CEF pre-conference workshop
Calibration and Validation of ABMs

Francesco Lamperti

July 03, 2023

Sant'Anna School of Advanced Studies - SSSA
and
RFF-CMCC European Institute on Economics and the Environment - EIEE

Agenda of the workshop

- 1st Half (8:30-10:30): The basics
    - General introduction to ABM models
    - Validation challenges and approaches
    - "Simple" estimation methods: simulates method of moments, kernel density-based approaches
- 2nd Half (10:45-12:15): Scaling up
    - Meta-modelling and budgeted search
    - Bayesian estimation with Gaussian surrogates

- **Instructors:**
    - Sylvain Barde (University of Kent)
    - Francesco Lamperti (Sant'Anna Pisa and EIEE)

Table of Contents

Key references

- Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., & Roventini, A. (2019). Validation of agent-based models in economics and finance. Computer simulation validation: fundamental concepts, methodological frameworks, and philosophical perspectives, 763-787.

- Lux, T., & Zwinkels, R. C. (2018). Empirical validation of agent-based models. In Handbook of computational economics (Vol. 4, pp. 437-488). Elsevier.

Computational economics, complexity and ABMs

Computational economics

- **Computational Economics** (CE) is a field of economic theory that explores the intersections between *economics* and *computation*

- In economics and econometrics, we study objects called models to describe economies and learn about their properties

- Sometimes these models can be *analytically solvable*, but often they can not. In those cases we need computation to solve them

- **Computing** means using algorithms (written in a code) that can be employed to solve problems that do not lend themselves to analytical solutions

Economics vis-a-vis other disciplines
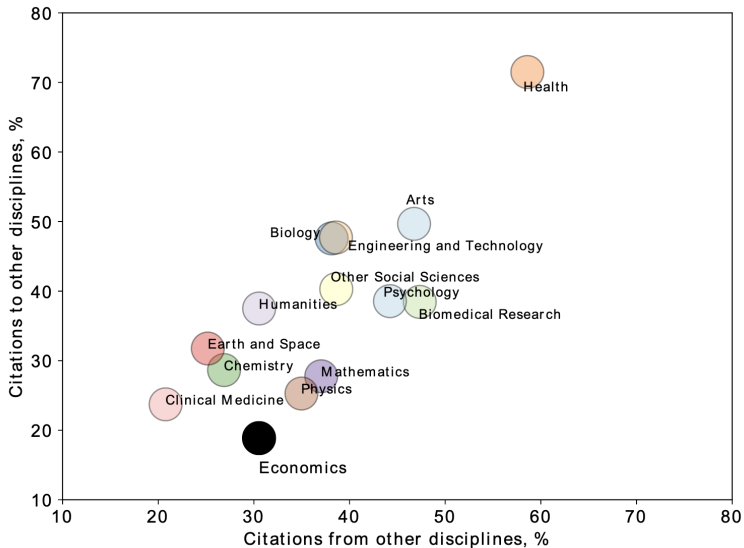
Agent-Based Computational Economics

- Most applications of CE are "*numerical*": we use computation to find solutions to analytically-stated problems that otherwise cannot be analytically solved
    - equilibrium relations in economic models
    - equilibrium solutions of economic models
    - properties of estimators in econometric models
    - ...

- This is not what agent-based modelling is about! Rather, Agent-based Computational Economics (ACE) deals with "**computationally constructive**" approaches to economics

- ACE relies on computations that can help economists to build from scratch and then analyze models of the economy

ACE

- **What is ACE?** ACE is the computational study of economic processes modeled as dynamic systems of interacting agents (Tesfatsion, 2006, p. 835)

- **Stress on:**
    - Computational study
    - Realistic micro behaviours
    - Dynamic systems of Interacting Agents
    - Emergent macro properties

- **In other words...** ACE is the computational study of economies as *complex evolving systems*

Complex systems and ABMs

- Many definitions for "complex system"

- One of them:
  "A system that involves **numerous interacting agents** whose aggregate behaviors are to be understood. Such **aggregate activity** is nonlinear, hence it **cannot simply be derived from summation of individual components behavior**" (Jerome Singer)

- **Agent-based (economic) models (ABMs)** are collections of behavioural rules and interaction rules which allow simulating the economy from the bottom-up (i.e. from micro to macro)

- ABMs are tools for the analysis of complex systems

Aggregation and feedbacks

- ABMs typically entail
  - Micro-to-macro aggregation
  - Macro-to-micro feedbacks
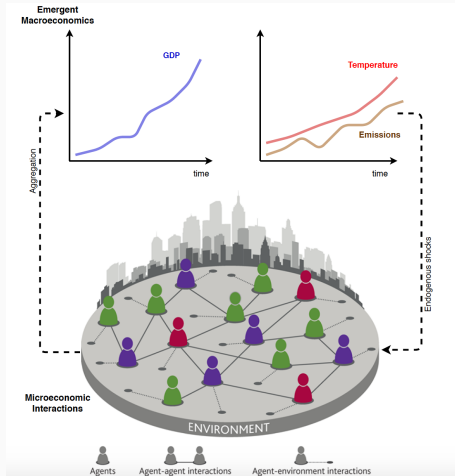
The convenience of linearity

- Consider a system S made of N units and evolving over time t=1,2,3,...

- We observe a property x on each unit $i = 1, 2, 3, ..., N$ at every time period t

- A unit i whose property is x at time produces a "behavior" $f(x_i, x_{-i}, t)$, where $f(\cdot)$ is a function

- The *average state of the system* at time t is the the mean of observed $x_{i,t}$:

$$\bar{x} = \frac{1}{N} \sum_i x_{i,t}$$

- average behaviour VS behaviour of the average; in general:

$$\frac{1}{N} \sum_i f(x_{i,t}, x_{-i,t}) \neq f(\bar{x})$$

- special case giving equality: linearity of $f(\cdot)$

Key features of ABMs

- **A tool to model economies where agents...**

    - Are boundedly-rational entities
    - Directly interact in non trivial networks
    - Might be persistently heterogeneous (e.g. in temrs of expectations)
    - Influence each other (within and outside markets)

- **A bottom-up approach ...**

    1. Modeling agents behaviors and their interactions first
    2. No top-down constraints such as representative agent, consistency of beliefs, rational expectations, equilibrium, ...
    3. Let it run and see what happens
    4. Statistical analysis of models output
    5. Matching with empirical data

Key features of ABM's output

- **Stochastic** (e.g. agents' behaviour has a random component)

- **Non-linear**, in different ways
    - the average behaviour is NOT the average of behaviours
    - relationships between aggregate variables may NOT be approximated by a linear system
    - threshold effects, multiple "steady states", "self-organized criticality"

- **Path-dependent**
    - initial conditions matter
    - history of state of the worlds matter

ACE models and their critiques

Structural critiques to ACE models

- no optimizing behaviour;
- no rational expectations (issues with the Lucas critique);
- no required partial/general equilibrium;
- lack of clarity in transmission mechanisms;

Are critiques based on these arguments well-founded?

- not in general (Gigerenzer and Brighton, 2009; Kirman, 1989, 1992);
- only partially (Moneta, 2005; Meisser, 2017)
-

A discussion on the primitives

The departure from the axiomatic approach and the optimizing behaviour casts doubts on ACE models from a model-consistency perspective. ACE abandons the *primitives* of traditional economic models.

But once the axioms are abandoned, what psychological motives determine agents' choices?

Heuristics

Any approach to problem solving or self-discovery that employs a practical method that is not guaranteed to be optimal, perfect, or rational, but is nevertheless sufficient for reaching an immediate, short-term goal or approximation. (see also Gigerenzer and Brighton, 2009).

In ACE models, heuristics are the *psychological primitives*. But they suffer of the wilderness of bounded rationality problem (i.e. which heuristic to choose?)

An expedition in the wilderness of bounded rationality

How to select among all the possible heuristics?

> *There is only one way to be rational. But there are infinitely many different ways to be irrational.*
>
> *Hommes (2013)*

Developing sub-literatures in the behavioural economics domain:

- empirical micro evidence (Dosi, 2004);
- management science approach (Dawid and Harting, 2012);
- experimental economics (Hommes, 2013).

Dawid and Delli Gatti (2018) have recently classified the behavioural rules adopted in ABM under a bunch of common general principles.

The inductive nature of ACE models

Methodological distinction

- *traditional models* follow the deductive method: from the axioms and the optimization problem, result are deduced under the form of theorems and proofs;

- *ACE models* follow the inductive method: the behavioural rules induce the generation of aggregate emergent properties, to which an economic argumentation can only be attached by means of statistical or anecdotal evidence.

ACE models do not display a homotetic behaviour between micro- and macro- dynamics. This creates a difficulty in explaining the results from an Agent-Based Model (ABM).

Technical critiques to ACE models

Aside from the fundamentals differences, there are some technical concerns:

- large numbers of free parameters – somehow too many;
- simulations with not credible values;

The last two are **technical and valuable critiques** (acknowledged also by AB modellers)

However:

- many parameters in ABM are *deep* ;
- there exist tools to estimate ACE models;
- ABM are not the unique ones facing these problems (Fagiolo and Roventini, 2017).

Towards validation of ABMs

Verification and validation (V&V)

Model Verification

Replies to the question: *are we building the model right?*

It is a formal requirement on the efficacy of the code to be correctly representing the conceptual model under study (i.e. the economic forces at stake) (Yilmaz, 2006; Naylor and Finger, 1967). All models should be verified.

Model Validation

Replies to the question: *are we building the right model?*

It consists upon measuring and evaluating the extent to which the model that we have designed is a good representation of the reality (Marks, 2013). Not all models are built to be validated.

Parameter validation
Replies to the question: *Are the exogenous inputs for the model empirically meaningful and appropriate for the purpose at hand?* Examples of exogenous model inputs include initial state conditions, functional forms, random shock realizations, data-based parameter estimates, and/or parameter values imported from other studies.

Process validation
Replies to the question: How well do the physical, biological, institutional, and social processes represented within the model reflect real-world aspects important for the purpose at hand? Are all process specifications consistent with essential scaffolding constraints, such as physical laws, stock-flow relationships, and accounting identities?

The choice of parameters: Calibration and estimation

Model Calibration
Finding the model parameters' values that allow the model to resemble as close as possible some properties of the data (ex-ante, it does not pretend that the model is a representation of the true DGP).

Model Estimation
Finding the model parameters' values that are most likely to have generated the observed empirical data series (ex-ante, it assumes that the model is the true representation of the DGP).

Note: The literature is often blurred in addressing such difference. De facto, they often converge to overlapping exercises.

From input to output validation

Conceptualizing output validation (Marks, 2013)

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

From input to output validation

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

1. If there is no intersection between M and S ($Q = \emptyset$), then the model is **useless**

From input to output validation

Conceptualizing output validation (Marks, 2013)

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

1. If there is no intersection between M and S ($Q = \emptyset$), then the model is **useless**
2. If the intersection Q is not null, then the model is useful, to some degree. In general, the model will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. That is, the model is both **incomplete** and **inaccurate**.

From input to output validation

Conceptualizing output validation (Marks, 2013)

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

1. If there is no intersection between M and S ($Q = \emptyset$), then the model is **useless**
2. If the intersection Q is not null, then the model is useful, to some degree. In general, the model will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. That is, the model is both **incomplete** and **inaccurate**.
3. If M is a proper subset of S ($M \subset S$), then all the models behaviours are correct (match historical behaviours), but the model doesnt exhibit all behaviour that historically occurs. The model is **accurate but incomplete**.

From input to output validation

Conceptualizing output validation (Marks, 2013)

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

1. If there is no intersection between M and S ($Q = \emptyset$), then the model is **useless**
2. If the intersection Q is not null, then the model is useful, to some degree. In general, the model will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. That is, the model is both **incomplete** and **inaccurate**.
3. If M is a proper subset of S ($M \subset S$), then all the models behaviours are correct (match historical behaviours), but the model doesnt exhibit all behaviour that historically occurs. The model is **accurate but incomplete**.
4. If S is a proper subset of M ($S \subset M$), then all historical behaviour is exhibited, but the model will exhibit some behaviours that do not historically occur. The model is **complete but inaccurate**.

From input to output validation

Conceptualizing output validation (Marks, 2013)

- Let set S be the observed outputs of the real-world system
- Let set M be the exhibited outputs of the model in a comparable number of periods
- Let set Q be the intersection, if any, between the set M and the set S, $Q := M \cap S$. We can characterise the model output in five cases.

1. If there is no intersection between M and S ($Q = \emptyset$), then the model is **useless**
2. If the intersection Q is not null, then the model is useful, to some degree. In general, the model will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. That is, the model is both **incomplete** and **inaccurate**.
3. If M is a proper subset of S ($M \subset S$), then all the models behaviours are correct (match historical behaviours), but the model doesnt exhibit all behaviour that historically occurs. The model is **accurate but incomplete**.
4. If S is a proper subset of M ($S \subset M$), then all historical behaviour is exhibited, but the model will exhibit some behaviours that do not historically occur. The model is **complete but inaccurate**.
5. If the set M is equivalent to the set S ($M \equiv S$), then the model is complete and accurate.

Classes of models

Demonstration models
Existence proofs for phenomena of interest. They are also called *exploration models* because their primary objective to unveil the existence of some mechanisms; typically answer the question *what if* ? (Schelling, 1971, e.g.)

Descriptive models
Attempt to track real world phenomena. Their primary aims are that of qualitatively/quantitatively mimic the dynamics of some real event and to explain it by means of the economic forces embedded in the model. (Dosi et al., 2015, e.g.)

Both types of models require verification, but only the second class requires validation (and credible values).

Parameters' space exploration

Parameters' space exploration (*sensitivity analysis*)

Replies to the question: *how the model's outcome is affected by a marginal variation in the parameters?*

Represents a sort of "model elasticity" w.r.t. a change in the parameters (Salle and Yldzolu, 2014).

Trade-off:

- **robust model**: ideally you don't want that all the results change as soon as you change one parameter (unless that's the scope of the parameter within the model);

- **tautological model**: ideally you don't want a model whose outcome is equivalent for all the values of a parameter (in that case it is likely that the parameter does not play any role and can be dropped).[1]

---

[1]Assuming that a more parsimonious model is always preferred (KISS vs. KIDS).

The challenges

Ideally, we want a descriptive model to be

- **verified**: we have written the code well, the software actually represents what it is meant to
- **valid**:
  - parameters are correctly calibrated/estimated
  - initial conditions are correctly identified
  - external shocks are correcly sampled
  - processes (e.g. interactions) are correctly represented
  - output is accurate and complete
- **robust**: delivers a range of dynamics, whose dependence on inputs is correctly indentified

Model stationarity and model ergodicity

Model Stationarity
A model is said to be stationary if, after a certain period T_s, the main variable(s) under scrutiny converges to a time-invariant distribution;

Model Ergodicity
A model is said to be ergodic if, after a certain period T_s, the main variable(s) under scrutiny converges to a time-invariant distribution and such distribution is equivalent for all the Monte Carlo simulations.

Stylized facts

Stylized fact (SF)
An empirical/statistical regularity.

> *Facts as recorded by statisticians are always subject to numerous snags and qualifications, and for that reason are incapable of being summarized.*
>
> *Kaldor (1957)*

SF can be characterized by different degrees of:

- robustness;
- generality;
- easiness of generation.

Replication of SF does not mean the model is a descriptive one, but a descriptive model shall replicate (at least some) SF (Brock, 1999; Buchanan, 2012; Fagiolo et al., 2019; Guerini and Moneta, 2017).

Estimation of ACE models

Indirect inference

A simulation-based method for estimating the parameters of economic models whose likelihood function is intractable or unknown (Gourieroux et al., 1993).

General procedure

- simulate data from the economic model for different values of its parameters;
- use an *auxiliary model* to form a criterion function;
- determine the parametrization at which the model and the observed data look most similar.

The objective is that of choosing the parameter vector such that the observed data and the simulated data look most similar from the point of view of the chosen auxiliary model.

A formal description

The ACE model

Let:

$$y_t = \mathcal{G}(y_{t-1}, x_t, u_t; \theta), \quad t = 1, \ldots, T, \quad y_0 = \hat{y}_0, \quad u_t \sim iid \tag{1}$$

describe the DGP (i.e. our ACE model). $\mathcal{G}$ can be interpreted as a probability density function for y_t conditional on y_{t-1} and x_t, depending also on the k-dimensional parameter vector θ.

An econometrician would approach the problem by using the observed data to estimate the k-dimensional parameter vector $\hat{\theta}$.

Standard inference tools – e.g. Maximum Likelihood Estimation (MLE) – require that $\mathcal{G}$ has an analytical closed form such that a likelihood function can be derived and the optimal θ^* can be estimated.

Indirect inference helps when $\mathcal{G}$ is too complex and $\mathcal{G}', \mathcal{G}''$ cannot be computed. That's why it is particularly suited for ABM, where $\mathcal{G}$ typically represents a high-dimensional, non-linear model.

Indirect inference

The auxiliary model (on actual data)

Let:

$$\mathcal{F}(y_t|y_{t-1}, x_t; \beta) \tag{2}$$

be an auxiliary model (sufficiently simple) depending on the p-dimensional parameter vector β (with $p \geq k$) which can also be incorrectly specified.

The parameters $\hat{\beta}$ can be estimated using the actual data as:

$$\hat{\beta} = \underset{\beta}{argmax} \sum_{t=1}^{T} log \mathcal{F}(y_t|y_{t-1}, x_t; \beta) \tag{3}$$

Indirect inference

The auxiliary model (on simulated data)

The parameters β of the auxiliary model $\mathcal{F}$ can also be estimated over the simulated data.

Procedure:

- draw the disturbances $u_{m,t}$, with $m = 1, \ldots, M$ and $t = 1, \ldots, T$ from a defined probability distribution;

- pick a parameter vector θ and iterate the model $\mathcal{G}(\theta)$ to generate a random sequence of the variables of interest $\tilde{y}_{m,t}(\theta)$;

- maximize the average of the log-likelihood across the M simulations to obtain:

$$\tilde{\beta}(\theta) = \underset{\beta}{argmax} \sum_{m=1}^{M} \sum_{t=1}^{T} log \mathcal{F}(\tilde{y}_{m,t}(\theta) | \tilde{y}_{m,t-1}(\theta), x_t; \beta) \; ; \tag{4}$$

- select the values θ that minimizes the distance between $\hat{\beta}$ and $\tilde{\beta}(\theta)$.

Similarity measure

When using indirect inference, it is necessary to choose a metric for measuring the distance between the parameters of the auxiliary model estimated respectively with real and simulated data.

There are three possibilities corresponding to the classical trinity of testing:

$$\hat{\theta}^{WA} = \underset{\theta}{argmin} \left[\left(\hat{\beta} - \tilde{\beta}(\theta) \right)' W \left(\hat{\beta} - \tilde{\beta}(\theta) \right) \right]$$

$$\hat{\theta}^{LR} = \underset{\theta}{argmin} \left[\sum_{t=1}^{T} log \mathcal{F}(y_t | y_{t-1}, x_t; \hat{\beta}) - \sum_{t=1}^{T} log \mathcal{F}(y_t | y_{t-1}, x_t; \tilde{\beta}(\theta)) \right]$$

$$\hat{\theta}^{LM} = \underset{\theta}{argmin} \left[S(\theta)' V S(\theta) \right], \qquad S(\theta) = \sum_m \sum_t \frac{\partial}{\partial \beta} \left[log \mathcal{F} \left(\tilde{y}_t^m(\theta) | \tilde{y}_{t-1}^m, x_t; \hat{\beta} \right) \right]$$

where W and V are two positive-definite weighting matrices.

Pros and cons for ACE applications

Advantages

- relies on estimators with well known statistical properties;
- even if the auxiliary model is incorrect, the estimation can be correct;
- it is the natural extension of a classical econometric problem.

Drawbacks

- the estimation error increases with the dimensionality of the problem;
- requires the simulation of the true model over a large set of parameter values and with many Monte Carlo runs (might be computationally heavy);
- results on the optimal choice of θ might depend upon:
  - the chosen objective variables;
  - the chosen distance measure;
  - the number of Monte Carlo replications.

Particular cases

The **Simulated Minimum Distance** (SMD - Hsiao, 1989) and the **Method of Simulated Moments** (MSM - McFadden, 1989) can be seen as special cases of the indirect inference, where the auxiliary models upon which the estimation exercise is constructed is:

- one or few parameters estimated from a simple econometric model involving the variables of interest (e.g. the Pareto-tail parameter of the size distribution of firms);
- one or few low order moments concerning the variables of interest (e.g. average growth rate of GDP).

Estimation of small-scale ABM

Gilli and Winker (2001)

Gilli and Winker (2001) have been the pioneers in this field of research by applying the indirect inference method to ACE models.

Main features

- Simple ABM of the foreign exchange market (the Kirman, 1991, *ant model*);
- Dataset on the daily log-returns of the DM/US\$ exchange rate;
- Two key data properties to be matched:
  - the excess kurtosis at daily frequency k_d;
  - the volatility clustering parameter α_1.

The description of the model

Two types of agents:

- fundamentalist: $E^f[\Delta S_{t+1}] = \nu(\bar{S} - S_t)$;
- chartist: $E^c[\Delta S_{t+1}] = S_t - S_{t-1}$.

An agent might change rule due to:

- random mutation, with a fixed probability ε;
- conviction, with a fixed probability δ and after having interacted with a second individual.

The market price is determined as:

$$E^m[\Delta S_{t+1}] = w_t E^f[\Delta S_{t+1}] + (1 - w_t)E^c[\Delta S_{t+1}]. \tag{5}$$

Simulated Minimum Distance

The indirect inference procedure for the estimation of the two parameters (ε, δ) solves the following minimization problem:

$$f = |\bar{\bar{k}}_d - \hat{k}_d| + |\bar{\bar{\alpha}}_1 - \hat{\alpha}_1| \tag{6}$$

where $\bar{\bar{k}}$ and $\bar{\bar{\alpha}}_1$ are the two data properties, but estimated on the simulated data and averaged over $10K$ Monte Carlo replications; $\hat{k}_d, \hat{\alpha}_1$ are instead estimated on real data.

The algorithm

Algorithm 1 Indirect estimation procedure

1: Give $x^{(0)} \in \mathbf{R}^n$ starting values of parameters to be estimated
2: **while** not converged **do**
3:    Determine successive vectors x (defined by optimization algorithm)
4:    **for** each x **do**
5:       Initialize random variable generators with fixed seed
6:       **for** $i = 1 : nr\ ep$**do**
7:          Generate random sequences for price simulation
8:          Simulate price path $p^{(i)}$ and returns $r^{(i)}$
9:          Compute $\hat{\alpha}_{1_i}$ and k_{d_i}
10:      **end for**
11:      Truncate tails (10%) of the distribution of $\hat{\alpha}_{1_i}$ and k_{d_i}, $i = 1, \ldots, nr\ ep$
12:      Compute means $\overline{\hat{\alpha}}_1$ and $\overline{k}_d$ of truncated distributions
13:      Evaluate objective function
          $f = \mid \overline{k}_d - k_{d_{\mathrm{emp}}} \mid + \mid \overline{\hat{\alpha}}_1 - \alpha_{1_{\mathrm{emp}}} \mid$
14:   **end for**
15: **end while**

Figure 1: The Gilli and Winker (2001) algorithm for indirect inference of the two parameters (ε, δ).
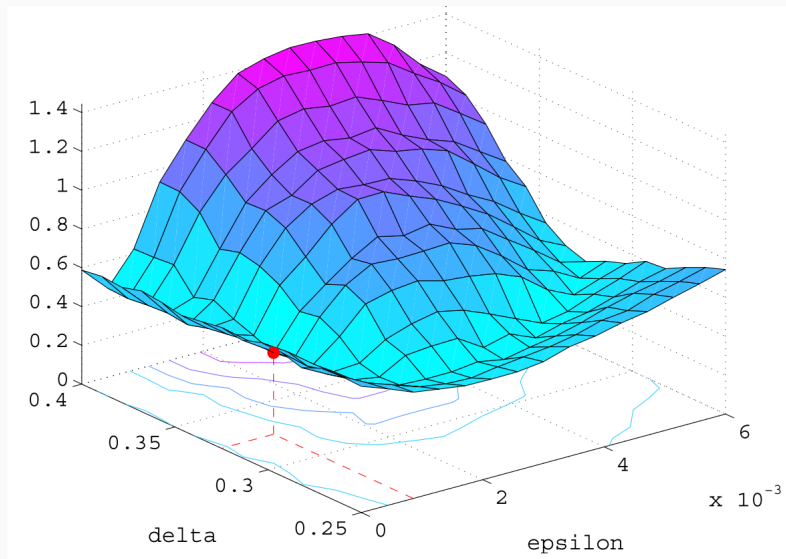
Surface plot

Figure 2: The Gilli and Winker (2001) surface plot of f over the (ε, δ) space.

What does this exercise tell us?

Optimal estimate is $(\varepsilon^*, \delta^*) = (0.00086, 0.325)$.

Economic implications

In the Kirman (1991) model the dynamic converges to a coexistence of fundamentalist and chartist agents if $\varepsilon < (1 - \delta)(N - 1)$.[2]

Hence, bringing the data to the model might suggest that the DGP of the empirical observation is better approximated by a model where the switching between the fundamentalist and chartist trading strategies matters.

The estimation seems to conclude that the relative fractions of fundamentalist and chartist traders aren't constant over time.

---

[2]Where $N = 100$ is the number of traders.

Limitations

Gilli and Winker (2001) recognize that:

- the two moments do not provide a sufficient description of the exchange rate dynamics (hence this estimation in hardly a test on the theory of exchange rate determination);
- further features (additional moments) shall be taken into account to better discriminate between different models;
- including other moments would however require more computational power (which was limited in 2001, but is less of a problem today for simple small-scale models).

In addition:

- the truncation of the 10% tails can be debated;
- the distance f is additive with same weights of the two objective. One can think of many possible functional forms or at different weights combinations;
- there are other parameters which have not been estimated. What about, for example, the effects of the parameter ν?

The heritage of Gilli and Winker (2001)

Gilli and Winker (2001) set the ground for the development of the literature in the estimation of ACE models.

The main algorithmic procedure adopted has been extensively used with some variations and improvements:

- Alfarano et al. (2005): simpler model with an analytical closed form solution for the distribution of returns and, therefore, direct and parametric (rather than indirect) estimation of the parameters;

- Franke and Westerhoff (2012): two models (with at most 8 parameters) estimated with respect to 9 selected moments by minimization of a weighted quadratic loss function; introduction of the *moment coverage ratio (MCR)* concept;

- Kukacka and Barunik (2017): estimation of the Brock and Hommes (1998) model by means of non-parametric simulated maximum likelihood estimation (NPSMLE).

Grazzini et al. (2017)

Introduction of Bayesian techniques to estimate ABM with applications to the Cliff and Bruten (1997) model as well as to a New Keynesian model (mid-scale).

Advantages

- does not require pre-selected moments to evaluate the model quality;
- incorporates prior knowledge and copes better with uncertainty;
- uses all the information coming from the data.

Drawbacks

- can mask identification issues by adding curvature to possibly flat likelihood functions;
- it is computationally demanding since the likelihood estimation might require as much time as the simulation of the model;
- when approximating the likelihood, the extent of the first advantage is reduced.

Bayes theorem

Bayes theorem applied to the real data, as possibly derived by the model as a DGP writes:

$$p(\theta|y) \propto \mathcal{L}(\theta; y) \cdot p(\theta) \tag{7}$$

where:

- $p(\theta|y)$ is the *posterior* distribution of the parameters θ;
- $\mathcal{L}(\theta; y) = p(y|\theta)$ is the likelihood of observing the empirical data y, conditional on the simulation parameter vector θ;
- $p(\theta)$ is the *prior* distribution of the parameters θ;

Major difference with SMD
The outcome of the estimation is a multivariate distribution (univariate over each of the components of the parameter vector θ) and not a single point estimate.

Estimating the likelihood

The central issue of the Bayesian approach is the estimation of the likelihood function $\mathcal{L}(\theta; y)$.

Non-parametric estimation (slower)

$$p(y|\theta) = \mathcal{L}(\theta; y) = \prod_{t=1}^{T} g(y_t|\theta)$$

and the estimate $\hat{g}$ is performed trough Kernel density estimation (KDE) using the simulated data after the statistical equilibrium has been reached.[3]

Parametric estimation (faster)

$$y_t = g^*(\theta) + \varepsilon_t$$

The functional form of g^* is here assumed to have some convenient form (e.g. Gaussian) but if the assumption is wrong, an estimation bias is likely to arise.

---

[3]Ergodicity and stationarity of the model are required.

Iteration over θ

Once the likelihood is estimated, the application of Bayes theorem in equation 7 allows one to obtain a density for the posterior at one given value of θ.

But to recover the whole posterior distribution one shall sample for different $\theta \in \Theta$ (as in the standard indirect inference approach) and simulate the models at many different values.

Thus, if θ has length K and each element θ_k is sampled at J different values, we need to estimate the likelihood a number of times equal to $N_{\mathcal{L}} = J^K$.

All in all, the model requires $N = N_{\mathcal{L}} \cdot M$ independent simulations and an equivalent number of calculations for the likelihood function.

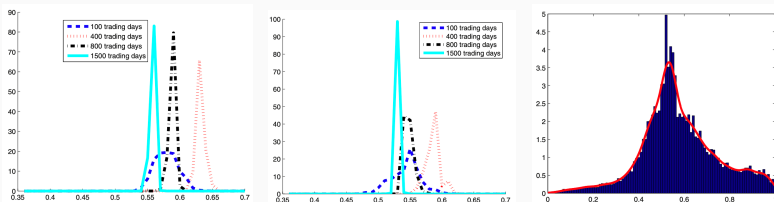# Estimation with parametric and non-parametric strategies

Figure 3: Posterior distribution of the parameter β under grid search and Gaussian parametric estimation, grid search with KDE and MCMC search with KDE.

Results

- Parametric Gaussian: highest precision;
- KDE with grid search: quite precise but biased;
- KDE with MCMC search: quite precise but highly uncertain.
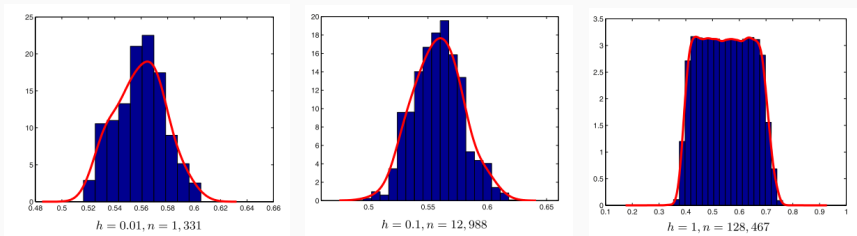
Estimation with the ABC strategy

Figure 4: Posterior distribution of the parameter β under Approximate Bayesian Computation algorithm, with KDE estimated at different bin-width values.

Result

- Approximate Bayesian Computation as good as the parametric one in the best case;
- but the results are sensitive to the KDE bin-width choice;
- in particular there is a trade-off between bin-width and estimation uncertainty.

Estimation of large-scale ABM

Estimation via method of simulated moments

To our knowledge, only three mid-scale models estimated via MSM:

- Bianchi et al. (2007): macroeconomic model;
- Barde (2020b): macroeconomic models comparison;
- Guerini et al. (2020): industry dynamics model.

The biggest issue remains the computational burden, which is accompanied by the *curse of dimensionality* (De Marchi, 2005).

Example: Calibrating the Dosi et al. (2015) model

- number of parameters: $K = 31$;
- number of Monte Carlo: $M = 100$;
- grid search of equispaced points for each parameter: $J = 20$;
- simulation time of a run: $T = 60s$.

Required independent simulations: $M \cdot J^K = 2.15 \cdot 10^{42}$ runs;
Required computational time: $T \cdot M \cdot J^K = 3.58 \cdot 10^{40}$ hours.

Meta-models

At this stage, is clear that the two most computationally intensive tasks for estimating an ABM are:

- running the model;
- computing the likelihood / evaluating the distance.

One approach to fasten the times of evaluation is that of replacing the model with a *surrogate* (a.k.a. meta-model) and evaluating its quality.

What is a meta-model?
Is a model/function which approximates the relation between the inputs of the complete ABM with its output. Allows one to evaluate the variation in the model outcomes over the parameter space without simulating the model.

Two possible approaches:

- parametric (see Salle and Yldzolu, 2014; Dosi et al., 2018)
- non-parametric (see Lamperti et al., 2018)

Procedure

To carry out the exercise and evaluate the performance of the model the following procedure is followed:

1. draw a relatively large set of points using any standard sampling routine (e.g. NHLH) to proxy the whole parameter space;

2. employ a small random subset of these points, simulate the ABM and classify them (positive or negative label) to learn a surrogate model (i.e. a classifier system);

3. use the surrogate model to predict the probability of all the other non-classified points to be positively labelled;

4. select a small random subset of the points classified by the surrogate (with high probabilities of being positive) and evaluate them against the ABM output;

5. repeat steps 2 to 4 until a satisfactory level of performance is achieved or until a predefined number of iterations (a.k.a. *budget*) is completed.

Crucial Steps

Definition of a positive calibration

A positive calibration is a parameter vector $\theta \in \Theta$ such that the model's outcome satisfies a predefined calibration criterion.

Selection of the learning algorithm

Usage of XGBoost (a gradient boosting algorithm) that produces a statistical model that classifies/labels the sampled points.

Surrogate performance

Binary case: $\underset{\mathcal{F}}{argmax}\ F_1 = \frac{2 \cdot \text{true positives}(\theta;\mathcal{F})}{2 \cdot \text{true positives}(\theta;\mathcal{F}) + \text{false positives}(\theta;\mathcal{F}) + \text{false negatives}(\theta;\mathcal{F})}$

Real-valued case: $\underset{\mathcal{F}}{argmin}\ MSE = \frac{\sum_{i=1}^{N}(\hat{y}_i(\theta;\mathcal{F}) - y_i(\theta))^2}{N}$

Out-of-sample evaluation

Use of the true positive rate $TPR = \frac{\text{number of correctly predicted positives}}{\text{number of positives in the pool}}$

Application to the Fagiolo and Dosi (2003) island model

Surrogate model selection criteria

- average GDP growth above 2%;

- output growth rates with AEP distribution with $b \leq 1$.

- no exercise with real data; the aim of the paper is that of evaluating the goodness of fit of the surrogate (not of the Fagiolo and Dosi (2003) *island model*).

Results

- surrogate modelling is 3750 times faster than the fully-fledge *island model*;

- with a sufficient *budget*, the surrogate model can out-of-sample correctly classify 90% of the points.

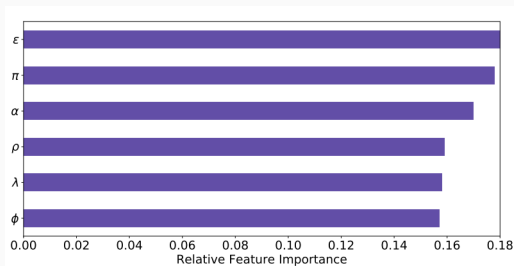# The meta-model as a descriptive device

Figure 5: Importance of each parameter in shaping behaviour of the Islands model according to the specified criteria.

Results

- the meta-model can be used to decompose the variance of the performace as a function of the parameters;
- here all the parameters are almost equally important;
- the parameter ε is slightly more relevant, meaning that small modifications of it can affect the calibration exercise to a larger extent.

Assessment of the strategy

Advantages

- fast due to the fact that one doesn't need to simulate the ABM many times;

- when the response surface to a variation in the parameters is very rugged, the surrogate is able to capture them;

- allows one also to assess the parameters importance and to avoid modifying parameters that do not affect the model's outcome.

Drawbacks

- cases in which the best meta-model is an insufficient proxy of the ABM;

- when the response surface to a variation in the parameters is smooth, parametric alternatives can be more efficient;

- the surrogate is difficult to interpret: for policy analysis one needs to resort to simulation of the ABM.

New distances and objective functions

## New metrics for validation of ACE model

In the recent years a series of new distances to evaluate the performance of an ACE model against real data have been developed.

Among them:

- Marks (2013);
- Barde (2017, 2020a)
- Lamperti (2018b)
- Guerini and Moneta (2017)

Current research is in the attempt of using these distances as minimization criteria for medium- and large-scale models in order to estimate them over a large number of parameters.

Are real and simulated data similar?
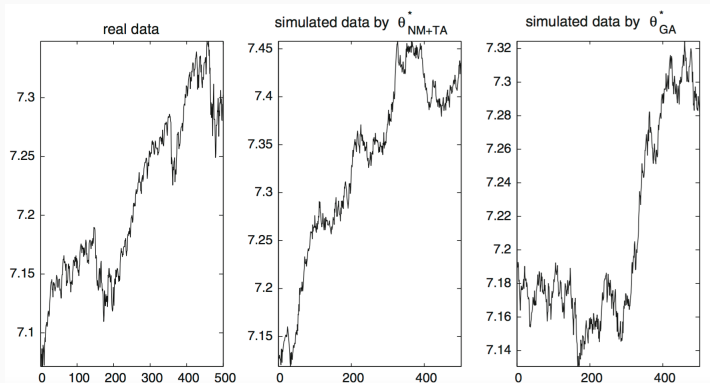
Source: Fabretti (2013)

The KL divergence

- Estimate the conditional distribution of patterns emerging from real and simulated data
- Compare the two distributions

A well known measure of the distance between distributions, with well known theoretical properties, is the Kullback-Leibler divergence:

$$D_{KL}(\mathbf{p}||\mathbf{q}) = \sum_{s \in S} p(s) \log \left(\frac{p(s)}{q(s)} \right), \tag{8}$$

The KL divergence measures the information loss when assuming that a distribution is q while the true one is p.

Alternative distance measures

A **set of criteria** has been developed to capture the distance between real and simulated data:

- the State Similarity Measure (SSM – Marks, 2013);
- the Markov Information Criterion (MIC – Barde, 2017, 2020a);
- the Generalized Subtracted L-divergence (GSL-div – Lamperti, 2018b,a).

Similarities and trade-offs:

- they are all built upon the KL divergence;
- they capture slightly different features;
- they have different computational burdens ($\text{time}_{\text{MIC}} > \text{time}_{\text{GSL}} > \text{time}_{\text{SSM}}$)

The GSL-div by Lamperti (2018b)

It measures the **distance between the regularity of behaviours** observed in the empirical data and that of simulated data.

It is constructed through a **simple algorithm**

1. Time series are symbolized into b symbols
2. Sequences of symbols (words) are observed using multiple rolling window of maximal length L
3. Distributions are compared using a well known information theoretic divergence measure (L-div)
4. They are iteratively aggregated using weights that penalize short windows

Parameters b and L need to be fine-tuned.

Symbolization
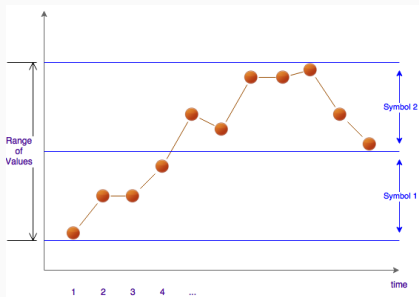
(a) b (precision of symbolization) $= 2$
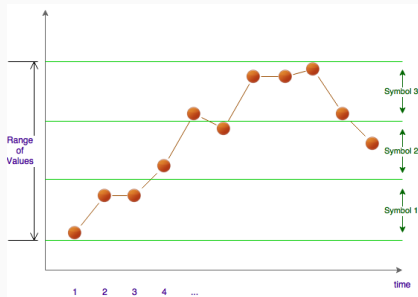
(b) b (precision of symbolization) $= 3$

Matching Causality by Guerini and Moneta (2017)

Is the mere replication of stylized facts or distribution a hard test for evaluating the model quality? If the answer is no, then one shall focus on **causal relationships**.
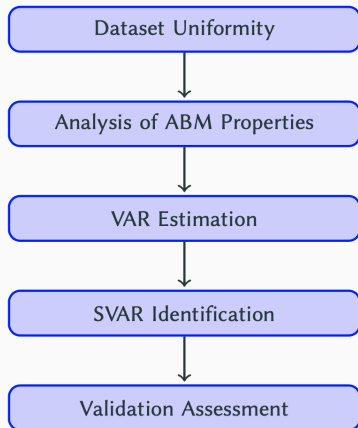
The intuition is that models with different causal structures, might generate equivalent statistical properties (i.e. the same stylized facts).

How to discriminate between them?

- estimate the causal structures emerging from real-word data and model-generated data;
- compare them in a meaningful way.

A new paper building upon this has recently been proposed by Martinoli et al. (2022).

Dataset Uniformity

↓

Analysis of ABM Properties

↓

VAR Estimation

↓

SVAR Identification

↓

Validation Assessment

Ergodicity and stationarity can be tested by batteries of Kolmogorov-Smirnov tests.

The SVAR identification can be done by modern causal search algorithms with relatively mild assumptions, without imposing **a-priori restriction**.

The **similarity** is computed using signs and magnitudes of the causal relationships estimated on the real and on the simulated data.

Application has been done at the moment on the Dosi et al. (2015) model.

Currently on-going: a comparison of Grazzini et al. (2017) and Guerini and Moneta (2017) on the Delli Gatti and Desiderio (2015) model.

Conclusions

Overall Evaluation

20 years after the Gilli and Winker (2001) seminal contribution:

- computational power has hugely increased expanding the possibilities for estimation and calibration of ACE models;
- many alternative nuances of the estimation method are now available, but wide applications by ACE modellers is still limited.

The quest for the optimal estimation strategy?
Probably not, as all methods have advantages and drawbacks.

Future research will likely aim at:

- specifying the statistical properties of the different estimators as well as the conditions for their usage to better define the applicability domains of the different methods;
- writing a broad open-source software embedding most of the methodologies to allow ABM economists and policy-makers to easily implement them.

A Validation Cookbook for AB Modellers

Ideal procedure for *descriptive models*:

1. **input validation**: build the model by grounding behavioural rules and interaction mechanisms on the empirical or experimental literature
2. **estimation**: decide the real-world data, the moments to match and estimate the model with the preferred method
3. **estimation robustness**: run Monte Carlo simulations and check if the model dynamics resembles the real world one according to some distance measure
4. **stylized facts replication**: check which stylized facts the model can replicate (in addition to the ones used for the estimation)
5. **causal structure validation**: check if the model causal structure resembles the one that emerges from the real data (if the model has policy implications, this is particularly relevant)
6. **sensitivity analysis**: run local sensitivity analysis or use a surrogate to globally explore the parameter space
7. **policy analysis**: run policy exercises and compare results to the baseline parametrization

Thanks for the attention.

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. ISSN 09277099. doi: 10.1007/s10614-005-6415-1.

Sylvain Barde. A practical, accurate, information criterion for nth order markov processes. *Computational Economics*, 50(2):281–324, Aug 2017. doi: 10.1007/s10614-016-9617-9. URL https://doi.org/10.1007/s10614-016-9617-9.

Sylvain Barde. Macroeconomic simulation comparison with a multivariate extension of the markov information criterion. *Journal of Economic Dynamics and Control*, 111:103795, 2020a. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2019.103795. URL http://www.sciencedirect.com/science/article/pii/S0165188919301927.

Sylvain Barde. Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion. *Journal of Economic Dynamics and Control*, 111, 2020b. ISSN 0165-1889. doi: 10.1016/j.jedc.2019.103795. URL https://www.sciencedirect.com/science/article/pii/S0165188919301927.

Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro A. Vagliasindi. Validating and calibrating agent-based models: A case study. *Computational Economics*, 30(3):245–264, 2007. ISSN 09277099. doi: 10.1007/s10614-007-9097-z.

W. Brock. Scaling in economics: a reader's guide. *Industrial and Corporate Change*, 8(3): 409–446, 1999. ISSN 0960-6491. doi: 10.1093/icc/8.3.409.

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, 1998.

Mark Buchanan. It's a (stylized) fact! *Nature Physics*, 8(1):3, 2012. ISSN 17452473. doi: 10.1038/nphys2191.

Dave Cliff and Janet Bruten. Minimal-intelligence agents for bargaining behaviors in market-based environments. Technical report, 1997.

Herbert Dawid and Domenico Delli Gatti. *Agent-Based Macroeconomics*, volume 4. Elsevier B.V., 2018. doi: 10.1016/bs.hescom.2018.02.006. URL https://doi.org/10.1016/bs.hescom.2018.02.006.

Herbert Dawid and Philipp Harting. Capturing firm behavior in agent-based models of industry evolution and macroeconomic dynamics. *Applied Evolutionary Economics, Behavior and Organizations, ed. by G. Bünsdorf, Edward-Elgar*, pages 103–130, 2012.

Scott De Marchi. *Computational and mathematical modeling in the social sciences*. Cambridge University Press, 2005.

Domenico Delli Gatti and Saul Desiderio. Monetary policy experiments in an agent-based model with financial frictions. *Journal of Economic Interaction and Coordination*, 10(2):265–286, 2015.

G. Dosi, M. C. Pereira, and M. E. Virgillito. On the robustness of the fat-tailed distribution of firm growth rates: a global sensitivity analysis. *Journal of Economic Interaction and Coordination*, 13(1):173–193, 2018. ISSN 18607128. doi: 10.1007/s11403-017-0193-4.

Giovanni Dosi. A very reasonable objective still beyond our reach: economics as an empirically disciplined social science. In *Models of a Man: Essays in Memory of Herbert Simon*. 2004.

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, 2015. ISSN 01651889. doi: 10.1016/j.jedc.2014.11.014. URL http://dx.doi.org/10.1016/j.jedc.2014.11.014.

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.

Giorgio Fagiolo and Giovanni Dosi. Exploitation, exploration and innovation in a model of endogenous growth with locally interacting agents. *Structural Change and Economic Dynamics*, 14(3):237–273, 2003.

References IV

Giorgio Fagiolo and Andrea Roventini. Macroeconomic policy in DSGE and agent-based models redux: New developments and challenges ahead. *Jasss*, 20(1), 2017. ISSN 14607425. doi: 10.18564/jasss.3280.

Giorgio Fagiolo, Mattia Guerini, Francesco Lamperti, Alessio Moneta, and Andrea Roventini. Validation of Agent-Based Models inEconomics and Finance. pages 763–787. 2019. doi: 10.1007/978-3-319-70766-2_31. URL http://link.springer.com/10.1007/978-3-319-70766-2{_}31.

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. ISSN 01651889. doi: 10.1016/j.jedc.2011.10.004. URL http://dx.doi.org/10.1016/j.jedc.2011.10.004.

Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1):107–143, 2009. doi: 10.1111/j.1756-8765.2008.01006.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2008.01006.x.

Manfred Gilli and Peter Winker. Indirect Estimation of the Parameters of Agent Based Models of Financial Markets. *SSRN Electronic Journal*, (November), 2001. doi: 10.2139/ssrn.300220.

References V

C Gourieroux, A Monfort, and E Renault. Indirect Inference C. Gourieroux; A. Monfort; E. Renault Journal of Applied Econometrics , Vol. 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques. (Dec., 1993), pp. S85-S118. *Journal of Applied Econometrics*, 8, 1993.

Jakob Grazzini, Matteo G. Richiardi, and Mike Tsionas. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77:26–47, 2017. ISSN 01651889. doi: 10.1016/j.jedc.2017.01.014.

M. Guerini and A. Moneta. A method for agent-based models validation. *Journal of Economic Dynamics and Control*, 82, 2017. ISSN 01651889. doi: 10.1016/j.jedc.2017.06.001.

Mattia Guerini, Philipp Harting, and Mauro Napoletano. Governance structure, technical change and industry competition. Technical report, 2020.

Cars Hommes. *Behavioral rationality and heterogeneous expectations in complex economic systems*. Cambridge University Press, 2013.

Cheng Hsiao. Consistent estimation for some nonlinear errors-in-variables models. *Journal of Econometrics*, 41(1):159–185, 1989. ISSN 03044076. doi: 10.1016/0304-4076(89)90047-X.

Nicholas Kaldor. A Model of Economic Growth. *The Economic Journal*, 67(268):591–624, 12 1957. ISSN 0013-0133. doi: 10.2307/2227704. URL https://doi.org/10.2307/2227704.

References VI

Alan Kirman. The intrinsic limits of modern economic theory: the emperor has no clothes. *The Economic Journal*, 99(395):126–139, 1989.

Alan Kirman. Epidemics of opinion and speculative bubbles in financial matkets. *Money and financial markets chap*, 17, 1991.

Alan P Kirman. Whom or what does the representative individual represent? *Journal of economic perspectives*, 6(2):117–136, 1992.

Jiri Kukacka and Jozef Barunik. Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control*, 85:21–45, 2017.

Francesco Lamperti. Empirical validation of simulated models through the gsl-div: an illustrative application. *Journal of Economic Interaction and Coordination*, 13(1):143–171, 2018a.

Francesco Lamperti. An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics*, 5:83–106, 2018b. ISSN 24523062. doi: 10.1016/j.ecosta.2017.01.006. URL https://doi.org/10.1016/j.ecosta.2017.01.006.

Francesco Lamperti, Andrea Roventini, and Amir Sani. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389, 2018. ISSN 01651889. doi: 10.1016/j.jedc.2018.03.011. URL https://doi.org/10.1016/j.jedc.2018.03.011.

References VII

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

Mario Martinoli, Alessio Moneta, and Gianluca Pallante. Calibration and Validation of Macroeconomic Simulation Models: A General Protocol by Causal Search. *LEM Workin Paper Series*, 33, 2022.

Daniel McFadden. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5):995, 1989. ISSN 00129682. doi: 10.2307/1913621.

Luzius Meisser. The code is the model. *International Journal of Microsimulation*, 10(3):184–201, 2017.

Alessio Moneta. Causality in macroeconometrics: some considerations about reductionism and realism. *Journal of Economic Methodology*, 12(3):433–453, 2005. doi: 10.1080/13501780500223742. URL https://doi.org/10.1080/13501780500223742.

Thomas H. Naylor and J. M. Finger. Verification of Computer Simulation Models. *Management Science*, 14(2):B–92–B–101, 1967. ISSN 0025-1909. doi: 10.1287/mnsc.14.2.b92.

Isabelle Salle and Murat Yldzolu. Efficient sampling and meta-modeling for computational economic models. *Computational Economics*, 44(4):507–536, 2014. ISSN 15729974. doi: 10.1007/s10614-013-9406-7.

Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2): 143–186, 1971.

Levent Yilmaz. Validation and verification of social processes within agent-based computational organization models. *Computational & Mathematical Organization Theory*, 12(4):283–312, 2006. doi: 10.1007/s10588-006-8873-y. URL https://doi.org/10.1007/s10588-006-8873-y.