, in Agent-Based models one may need to explore the parameter space in order to assess the impact of different parameters on the dynamics of the model and to perform policy analysis exercises. An increasing number of works have started to investigate the robustness[12] of a model by running Monte Carlo simulations under different parameter settings (Ciarli, 2012; Salle and Yıldızoğlu, 2014; Bargigli et al., 2016; Dosi et al., 2016b, 2017b,c). More on that in section 5.3.

4.2.2 Output Validation

As introduced in section 3, output validation is the process of evaluating the extent to which the outcome of a simulated model is a good representation of real world observations. The baseline evaluation process focussing on the replication of stylized facts has been naturally embedded in most of Agent-Based models, which are often designed to account for phenomena unexplained by analytically tractable models.

Recently, more sophisticated statistical techniques have been developed to satisfy more stringent output validation requirements. In particular, they try to account for the "unconditional object" critique in Brock (1999) and to better discriminate among different ABMs reproducing the same set of stylized facts.

For instance, Marks (2013) employs three similarity measures — the Kullback-Leibler, the State Similarity Measure and the Generalized Hartley Metric — to analyse and validate an ABM of brand rivalry in the general validation framework developed in Marks (2007). Barde (2016b,a) and Lamperti (2017, 2016) develop two new similarity measures based on information theoretic criteria. Guerini and Moneta (2017) instead measure similarity by comparing the causal relations entailed in a Structural Vector Auto-Regression model estimated on both real and simulated data. The approaches of Lamperti (2017) and Guerini and Moneta (2017) will be discussed in more details in sections 5.1 and 5.2.

Note that all these recently developed validation techniques focus only on aggregate time series, while most of ABMs have been been able to replicate *both* micro and macro stylized facts. We believe that the next challenge is to further extend the new approaches to validate ABMs also in terms of microeconomic behaviors.

5 A New Wave of Validation Approaches

As discussed in the previous section, the debate on ABM validation is still very open and a novel wave of approaches has recently blossomed, offering to modelers and policy makers additional tools for the analysis of their models. Quite an impressive share of such new developments has been carried out within

[12]For robustness of the model, we here mean the stability of the results to small variations of the parameters.

the Institute of Economics of Scuola Superiore Sant'Anna in Pisa, which has historically been at the forefront of Agent-Based modelling in economics and finance.[13] This section outlines and discusses these contributions in relation to existing gaps in the literature.

5.1 Validation as Replication of Time Series Dynamics

Output validation concerns the assessment of how successfully simulations from a model mirror the historical behavior of the real-world target system (cf. section 3). In practice, this amounts at evaluating the degree of similarity between two or more time series. In most applications, as the method of simulated moments and simulated minimum distance presented earlier, such a step is performed by computing vectors of summary statistics (moments) from the synthetic and real series, which are then compared using an objective function. As previously mentioned these approaches may suffer, on one side, from the arbitrary choice of moments and, on the other, from the poor representation that such statistics might offer about the behavior of complex time series.

In order to overcome such shortcomings, Lamperti (2017) has recently proposed a novel information theoretic criterion, called Generalized Subtracted L-divergence (*GSL-div*), that measures the degree of similarity between the dynamics observed in real data and those produced by the numerical simulation of a model. Contrary to simple summary statistics, the *GSL-div* has been constructed to compare time series on the basis of their patterns. Validation is achieved capturing the ability of a given model to reproduce the distributions of time changes (that is, changes in the process' values from one point in time to another) observed in the real-world series, without the need to resort to any likelihood function or to impose requirements of stationarity. The *GSL-div* provides a precise quantification of the distance between the model and data with respect to their dynamics in the time domain.[14]

The *GSL-div* can be estimated numerically following a simple, four-steps procedure.

1. Time series (both real and simulated) are symbolized.
2. Patterns of symbols are observed through rolling windows of different length $l = 1, .., L$.
3. Distributions of patterns, f_l, are estimated for each windows' length.
4. The distance between distributions from real and simulated data are evaluated through an information theoretic criterion and, finally, aggregated.

[13]Of course, many other institutions are contributing to the area, including but not limiting to the Catholic University of Milan (Italy), Polytechnic University of Marche (Italy), University of Bielefeld (Germany), University of Kiel (Germany), University of Turin (Italy), University of Kent (UK).

[14]An interesting similar criterion has been developed in Barde (2016b) for the class of Markov models.

Symbolization constrains a real-valued process to assume a finite set of ordered values $(1, 2, 3, ..., b)$, which are called symbols. Each observation is mapped to a symbol such that data are transformed into sequences of symbols having the same length as the original series. The precision of the symbolization, b, is controlled by the user.[15] The second step consists in the use of rolling windows of increasing length (l) to observe *words* of symbols in the synthetic and real series. For example, when $l = 2$, the word 12 indicates that the corresponding process has faced an up-ward movement. The choice of the symbolization precision allows the modeler to remove noise from the observations. For each $l = 1, ..., L$, one can obtain the distribution of words observed in the real and simulated data and to measure the discrepancy between the two. In particular, Lamperti (2017) uses a symmetric variant (Lin, 1991) of the Kullback and Leibler (1951)'s divergence. Finally, discrepancies can be aggregated over lengths of the words. Repeating these steps for multiple model runs (under the same parameter configuration), averaging and correcting for a systematic bias, one obtains a good approximation of the distance between the true probabilistic structure of the model and the data, as measured by the *GSL-div*.

The *GSL-div* has been tested to discriminate amongst different classes of stochastic processes, going from simple Auto Regressive Moving Average (ARMA) models to random walks with drifts and structural breaks. Systematic comparisons with alternative measures of fit commonly used for calibrating ABMs in economics and finance (e.g. mean squared error (MSE), distance between moments, etc.) has revealed that the *GSL-div* provides much more satisfactory performances (see table 1). Such results point the adequacy of the *GSL-div* to quantify the degree a simulation model mirrors real-world data.

Table 1: Comparison of model behavior through different criteria.

Class of models	Models in the contest	% of times true model identified			
		GSL	MSE	Dist. in Mean	Dist. in Variance
ARMA	9	88%	55%	44%	22%
Random Walks with Drift	12	92%	67%	58%	17%
Random Walks - break in Drift	12	100%	67%	50%	25%
Random Walks - break in Volatility	12	67%	58%	25%	33%

Note: the table reports the results of an exercise where each model has been used as a pseudo-real data generating process and compared with all the others in the same class. The number of models in each contest is reported in the second column. Different criteria have been employed to run the comparison and identify the closest to the pseudo-real data. GSL stands for Generalized Subtracted L-divergence. MSE stands for Mean Squared Error. Distance in mean consists in the difference between the sample mean in simulated and real data. The same apply for variance. Additional details in Lamperti (2017).

Lamperti (2016) applies the described approach to the analysis of a widely used financial market model with heterogeneous traders. He finds that the GSL-div can further improve the validation of the model with respect to criterion grounded on the minimization of the mean squared error as in Recchioni et al.

[15]The default value is five. See Lamperti (2017) for robustness tests.

(2015).

5.2 Validation as Matching of Causation

Since the foundation of the Econometric Society economists have been largely concerned with the identification of causal relationships among the variables characterizing our economic system. Discovering causal structures is a relevant task for at least two interconnected reasons: (i) it allows understanding and explaining the origin and propagation of phenomena that are observed at some point in time; (ii) it provides information on available policy channels to be used for impacting the system. Given such premises, Guerini and Moneta (2017) claim that models employed to provide policy prescriptions should match the causal relationships observed in the real systems they represent. They propose a procedure to validate a simulation model by estimating and comparing the causal structures incorporated in the model with those obtained from real-world data. In this manner, the procedure provides a possible solution to both the issues of confronting an ABM to empirical data and to compare different simulation models. Once a good matching between model-generated and real-world causal structures is achieved, the policy statements drawn from an ABM would get a more rigorous empirical support than those provided by the mere replication of stylized facts.

The causation matching approach proposed by Guerini and Moneta (2017) follows a sequential procedure that can be divided in five steps.

1. Data harmonization and preparation.
2. Analysis of ABM properties.
3. Estimation of Vector Auto Regressive (VAR) models.
4. Identification of the Structural Vector Autoregressive (SVAR) models.
5. Validation assessment.

In the first step, some simple transformations are performed to allow the comparison of empirical and artificial data (e.g. cutting simulated series to make them equally long as their real-world counterpart, removing trend, etc.). In the second step, the emergent properties of the series produced by the simulated model are analyses (e.g stationarity and ergodicity tests). In the third step, the reduced-form VAR model is estimated via ordinary least squares or accounting for co-integrated variables via the Johansen and Juselius (1990) procedure. In the fourth step, the structural form of the model is identified by means of the so-called PC (in case of Gaussian residuals, Spirtes et al., 2000) or VAR-LiNGAM (if residuals are non-Gaussian, Shimizu et al., 2006 and Hyvarinen et al., 2010) causal search algorithms.[16]. Finally, in the last step, the two estimated causal structures are compared according to simple distance measures.

The crucial part of such a procedure lies in the causal search step. In particular, residuals from the estimated VAR are collected and serve as input in the search for causal relationships, for both the PC and VAR-LiNGAM algorithms. When the PC is applied, feedbacks in the contemporaneous causal structure are excluded (dynamics feedbacks are of course conceivable). The PC algorithm also assumes causal sufficiency, i.e. there is no unmeasured variable which simultaneously affect two or more observed variables. The information obtained through the PC approach are generally not sufficient to provide full identification of the SVAR model, thereby requiring still a certain level of a priori theoretical knowledge. However, the VAR-LiINGAM algorithm solves the problem and provides a unique parameter set for the SVAR model simply requiring that residuals can be represented as a Linear Non-Gaussian Acyclic Model (LiNGAM), so that the contemporaneous causal structure can be thought as a set of directed relationships without cycles (further details in Guerini and Moneta, 2017 and Moneta et al., 2013). Once the causal structures are identified, a simple counter recording the percentage of the signs (positive or negative) of the causal relationships matched by the model is used for validation assessment. Alternatively, one can employ counters that account for the magnitude (size) of the relationships or both sign and magnitude.

Guerini and Moneta (2017) apply their approach to the well known K+S macroeconomic agent-based model developed in Dosi et al. (2015). Causal structures from model simulations are compared to those obtained from U.S. data for the period 1959-2014. Results show that the model is able to capture between 65% and 80% of the causal relations entailed by a SVAR estimated on real-world data (see table 2). Such a positive findings could be then compared to the results obtained when the procedure is also applied to different agent-based and DSGE models. In that, the causality-matching validation test is highly complementary to GSL-div employed to assess the replication of time series dynamics (cf. section 5.1 above).

5.3 Global Sensitivity Analysis via Kriging Meta-Modelling

Sensitivity analysis constitutes an open challenge for ABMs in economics and finance. The understanding of model's response to (possibly joint) changes in some parameter values or initial conditions is pivotal to assess the robustness of models' output as well as to draw robust implications from policy exercises. However, sensitivity analysis in ABM can often involve high computational costs stemming from simulating the model for many vectors of parameters, initial conditions and, possibly, seeds of the pseudo-random number generating process. Salle and Yıldızoğlu (2014) have been the first to propose the combination of

[16] VAR-LiGAM stands for Vector Auto Regressive Linear Non-Gaussian Acyclic Model.

Table 2: The table presents the performance of the K+S model (Dosi et al., 2015) to replicate causal links estimated from US macroeconomic data. The exercise has been repeated using different estimation methods (see first column). Mean (μ) and standard deviation (σ) of the three similarity measures (share of correctly replicated signs, share of correctly replicated magnitude, both) are reported.

Estimation Method	Similarity Type	μ	σ
VAR-OLS (all parameters)	sign-based	0.7892	0.0517
VECM-ML (all parameters)	sign-based	0.7385	0.0628
VAR-OLS (significant parameters)	sign-based	0.6490	0.1030
VECM-ML (significant parameters)	sign-based	0.7989	0.0689
VAR-OLS (all parameters)	size-based	0.7879	0.1073
VECM-ML (all parameters)	size-based	0.8111	0.0756
VAR-OLS (significant parameters)	size-based	0.7768	0.2783
VECM-ML (significant parameters)	size-based	0.8857	0.1535
VAR-OLS (all parameters)	conjunction	0.6748	0.0702
VECM-ML (all parameters)	conjunction	0.6323	0.1317
VAR-OLS (significant parameters)	conjunction	0.5928	0.2123
VECM-ML (significant parameters)	conjunction	0.7261	0.1881

design of experiments (DoE) and kriging meta-modelling to address the issue within the economics literature. The strategy they propose is straightforward. DoE allows to minimize the sample size of parameter configurations under the constraint on their representativeness. Based on the data collected through that sample, the original model is approximated with a meta-model, which is then employed to connect the parameters to the variables of interest at virtually zero computational costs.

Building on such an approach, Dosi et al. (2017c) provide a global sensitivity analysis for a relatively simple model of industry dynamics. Their procedure runs as follow:

1. employ nearly orthogonal latin hypercubes (NOLH) to sample the parameter space;
2. develop a Kriging meta-model (KMM) to approximate the original ABM;
3. perform Sobol variance decomposition to analyse the meta-model sensitivity to parameters;
4. draw three dimensional surfaces to represent the response of the variable of interest in the meta-model to changes in parameters.

The NOLH is a statistical technique for the generation of plausible sets of points from multidimensional parameter distributions exhibiting good space-filling properties (Cioppa and Lucas, 2007). It significantly improves the efficiency of the sampling process in comparison to traditional Monte Carlo approaches, requiring smaller samples and much less computation time to get an estimation of meta-model coefficients (Iooss et al., 2010). Despite being superior to Monte Carlo sampling, NOLH exhibit problems when the dimensionality of the parameter space is higher than forty.

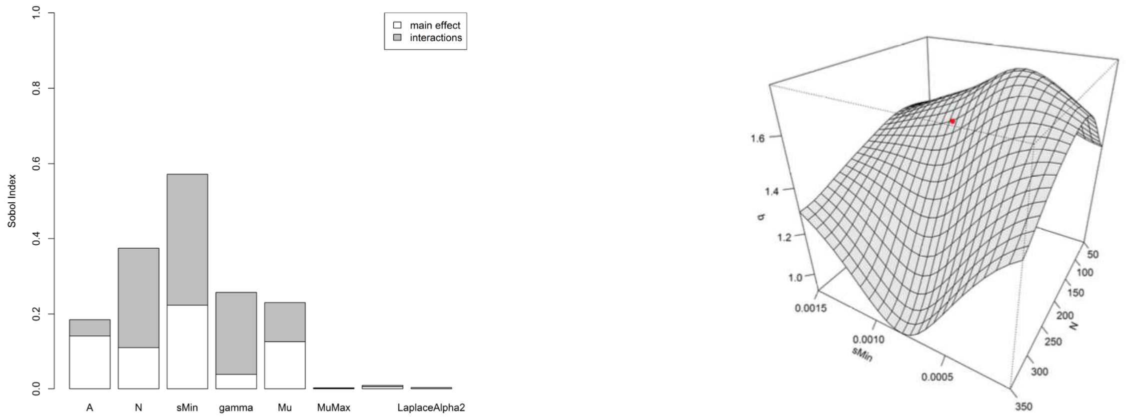
The meta-model is a simplified version of the original model that can be more parsimoniously run to evaluate the effect of inputs (parameters) on model's output. Kriging (or Gaussian process regression) is a simple and efficient method for investigating the behavior of simulation models (Krige, 1951; Van Beers and Kleijnen, 2004). The intuition behind it is that the original model's response to unknown points in the parameter space can be predicted by a linear combination of those observed in the closest points. As a consequence, the response function obtained through the KMM approach are a smooth approximation of the model output around the evaluation (known) points. Kriging meta-modeling provide the best linear unbiased prediction for such points (see ch. B.6 in Fischer and Getis, 2009). Coupling NOLH with KMM has been frequently used to approximate the output of computer simulation models (see e.g. McKay et al., 1979; Salle and Yıldızoğlu, 2014; Bargigli et al., 2016; Dosi et al., 2017b).

Sensitivity analysis has the objective of studying how uncertainty in the output of a model can be apportioned to different sources in its input (Saltelli et al., 2008). Given the constraints imposed by expensive simulations of a computational model, sensitivity analysis can be performed substituting the original model with its conveniently estimated meta-model (Kleijnen and Sargent, 2000; Wang and Shan, 2007). Following Saltelli et al. (2000), a Sobol decomposition form of variance-based global sensitivity analysis is selected. It decomposes the variance of a given output variable of the model in terms of the contributions of each input (parameter) variance (both in isolation and interacting it with every other input) by means of Fourier transformations. Sobol decomposition is attractive because it evaluates sensitivity across the whole parametric space and it allows for independent analyses of multiple output models (including non-linear and non-additive ones, see Saltelli and Annoni, 2010).

Finally, one can rank the importance of each parameter in explaining the variance of the output, and obtain response surfaces representing how the model (approximately) behave when input parameters are changed.

For example, Dosi et al. (2017c) study a model of industrial dynamics investigating how the distribution of firms' growth rates changes in response to different input variations ranging from the relevance of learning mechanisms to the strength of the selection process among competing firms. Figure 2 shows the response surface of the degree of skewness of firms' growth rates distribution (q) to the number of firm in the market (N) and the minimum market share ($sMin$) entailing the survival of a firm (right panel), and the relative importance of all model's parameter in affecting the selected output variable (left panel). Kriging and Sobol decompositions have also been successfully employed to the more complex K+S model agent-based model to study the impact of structural reforms in the labor market (Dosi et al., 2016b) and the emergence of hysteresis (Dosi et al., 2017b).

Figure 2: Sensitivity analysis and response surface from a Kriging meta-model. Source: Dosi et al. (2017c).



(a) Sensitivity analysis (Laplace)

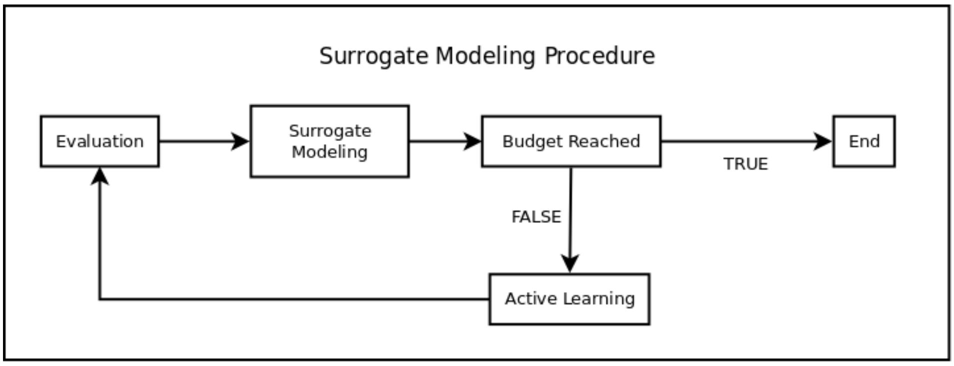
(b) 3D response surface (Laplace)

5.4 Parameter Space Exploration and Calibration via Machine Learning Surrogates

Kriging constitutes a valuable meta-modelling technique to approximate the behavior of an ABM in a given region of the parameter space, usually selected by the modeler because of some reasonable property or by economic intuition. However, one may need to extensively explore the parameter space to detect possible abrupt changes in the aggregate properties of the model or simply to have a general and precise overview of its behavior. Such broad explorations are usually infeasible in terms of computational costs. This potential issue is addressed in Lamperti et al. (2017b), who explicitly tackles parameter space exploration and calibration of ABMs combining supervised machine-learning and intelligent sampling to build a surrogate meta-model, which is then used to classify parameter vectors according to the behavior they produce. By providing a fast and accurate approximation of the original model behavior, the machine-learning surrogate dramatically reduce the computation time to perform large scale explorations of the parameter-space, while providing a powerful filter to gain insights into the complex functioning of agent-based models. Once the modeler has fixed the conditions that the output of her/his model should satisfy (e.g. some properties observed in real data), the surrogate single out the set of points in the parameter space that satisfy it.

The learning process of a surrogate occurs over multiple rounds (see figure 3). First, a large "pool" of parameterizations are drawn using a standard sampling routine, such as quasi-random Sobol sampling. Next, a very small subset of the pool is randomly drawn without replacement for evaluation in the original ABM, making sure to have at least one example of the user-desired behaviobehaviorr. These points are "labelled" according to the statistic measured on the output generated by the ABM and act as a "seed"

Figure 3: Surrogate modelling algorithm. Source: Lamperti et al. (2017b).



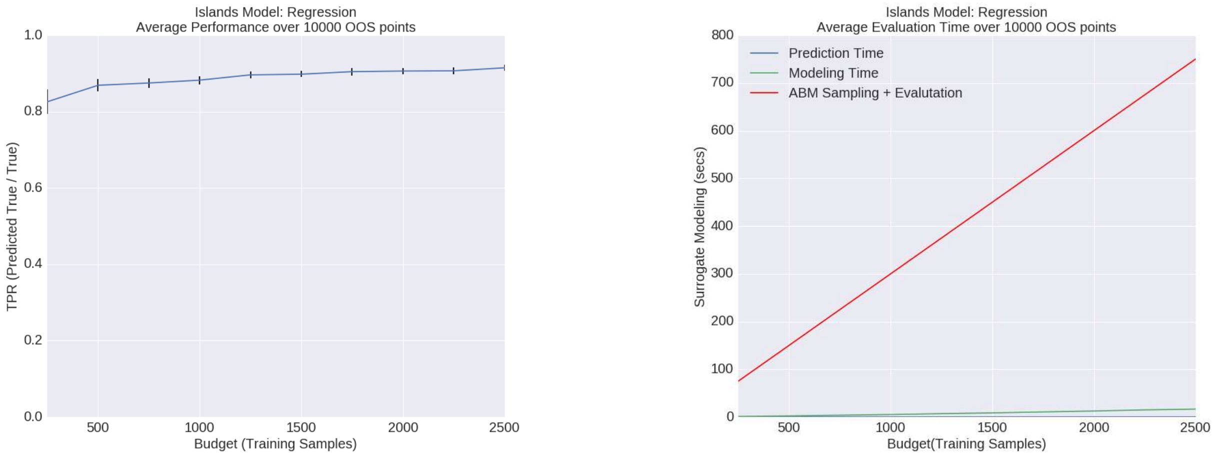
set of samples to initialize the surrogate model learned in the first round. This first surrogate is then exploited to predict the labels for unlabelled points remaining in the pool. Then, over multiple rounds, this process is repeated until a specified budget of evaluations is achieved. In each round, the surrogate directs which unlabelled points are drawn from the pool to maximize the performance of the surrogate learned in the next round. This semi-supervised "active" learning procedure incrementally improves the surrogate model, while maximizing the information gained over the ABM parameter space.

The crucial part of the job is finding a precise approximation of the original model, which has to be learnt over samples of points selected to minimize the computational effort of the overall procedure. In particular, the surrogate training procedure involves three decisions:

1. choose a machine-learning algorithm to act as a surrogate for the original ABM;
2. select a sampling procedure to draw samples from the parameters space in order to train the surrogatel;
3. select a score or criterion to evaluate the performance of the surrogate.

Extreme gradient boosted trees (XGBoost, see Chen and Guestrin, 2016) are used as the predefined surrogate learning algorithm employed to form a random ensemble of classification and regression trees (CART, cf. Breiman et al., 1984). This choice allows the surrogate to learn non-linear "knife-edge" properties, which typically characterize ABM parameter spaces. The sampling procedure builds a set of parameter vectors on which the agent-based model is actually evaluated in order to provide labelled data points for the training of the surrogate. Sets of parameter combinations are successively drawn according to a quasi-random Sobol sampling over the parameter space (Morokoff and Caflisch, 1994). Finally, the quality of the surrogate approximation is measured through the true positive ratio (TPR), a standard classification accuracy indicator computed as the number of parameter vectors correctly predicted (by the surrogate) to satisfy the user-specified conditions over the total number of parameters in the "pool" truly satisfying them.

Figure 4: Results from exploration of the parameter space for the Islands growth model. Source: Lamperti et al. (2017b).



Lamperti et al. (2017b) provide two applications of the machine-learning surrogate approach employing a financial ABM and a model of endogenous growth. Results are encouraging. In the growth model (Fagiolo and Dosi, 2003), parameter vectors delivering endogenous growth cum fat-tailed output growth-rate distributions are selected. growth rates exhibit fat-tails. They find that even for limited budget, the surrogate correctly classifies more than 80% parameter combinations (cf. figure 4, left panel) and computational costs are extraordinarily lower than those required by the original ABM (see figure 4, right panel).

6 Conclusions

Ten years after the influential article by Windrum et al. (2007), the issue of empirical validation of agent-based models (ABM) in economics and finance is still among the top items in the to-do list of researchers. This despite the fact that many advances have occurred, especially in the three key areas of: (i) calibration and estimation of model parameters; (ii) comparison of real world and artificial data; and (iii) parameter space exploration.

This Chapter has attempted to critically survey such a recent literature focusing on developments in the above three areas.

Notwithstanding the huge effort made in advancing the frontier in ABM validation techniques, the process of developing a complete and coherent validation toolbox is still on-going and some important issues are still to be better understood.

First, the pros and cons of each different validation methodology are still not completely laid out in the literature. This is a pity, as a sort of if-then map would be extremely useful for practitioners aiming at picking the right tool in each specific situation. Projects developing such a map would be very welcome in the community, although it is of course clear that each tools is aimed at a specific task, and no validation technique is more general than the others. Relatedly, validation software packages should be developed to ease the adoption of the different existing techniques.

Second, and most importantly, more research efforts should be devoted towards advancing hypothesis testing in ABM. In particular, more robust statistical tests should be developed in order to better characterizing model stationarity and ergodicity, and to better understand how the failure of these properties might affect the problems of estimation, calibration, validation and exploration.

To conclude, we believe that the development of better empirical-validation techniques is a never-ending process, which must naturally co-evolve together with the developments of new models, new statistical techniques and with the increase in computational power. In that, recent developments in machine-learning and the increase availability of big data could entail the next leap forward: machine learning offers indeed more flexible methodologies that allow one to minimize the number of assumptions when running an econometric model; big data, instead, allow one to perform more thorough comparisons of the model with the real world situations, by extending validation also to the micro-level. All in all, these extensions would allow ACE models to progress from Level 2 to Level 3 in the Axtell and Epstein (1994) classification (see Section 4).

Furthermore, validation of ABMs will never tell whether a model is a correct description of the complex, unknown and non-understandable real-world data generating process. However, in a Popperian fashion, ABM validation techniques should eventually allow researchers to understand whether a model is a bad description of it.

References

Simone Alfarano, Thomas Lux, and Friedrich Wagner. Estimation of Agent-Based Models: The Case of an Asymmetric Herding Model. *Computational Economics*, 26(1):19–49, 2005. doi: 10.1007/s10614-005-6415-1. URL https://ideas.repec.org/a/kap/compec/v26y2005i1p19-49.html.

Simone Alfarano, Thomas Lux, and Friedrich Wagner. Estimation of a simple agent-based model of financial markets: An application to Australian stock and foreign exchange data. *Physica A: Statistical Mechanics and its Applications*, 370(1):38–42, 2006. doi: 10.1016/j.physa.2006.04.0. URL https://ideas.repec.org/a/eee/phsmap/v370y2006i1p38-42.html.

Mikhail Anufriev, Te Bao, and Jan Tuinstra. Microfoundations for switching behavior in heterogeneous agent models: An experiment. *Journal of Economic Behavior & Organization*, 129(C):74–99, 2016. doi: 10.1016/j.jebo.2016.06.00. URL https://ideas.repec.org/a/eee/jeborg/v129y2016icp74-99.html.

Mikhail Anufriev and Cars Hommes. Evolutionary Selection of Individual Expectations and Aggregate Outcomes in Asset Pricing Experiments. *American Economic Journal: Microeconomics*, 4(4):35–64, November 2012. URL https://ideas.repec.org/a/aea/aejmic/v4y2012i4p35-64.html.

Tiziana Assenza, Domenico Delli Gatti, and Jakob Grazzini. Emergent dynamics of a macroeconomic agent based model with capital and credit. *Journal of Economic Dynamics and Control*, 50(C):5–28, 2015. doi: 10.1016/j.jedc.2014.07.00. URL https://ideas.repec.org/a/eee/dyncon/v50y2015icp5-28.html.

Tiziana Assenza, Peter Heemeijer, Cars Hommes, and Domenico Massaro. Individual Expectations and Aggregate Macro Behavior. Tinbergen Institute Discussion Papers 13-016/II, Tinbergen Institute, January 2013. URL https://ideas.repec.org/p/tin/wpaper/20130016.html.

Robert Axelrod. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, 1997.

Robert L Axtell and Joshua M Epstein. Agent-based modeling: Understanding our creations. *The Bulletin of the Santa Fe Institute*, 9(2):28–32, 1994.

Sylvain Barde. Direct comparison of agent-based models of herding in financial markets. *Journal of Economic Dynamics and Control*, 73(C):329–353, 2016a. URL https://ideas.repec.org/a/eee/dyncon/v73y2016icp329-353.html.

Sylvain Barde. A practical, accurate, information criterion for nth order markov processes. *Computational Economics*, pages 1–44, 2016b.

Sylvain Barde and Sander van der Hoog. An empirical validation protocol for large-scale agent-based models. Studies in Economics 1712, School of Economics, University of Kent, July 2017. URL https://ideas.repec.org/p/ukc/ukcedp/1712.html.

Leonardo Bargigli, Luca Riccetti, Alberto Russo, and Mauro Gallegati. Network calibration and metamodeling of a financial accelerator agent based model. Technical report, Università Politecnica delle Marche, 2016.

Stefano Battiston, J. Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G. Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. Complexity theory and financial regulation. *Science*, 351(6275):818–819, 2016. ISSN 0036-8075. doi: 10.1126/science.aad0299. URL http://science.sciencemag.org/content/351/6275/818.

Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro Vagliasindi. Validating and calibrating agent-based models: A case study. *Computational Economics*, 30:245–264, July 2007.

Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro Vagliasindi. Validation in agent-based models: An investigation on the cats model. *Journal of Economic Behavior & Organization*, 67:947–964, 2008a.

Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro A. Vagliasindi. Validation in agent-based models: An investigation on the CATS model. *Journal of Economic Behavior & Organization*, 67(3-4):947–964, September 2008b. URL https://ideas.repec.org/a/eee/jeborg/v67y2008i3-4p947-964.html.

H. Peter Boswijk, Cars H. Hommes, and Sebastiano Manzan. Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938–1970, June 2007. URL https://ideas.repec.org/a/eee/dyncon/v31y2007i6p1938-1970.html.

Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

Thomas Brenner and Claudia Werker. A taxonomy of inference in simulation models. *Computational Economics*, 30(3):227–244, 2007. ISSN 1572-9974. doi: 10.1007/s10614-007-9102-6. URL http://dx.doi.org/10.1007/s10614-007-9102-6.

William A. Brock. Scaling in Economics: A Reader's Guide. *Industrial and Corporate Change*, 8(3):409–446, September 1999.

William A. Brock and Cars H. Hommes. A rational route to randomness. *Econometrica*, 65(5):1059–1095, 1997. URL http://www.jstor.org/stable/2171879.

William A. Brock and Cars H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22(8-9):1235–1274, August 1998. URL https://ideas.repec.org/a/eee/dyncon/v22y1998i8-9p1235-1274.html.

Fabio Canova and Luca Sala. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449, May 2009. URL https://ideas.repec.org/a/eee/moneco/v56y2009i4p431-449.html.

Shu-Heng Chen, Chia-Ling Chang, and Ye-Rong Du. Agent-based economic models and econometrics. *The Knowledge Engineering Review*, 27(2):187–219, 2012. doi: 10.1017/S0269888912000136.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

Carl Chiarella, Xue-Zhong He, and Remco C.J. Zwinkels. Heterogeneous expectations in asset pricing: Empirical evidence from the S&P500. *Journal of Economic Behavior & Organization*, 105(C):1–16, 2014. URL https://ideas.repec.org/a/eee/jeborg/v105y2014icp1-16.html.

Tommaso Ciarli. Structural interactions and long run growth: an application of experimental design to agent-based models. *Revue de l'OFCE*, 124:295 – 345, 2012.

Thomas M. Cioppa and Thomas W. Lucas. Efficient nearly orthogonal and space-filling latin hypercubes. *Technometrics*, 49(1):45–55, 2007. ISSN 00401706. URL http://www.jstor.org/stable/25471274.

Herbert Dawid, Philipp Harting, Sander van der Hoog, and Michael Neugart. A Heterogeneous Agent Macroeconomic Model for Policy Evaluation: Improving Transparency and Reproducibility, 2016.

Eelke de Jong, Willem F.C. Verschoor, and Remco C.J. Zwinkels. Heterogeneity of agents and exchange rate dynamics: Evidence from the EMS. *Journal of International Money and Finance*, 29(8):1652–1669, December 2010. URL https://ideas.repec.org/a/eee/jimfin/v29y2010i8p1652-1669.html.

Marco Del Negro and Frank Schorfheide. How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models. *Economic Review*, (Q 2):21–37, 2006. URL https://ideas.repec.org/a/fip/fedaer/y2006iq2p21-37nv.91no.2.html.

G. Dosi, M. Pereira, A. Roventini, and M. E. Virgilito. When more flexibility yields more fragility: the microfoundations of keynesian aggregate unemployment. *Journal of Economic Dynamics & Control*, 81:162–186, 2017a.

Giovanni Dosi, Giogio Fagiolo, Mauro Napoletano, and Andrea Roventini. Income distribution, credit and fiscal policies in an agent-based keynesian model. *Journal of Economic Dynamics and Control*, 37(8):1598–1625, August 2013.

Giovanni Dosi, Giorgio Fagiolo, Mauro Napoletano, Andrea Roventini, and Tania Treibich. Fiscal and monetary policies in complex evolving economies. *Journal of Economic Dynamics and Control*, 52:166–189, March 2015.

Giovanni Dosi, Giorgio Fagiolo, and Andrea Roventini. Schumpeter meeting keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control*, 34(9):1748 – 1767, June 2010.

Giovanni Dosi, Mauro Napoletano, Andrea Roventini, and Tania Treibich. Micro and macro policies in the Keynes+Schumpeter evolutionary models. *Journal of Evolutionary Economics*, forthcoming:1–28, 2016a. ISSN 1432-1386. doi: 10.1007/s00191-016-0466-4. URL http://dx.doi.org/10.1007/s00191-016-0466-4.

Giovanni Dosi, Marcelo C Pereira, Andrea Roventini, and Maria Enrica Virgillito. The effects of labour market reforms upon unemployment and income inequalities: an agent based model. LEM Working Papers 2016/27, Scuola Superiore Sant'Anna, 2016b.

Giovanni Dosi, Marcelo C Pereira, Andrea Roventini, and Maria Enrica Virgillito. Causes and consequences of hysteresis: Aggregate demand, productivity and employment. LEM Working Papers 2017/07, Scuola Superiore Sant'Anna, 2017b.

Giovanni Dosi, Marcelo C Pereira, and Maria Enrica Virgillito. On the robustness of the fat-tailed distribution of firm growth rates: a global sensitivity analysis. *Journal of Economic Interaction and Coordination*, pages 1–21, 2017c.

Joshua M. Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

Annalisa Fabretti. On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, 8(2):277–293, 2013.

G. Fagiolo and G. Dosi. Exploitation, exploration and innovation in a model of endogenous growth with locally interacting agents. *Structural Change and Economic Dynamics*, 14:237–73, 2003.

Giorgio Fagiolo and Andrea Roventini. Macroeconomic Policy in DSGE and Agent-Based Models. *Revue de l'OFCE*, 0(5):67–116, 2012. URL https://ideas.repec.org/a/cai/reofsp/reof_124_0067.html.

Giorgio Fagiolo and Andrea Roventini. Macroeconomic Policy in DSGE and Agent-Based Models Redux: New Developments and Challenges Ahead. *Journal of Artificial Societies and Social Simulation*, 20(1), 2017. URL https://ideas.repec.org/a/jas/jasssj/2016-78-2.html.

Doyne J. Farmer and Duncan Foley. The economy needs agent-based modelling. *Nature*, 460:685–686, 2009.

Jesús Fernández-Villaverde, Juan F. Rubio Ramírez, and Frank Schorfheide. Solution and Estimation Methods for DSGE Models. NBER Working Papers 21862, National Bureau of Economic Research, Inc, January 2016. URL https://ideas.repec.org/p/nbr/nberwo/21862.html.

Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez. Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies*, 74(4):1059–1087, 2007. URL https://ideas.repec.org/a/oup/restud/v74y2007i4p1059-1087.html.

Manfred M Fischer and Arthur Getis. *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Science & Business Media, 2009.

Reiner Franke. Applying the method of simulated moments to estimate a small agent-based asset pricing model. *Journal of Empirical Finance*, 16(5):804–815, December 2009. URL https://ideas.repec.org/a/eee/empfin/v16y2009i5p804-815.html.

Reiner Franke and Frank Westerhoff. Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control*, 36(8):1193–1211, 2012. URL https://ideas.repec.org/a/eee/dyncon/v36y2012i8p1193-1211.html.

Edoardo Gaffeo, Domenico Delli Gatti, Saul Desiderio, and Mauro Gallegati. Adaptive Microfoundations for Emergent Macroeconomics. *Eastern Economic Journal*, 34(4):441–463, 2008. URL https://ideas.repec.org/a/pal/easeco/v34y2008i4p441-463.html.

David Goldbaum and Bruce Mizrach. Estimating the intensity of choice in a dynamic mutual fund allocation decision. *Journal of Economic Dynamics and Control*, 32(12):3866–3876, December 2008. URL https://ideas.repec.org/a/eee/dyncon/v32y2008i12p3866-3876.html.

C Gourieroux, A Monfort, and E Renault. Indirect Inference. *Journal of Applied Econometrics*, 8(S):85–118, 1993. URL https://ideas.repec.org/a/jae/japmet/v8y1993isps85-118.html.

Jakob Grazzini and Matteo Richiardi. Estimation of ergodic agent-based models by simulated minimum distance. *Journal of Economic Dynamics and Control*, 51(C):148–165, 2015. doi: 10.1016/j.jedc.2014.10.00. URL https://ideas.repec.org/a/eee/dyncon/v51y2015icp148-165.html.

Jakob Grazzini, Matteo G. Richiardi, and Mike Tsionas. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77(C):26–47, 2017. URL https://ideas.repec.org/a/eee/dyncon/v77y2017icp26-47.html.

Mattia Guerini. Is the friedman rule stabilizing? some unpleasant results in a heterogeneous expectations framework. *Department of Economics and Finance Working Papers, Unicatt (Milan)*, (03), 2013.

Mattia Guerini and Alessio Moneta. A method for agent-based models validation. *Journal of Economic Dynamics and Control*, 2017.

Mattia Guerini, Mauro Napoletano, and Andrea Roventini. No man is an island: The impact of heterogeneity and local interactions on macroeconomic dynamics. *Economic Modelling*, 2017. ISSN 0264-9993. doi: https://doi.org/10.1016/j.econmod.2017.05.004. URL http://www.sciencedirect.com/science/article/pii/S0264999317307411.

Lars Peter Hansen and James J Heckman. The empirical foundations of calibration. *The Journal of Economic Perspectives*, 10(1):87–104, 1996.

Samer Hassan, Juan Pavon, and Nigel Gilbert. Injecting data into simulation: Can agent-based modelling learn from microsimulation. In *World Congress of Social Simulation*, 2008.

Cars Hommes. The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, 35(1):1–24, January 2011. URL https://ideas.repec.org/a/eee/dyncon/v35y2011i1p1-24.html.

Cars Hommes. *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Number 9781107564978 in Cambridge Books. Cambridge University Press, February 2013. URL https://ideas.repec.org/b/cup/cbooks/9781107564978.html.

Aapo Hyvarinen, Kun Zhang, Shohei Shimizu, and Patrick O. Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11:1709–1731, 2010.

Bertrand Iooss, Loïc Boussouf, Vincent Feuillard, and Amandine Marrel. Numerical studies of the metamodel fitting and validation processes. *arXiv preprint arXiv:1001.1049*, 2010.

Soren Johansen and Katerina Juselius. Maximum likelihood estimation and inference on cointegration. with application to the demand for money. *Oxford Bullettin of Economics and Statistics*, 52:169 – 210, 1990.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–291, 1979.

Alan Kirman. *Epidemics of opinion and speculative bubbles in financial markets*, pages 354–368. Blackwell, 1991.

Jack PC Kleijnen and Robert G Sargent. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120(1):14–29, 2000.

Daniel G Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.

Jiri Kukacka and Jozef Barunik. Estimation of financial agent-based models with simulated maximum likelihood. 2016.

S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:49–86, 1951.

Francesco Lamperti. Empirical validation of simulated models through the gsl-div: an illustrative application. Technical report, LEM Working Papers, Scuola Superiore Sant'Anna, 2016.

Francesco Lamperti. An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics*, 2017.

Francesco Lamperti, Giovanni Dosi, Mauro Napoletano, Andrea Roventini, and Alessandro Sapio. Faraway, so Close: Coupled Climate and Economic Dynamics in an Agent-Based Integrated Assessment Model. LEM Papers Series 2017/12, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, March 2017a. URL https://ideas.repec.org/p/ssa/lemwps/2017-12.html.

Francesco Lamperti, Andrea Roventini, and Amir Sani. Agent-Based Model Calibration using Machine Learning Surrogates. LEM Papers Series 2017/11, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, March 2017b. URL https://ideas.repec.org/p/ssa/lemwps/2017-11.html.

David A. Lane. Artificial worlds and economics, part ii. *Journal of Evolutionary Economics*, 3(3):177–197, Sep 1993. doi: 10.1007/BF01200867.

Sandrine Jacob Leal, Mauro Napoletano, Andrea Roventini, and Giorgio Fagiolo. Rock around the clock: An agent-based model of low- and high-frequency trading. *Journal of Evolutionary Economics*, 26(1):49–76, March 2016. URL https://ideas.repec.org/a/spr/joevec/v26y2016i1p49-76.html.

Blake LeBaron and Leigh Tesfatsion. Modeling macroeconomies as open-ended dynamic systems of interacting agents. *American Economic Review*, 98(2):246–50, May 2008. doi: 10.1257/aer.98.2.246. URL http://www.aeaweb.org/articles?id=10.1257/aer.98.2.246.

Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behrooz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey Voinov, J. Gary Polhill, Zhanli Sun, and Dawn C. Parker. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, 18(4):4, 2015. doi: 10.18564/jasss.2897. URL http://jasss.soc.surrey.ac.uk/18/4/4.html.

Roberto Leombruni, Matteo Richiardi, Nicole J. Saam, and Michele Sonnessa. A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation*, 9(1):15, 2006. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/9/1/15.html.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37: 145–151, 1991.

Franco Malerba, Richard Nelson, Luigi Orsenigo, and Sidney Winter. 'history-friendly' models of industry evolution: the computer industry. *Industrial and Corporate Change*, 8(1):3, 1999. doi: 10.1093/icc/8.1.3. URL +http://dx.doi.org/10.1093/icc/8.1.3.

S.M. Manson. *Validation and verification of multi-agent systems, in Complexity and Ecosystem Management*. Cheltenham: Edward Elgar, edited by m.a. janssen edition, 2002.

Robert Marks. Validating Simulation Models: A General Framework and Four Applied Examples. *Computational Economics*, 30(3):265–290, October 2007. doi: 10.1007/s10614-007-9101-7. URL https://ideas.repec.org/a/kap/compec/v30y2007i3p265-290.html.

Robert E Marks. Validation and model selection: Three similarity measures compared. *Complexity Economics*, 2(1): 41–61, 2013.

Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

N. Metropolis and S. Ulam. The monte carlo method. *Journal of American Statistical Association*, 44:335–341, 1949.

Alessio Moneta, Doris Entner, Patrick Hoyer, and Alex Coad. Causal inference by independent component analysis: Theory and applications. *Oxford Bullettin of Economics and Statistics*, 75(5):705–730, 2013.

William J Morokoff and Russel E Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251–1279, 1994.

Alessia Paccagnini. DSGE Model Validation in a Bayesian Framework: an Assessment. MPRA Paper 24509, University Library of Munich, Germany, May 2010. URL https://ideas.repec.org/p/pra/mprapa/24509.html.

Paolo Pellizzari and Arianna Dal Forno. A comparison of different trading protocols in an agent-based market. *Journal of Economic Interaction and Coordination*, 2(1):27–43, June 2007. doi: 10.1007/s11403-006-0016-5. URL https://ideas.repec.org/a/spr/jeicoo/v2y2007i1p27-43.html.

Donovan Platt and Tim Gebbie. Can Agent-Based Models Probe Market Microstructure? Papers 1611.08510, arXiv.org, November 2016. URL https://ideas.repec.org/p/arx/papers/1611.08510.html.

Lilit Popoyan, Mauro Napoletano, and Andrea Roventini. Taming macroeconomic instability: Monetary and macro-prudential policy interactions in an agent-based model. *Journal of Economic Behavior & Organization*, 134(C): 117–140, 2017. doi: 10.1016/j.jebo.2016.12.01. URL https://ideas.repec.org/a/eee/jeborg/v134y2017icp117-140.html.

Maria Cristina Recchioni, Gabriele Tedeschi, and Mauro Gallegati. A calibration procedure for analyzing stock price dynamics in an agent-based framework. *Journal of Economic Dynamics and Control*, 60:1 – 25, 2015. ISSN 0165-1889. doi: http://dx.doi.org/10.1016/j.jedc.2015.08.003. URL http://www.sciencedirect.com/science/article/pii/S0165188915001517.

Robert Rosen. *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations.* Oxford: Pergamon, 1985.

Isabelle Salle and Murat Yıldızoğlu. Efficient sampling and meta-modeling for computational economic models. *Computational Economics*, 44(4):507–536, 2014.

Andrea Saltelli and Paola Annoni. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517, 2010.

Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer.* John Wiley & Sons, 2008.

Andrea Saltelli, Stefano Tarantola, and Francesca Campolongo. Sensitivity analysis as an ingredient of modeling. *Statistical Science*, pages 377–395, 2000.

Thomas C. Schelling. Models of segregation. *The American Economic Review*, 59(2):488–493, 1969. ISSN 00028282. URL http://www.jstor.org/stable/1823701.

Thomas C. Schelling. Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2):143–186, 1971. doi: 10.1080/0022250X.1971.9989794.

Davide Secchi and Raffaello Seri. Controlling for false negatives in agent-based models: a review of power analysis in organizational research. *Computational and Mathematical Organization Theory*, 23(1):94–121, 2017.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvarinen, and Antti J. Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

Herbert A. Simon. Bounded rationality and organizational learning. *Organization Science*, 2(1):125–134, 1991. doi: 10.1287/orsc.2.1.125.

Peter Spirtes, Clark Glymur, and Richard Scheines. *Causation, Prediction, and Search.* MIT Press, 2000.

Leigh Tesfatsion. Chapter 16 agent-based computational economics: A constructive approach to economic theory. *Handbook of Computational Economics*, 2:831 – 880, 2006. ISSN 1574-0021. doi: http://dx.doi.org/10.1016/S1574-0021(05)02016-2.

A. Turrell. Agent-based models: understanding the economy from the bottom up. Quarterly bulletin Q4, Bank of England, 2016.

Wim CM Van Beers and Jack PC Kleijnen. Kriging interpolation in simulation: a survey. In *Simulation Conference, 2004. Proceedings of the 2004 Winter*, volume 1. IEEE, 2004.

G Gary Wang and Songqing Shan. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design*, 129(4):370–380, 2007.

Claudia Werker and Thomas Brenner. Empirical calibration of simulation models. Papers on Economics and Evolution 0410, 2004. URL http://hdl.handle.net/10419/88244.

Frank H. Westerhoff and Roberto Dieci. The effectiveness of keynes–tobin transaction taxes when heterogeneous agents can trade in different markets: A behavioral finance approach. *Journal of Economic Dynamics and Control*, 30(2):293 – 322, 2006. ISSN 0165-1889. doi: http://doi.org/10.1016/j.jedc.2004.12.004. URL http://www.sciencedirect.com/science/article/pii/S0165188905000345.

Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8, 2007. URL http://jasss.soc.surrey.ac.uk/10/2/8.html.

Peter Winker and Manfred Gilli. Validation of agent-based models of financial markets. *IFAC Proceedings Volumes*, 34(20):401–406, 2001.

Peter Winker and Manfred Gilli. Applications of optimization heuristics to estimation and modelling problems. *Computational Statistics & Data Analysis*, 47(2):211–223, September 2004. URL https://ideas.repec.org/a/eee/csdana/v47y2004i2p211-223.html.