A meso-level empirical validation approach for agent-based computational economic models drawing on micro-data: a use case with a mobility mode-choice model

Alperen Bektas[1] · **Valentino Piana**[1] · **René Schumann**[1]

Abstract

The complex nature of agent-based modeling may reveal more descriptive accuracy than analytical tractability. That leads to an additional layer of methodological issues regarding empirical validation, which is an ongoing challenge. This paper offers a replicable method to empirically validate agent-based models, a specific indicator of "goodness-of-validation" and its statistical distribution, leading to a statistical test in some way comparable to the p value. The method involves an unsupervised machine learning algorithm hinging on cluster analysis. It clusters the ex-post behavior of real and artificial individuals to create meso-level behavioral patterns. By comparing the balanced composition of real and artificial agents among clusters, it produces a validation score in [0, 1] which can be judged thanks to its statistical distribution. In synthesis, it is argued that an agent-based model can be initialized at the micro-level, calibrated at the macro-level, and validated at the meso-level with the same data set. As a case study, we build and use a mobility mode-choice model by configuring an agent-based simulation platform called BedDeM. We cluster the choice behavior of real and artificial individuals with the same ex-ante given characteristics. We analyze these clusters' similarity to understand whether the model-generated data contain observationally equivalent behavioral patterns as the real data. The model is validated with a specific score of 0.27, which is better than about 95% of all possible scores that the indicator can produce. By drawing lessons from this example, we provide advice for researchers to validate their models if they have access to micro-data.

✉ Alperen Bektas
    alperen.bektas@unine.ch

Extended author information available on the last page of the article

Published online: 12 May 2021

Introduction

Modeling economies as complex systems has been attracting many scholars (Hamill and Gilbert 2016). Agent-based (AB) models are one of the modeling tools for complex systems, which can provide a realistic way to model economies; thus, their usage has been growing in the field of economics (as well as in other disciplines) during the last 3 decades (Fagiolo et al. 2019; Hamill and Gilbert 2016). AB models consist of autonomous and decentralized entities (agents); each can have dynamic behavior and heterogeneous characteristics (Geanakoplos et al. 2012). The dynamic behavior of heterogeneous agents is governed by decision-making mechanisms (rules) derived from established empirical and theoretical foundations (Dawid et al. 2014). Thus, agents do not necessarily make decisions based on the assumption of a representative agent who is intertemporally optimizing an objective function under rational expectations (Colander et al. 2008). The uses of these models in economics are collected under a common umbrella that we refer to as agent-based computational economics (ACE) (Tesfatsion 2002).

AB models have certain features that distinguish them from neoclassical ones (Arthur 1994). Economists often point to such features as a reason to use them (Hamill and Gilbert 2016). First of all, AB models have a bottom–up perspective. The macro-dynamics in these models are the emergent properties of micro-level interactions and agents' behavior, which is not constrained with equilibrium and hyper-rationality (Heckbert et al. 2010). These emergent properties at the macro-level can be used to analyze complex and decentralized systems quantitatively (Duffy 2006). As Arthur (2006) states, emerging properties often feedback micro-level decisions, which leads to a perpetual novelty in the behavior. Thanks to the bottom–up perspective, AB models are capable of modeling each individual's micro-behavior separately, which allows us to have a high level of heterogeneity (Dawid et al. 2012). Secondly, AB models can contain non-trivial interactions, which were governed by ex-ante defined rules of behavior. These interactions are often non-linear, which makes tracing of the emergent macro-patterns harder (Windrum et al. 2007). The interactions can lead to having information and adaptation, which make AB models realistic, as individual decisions (in real world) are largely based on incomplete information and preferences, which indicates that decision-making can evolve in case new information comes (Farmer and Foley 2009).

It is an asset for AB models (like other economic models) demonstrating how well the model Data Generating Process (mDGP) represents the real-world Data Generation Process (rwDGP) (Fagiolo et al. 2007; Klügl 2008; Bianchi et al. 2007; Murray-Smith 2015; Beisbart and Saam 2019). One way to do that is to compare the data generated by the mDGP and the rwDGP statistically; we call this procedure empirical validation (Windrum et al. 2007). AB models favor more descriptive accuracy than analytical tractability, contrary to neoclassical ones due to the potential existence (by no means necessary) of non-linearities, macro-micro feedback, heterogeneous interactions (Fagiolo et al. 2007). That makes the relationship

and the comparison of AB model-generated data and real data problematic, which leads to complexity and consequently, methodical problems regarding the empirical validation of AB models (Heckbert et al. 2010). Although there are contributions in the last decade, such as Barde (2020), Lamperti (2018a), and Guerini and Moneta (2017), we still do not have standardized empirical validation methods for AB models that inevitably lead to a lack of robustness in terms of validation (Fagiolo et al. 2019). That was recognized by AB modelers themselves and shown as one of the reasons for the reluctance of neoclassical economists to move AB camp, even though they recognized the significance of AB critique (e.g., heterogeneity, learning, interactions, etc.) and try to update their models accordingly (Windrum et al. 2007).

Previous research recommends the involvement of machine learning techniques as for empirical validation methods (Fagiolo et al. 2019; Barde and Van Der Hoog 2017), which allows us to perform more thorough comparisons of mDGP generated data and rwDGP generated data. The present paper has been motivated by this research and proposes an unsupervised machine-learning algorithm,[1] specifically *cluster analysis* (Russell and Norvig 2002), as an empirical validation method. The method focuses on the AB models that use micro-data as input and produce results accordingly to address questions from the real world. It aims to compare model-generated data and real data at the meso-level. To do this, it suggests clustering the ex-post behavior of real individuals and artificial agents, who have the same ex-ante given characteristics. Then, it quantitatively assesses how well the clusters are overlapping in a multidimensional latent space. Thus, the behavioral patterns in model-generated data and real data are compared. The method is discussed in the next section in detail. To apply the method as a case study, we build an AB model through configuring an AB simulation platform called Behavior Driven Demand Model (BedDeM) (Nguyen and Schumann 2019). The model and its features are explained in "Case study" thoroughly.

The rest of the paper is organized as follows. In "Methods", which consists of three subsections, we first discuss the theoretical background of the validation of AB models in light of existing literature. Then, we touch on the recently introduced validation approaches. After that, we explain the proposed method and discuss how it could expand the existing literature. In "Case study", we build an AB model to apply the method as a case study. "Results" shows and interprets the validation results of the case study. "Discussion" discusses the value of the method and its applicability to other AB models. It gives practical advice for the researchers who want to apply this validation method to their AB models. It also discusses what kind of AB models could be assessed by the method and provide some example models for the sake of clarity. Finally, the paper ends with the future works and conclusions sections.

---

[1] Unsupervised machine learning detects previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. Cluster analysis is considered as one of the most common unsupervised machine-learning algorithms (Kassambara 2017).
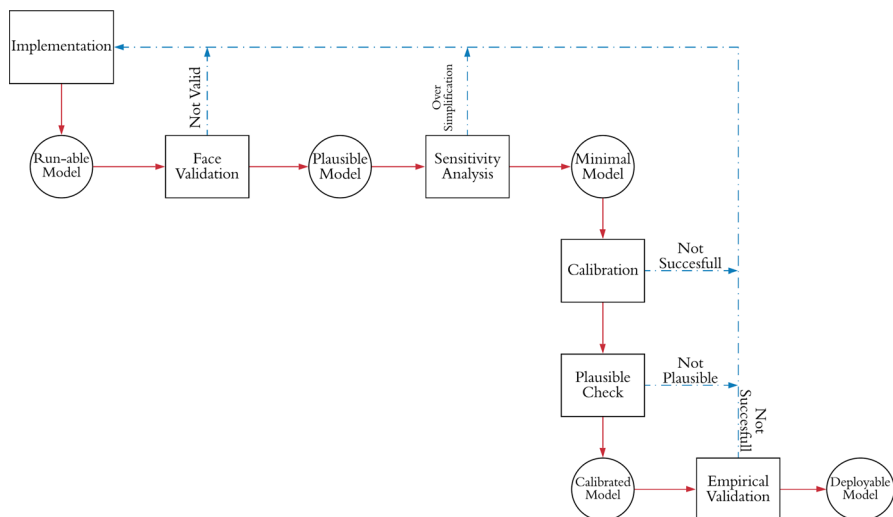
Fig. 1 A general procedure to validate AB models (Klügl 2008)

Methods

Theoretical background of validation of AB models

In this section, we follow a general-to-specific way to discuss the validation of AB models in the light of existing literature. First, we introduce the types of validation techniques (stages) for AB models in general terms. We utilize a procedure to validate AB models, which was introduced by Klügl (2008), and discuss the validation stages that are ordered in that procedure. Then, we discuss one of these stages (the last one) called empirical validation in detail, since we introduce a novel method for that stage in this paper.

One of the major valuable aspects of using AB models is to explain and understand a real-world phenomenon that is costly and sometimes difficult to analyze in real world (e.g., field experiments, real laboratory experiments, etc.) (Xiang et al. 2005). As Farmer and Foley (2009, p. 686) state, "AB models allow for the creation of a kind of artificial (virtual) universe in which many players act in complex and realistic ways". Thus, such models enable to analyze—in silico—the future status of the original system under novel conditions. Assessing how well the artificial universe (i.e., AB models) represents a proportion of the original system (i.e., a part of the real world that is aimed to be modeled) is an asset for models that potentially makes the modeling results more credible (Klügl 2008). This assessment is called validation in the literature (Windrum et al. 2007; Bianchi et al. 2007). If the model is validated, the answers derived from the model can be utilized to answer questions directed to the original system (Klügl 2008).

Klügl (2008) introduced a framework (see Fig. 1) that places different validation stages in an order to validate AB models. Some stages in the framework are

also discussed in Balci (1994) separately (i.e., without being a part of a framework). The framework starts with face validity. In that stage, the modelers are supposed to contact to domain experts to assess whether the model behaves reasonably. The experts provide subjective judgments on the accuracy of the model. Sensitivity analysis comes next, where the impact of different parameters on the model output is assessed. It is assumed that the relationship between a parameter and the output occurring in the model should occur similarly in the original system as well. Once such impacts are analyzed, then the appropriate values are assigned in calibration for the parameters. Calibration aims for finding the "optimal" parameter set, which resembles the output of the model to the output from the original system. In general, AB model parameters are calibrated to aggregated (macro) patterns (Guerini and Moneta 2017). The plausibility check comes after calibration, where human experts assess the plausibility of the model outcome (e.g., dynamics and trends of the different output values of model runs). It is technically the same as the previously discussed face validity, as Klügl (2008) states. Finally, statistical tests are applied to compare model-generated data and real data as named empirical validation.

Empirical validation is the last stage of the procedure in Fig. 1 and aims to compare the data coming from the rwDGP and the mDGP statistically. Assume that we have real data generated by the rwDGP, which contains different data points in a time-series. The data points can be at the micro-level as the expression in (1) denotes (Pyka and Fagiolo 2007; Windrum et al. 2007), where I represents the population of individuals whose heterogeneous behaviors are observed and contained in the vector of z in a finite time-series of n. For instance, for a mobility mode-choice model, z would be individual level mobility mode choice behavior:

$$(z)_i = \{z_{i,t}, t = t_0, \ldots, t_n\} \quad i \, \epsilon \, I, n \, \epsilon \, \mathbb{N} \tag{1}$$

$$(Z) = \{Z_t, t = t_0, \ldots, t_n\} \quad n \, \epsilon \, \mathbb{N}. \tag{2}$$

The data points that the rwDGP generates at the micro-level can be aggregated to obtain macro-data points, as denoted in (2) (Pyka and Fagiolo 2007; Windrum et al. 2007), where the vector of Z contains macro-data points of a population (i.e., I) over a time series. For instance, a household's consumption behavior is represented by a micro-level data point, while the aggregation of all households in a population I is represented by a macro-level data point, which can then be used as a component of the GDP. Modelers aim to approximate values for the vector of z or Z for which finding the optimal micro (θ, e.g., agent preferences) and macro (Θ, e.g., the environment) parameters is needed for calibration. Once the optimal parameters are set, which is the one step before the empirical validation in Fig. 1, then the output of the model can be compared empirically to real data from the original system (Fagiolo et al. 2007; Guerini and Moneta 2017). As Klügl (2008, p. 6) states, "calibration and validation must use different data sets for ensuring that the model is not merely tuned to reproduce given data, but may also be valid for inputs that it was not given to before". However, having two data sets from the same original system is not often possible. In such cases, the available data can be used on all available levels, as Klügl (2008) asserts. For instance, a model can use micro-data as input,

be calibrated at the macro-level, and be validated at the meso-level. Therefore, the same data set can be exploited at different levels without over-fitting.

Related works

In this section, we first discuss recently introduced validation methods. Then, we explain why our method is related to the discussed methods and how it could expand them for the sake of readers.

Lamperti (2018b) has offered an information theoretic criterion called General Subtracted L-divergence (*GSL-div*) as a validation method for AB models. The method measures the similarity between model-generated and real-world time-series. It assesses the extend of models' capability to mimic patterns (e.g., distribution of time-series such as changes of values from one point in time to another) occurring in real-world time-series. It is related to our method, because our method aims to compare the similarity among patterns occurring in real data and model-generated data as well. However, GSL-div focuses only on aggregated time-series data as Fagiolo et al. (2019) indicate, while our method focuses rather on meso-level behavioral patterns that are constructed by micro attributes. We discuss the advantages of the meso-level approach later. The authors state that the GSL-div can overcome certain shortcomings of the method of simulated moments (MSM), e.g., it does not need to resort to any likelihood function and provides a better representation about the behavior of complex time-series. Their method could be applied technically to any AB model that produces time-series data. Detailed explanation of the method, illustrative examples, and case studies can be found in Lamperti (2018a, 2018b).

Barde (2020, 2016) has introduced another information theoretic criterion as a validation method for AB models. The method is called Markovian information criterion (MIC). It follows the minimum description length (MDL) principle, which hinges on the efficiency of data compression to measure the accuracy of models' output (Grünwald and Grunwald 2007). It first uses model-generated data to create a Markov transition matrix for the model, and then uses the real data to produce a log score for the model on the data. The method uses the Kullback–Leibler (KL) divergence to measure the distance between real and model-generated data; thus, the accuracy of the mDGP is assessed. As the author states, the method does not include estimation; instead, it is applied to already calibrated models to assess their output. It is related to our method from that aspect. However, similar to GSL-div, the application level of our method is different than MIC as we explain in detail in the following section.

Grazzini and Richiardi (2015) discuss estimation methods for dynamic stochastic general equilibrium modeling (DSGE) models and analyze whether such models can also be applied to AB models. The authors mention the simulated minimum distance (SMD) methods, such as the method of simulated moments (MSM), as a natural approaches to the estimation of AB models. Such methods aim for estimating model parameters by minimizing the distance between the aggregates between model output and real data. Our approach differs from these methods, because it

focuses on the last step of the procedure of Klügl (2008) (see Fig. 1). In other words, it is applied to already calibrated models, similarly to the method of Barde (2016). Thus, the estimation methods in the class of SMD can be only complementary to our method. As we discussed in the future works section, in a future paper, we plan to couple an SMD method with our method to apply together on an AB model.

Differently from the previously discussed methods, Guerini and Moneta (2017) offer a method that aims to compare causal relationships in model-generated data and real-world data to validate AB models. The method hinges on estimating Structural Vector Autoregressive (SVAR) models through real and artificial time-series and comparing them to get a validation score. Our method does not rely on time-series and we compare relationships at meso-level, while the method of Guerini and Moneta (2017) focuses only on aggregate time-series.

To conclude, as Fagiolo et al. (2019, p. 14) state in their critical review, "all these recently developed validation methods focus only on aggregate time-series, while most of AB models have been able to replicate both micro and macro stylized facts". Some of the discussed methods could be applied in principle at the micro-level, but there is no "proof-of-concept" yet. Besides, applications of such methods at the micro-level could lead to over-fitting if a model gets micro-data as input and its parameters are estimated to fit individual behavior one-to-one (e.g., fitting behavior of artificial agent to its real counterpart). Considering the increasing availability of micro-data, the number of AB models using micro-data as input increases (Macal and North 2014; Hamill and Gilbert 2016). Therefore, in this paper, we offer a meso-level validation method for the models drawing on micro-data. The method involves an unsupervised machine-learning algorithm along the lines suggested by Fagiolo et al. (2019) and Barde (2016). They represent contributions regarding machine-learning involvement on the side of estimation (van der Hoog 2019). However, such involvement is still lacking on the side of validation. Our method could expand the existing validation methods towards the direction of machine-learning and encourage future contributions. The further text is structured as follows: we discuss the overall concept of our method in the next section in detail. We also discuss for what kind of AB models the method could be applied and provide some example models from recent research in "An overview of AB models that might be validated with our method".

The overall concept of the meso-level validation method

This section introduces a meso-level empirical validation method for AB models drawing on micro-data first as a broad methodological choice, and then, we describe it in detail. In broad terms, we sharply distinguish the different phases and goals of the relationship between real (empirical) data and the model in the following way: the meso-level is exclusively used for validation, whereas the micro-level is used for input micro-data into the agents in terms of parameters (not of outcomes of their decision-making process, because this could lead to over-fitting) and the macro-level for calibration. By this distinction, we radically eliminate any source of overlap between what is given to the model as input, what is used for calibrating its overall
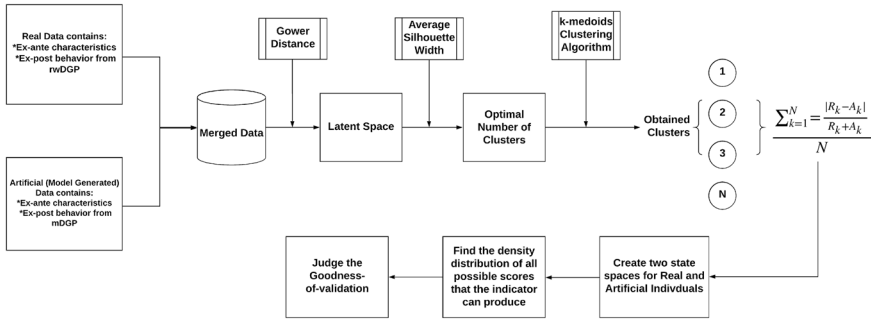
Fig. 2 The overall concept of the proposed empirical validation method

results and macro–micro-parameters, and what is used for validation. More specifically, our method consists of sequential steps for which we created an overall concept as in Fig. 2. We explain each step one after another, according to their sequence in the concept. The main goal of the concept is to compare model-generated data and real data at the meso-level to understand how well the mDGP can produce the behavioral patterns that occur in real data. It produces a quantitative score in a spectrum according to which we can assess the validity.

The overall concept gets two data sets as input. The first data set contains information regarding the ex-ante characteristics of artificial individuals (i.e., agents) and their ex-post behavior, generated by the mDGP. The second data set involves information regarding the ex-ante characteristics of real individuals and their ex-post behavior, generated by the rwDGP. Both data sets contain information at the individual level, since the mDGP of AB models produce data at the individual level (i.e., micro-level). Individuals are clustered according to their characteristics and behavior in the data sets, and these clusters are compared at meso-level quantitatively. An essential point for the comparison according to the method is that the real data should be the one that is used to initialize the model. In this case, individuals in real data are mapped to artificial agents one-to-one; thus, the number of real and artificial individuals becomes equal, which is a prerequisite to apply the validation method. The data sets can differ in what model-wise is an ex-post behavior, because an artificial agent might behave differently than a real individual with the same characteristics. The variables constituting the ex-ante characteristics should ideally be the ones influencing the ex-post behavior. By having this, the clusters involve a combination of the variables in individuals' characteristics and consequent behavior. Hence, by comparing clusters, we can study the behavioral patterns (e.g., the relationship between the characteristics and the behavior) in model-generated data and real data.

Instead of clustering artificial and real data sets separately, we merge them as indicated in Fig. 2, and cluster them together to analyze the balance in the clusters (i.e., how many real and how many artificial individuals are in each cluster). Individuals in the merged data are placed in a multidimensional latent space based on their attributes (i.e., ex-ante characteristics and ex-post behavior). The latent space

is represented by a symmetric distance matrix.[2] Several metrics exist to create that matrix, such as Euclidean, Manhattan, Gower, etc. (Bektas and Schumann 2019a). In the overall concept, we utilize the Gower distance metric, since it can handle different column types[3] (e.g., categorical, numerical, ordinal, etc.) to place instances in the latent space (Gower 1971). For instance, the merged data might contain some attributes of households that can be categorical such as income level, or numerical such as age. Gower distance can determine the positions of the individuals in the latent space based on these columns without any transformation, while other metrics such as Euclidean accepts only numerical ones (Bektas and Schumann 2019a):

$$\text{Sil} = \frac{b_i - a_i}{\max\{a_i, b_i\}}. \tag{3}$$

As for the clustering algorithm, we utilized the *k*-medoids clustering algorithm, since it is compatible with the latent space created by the Gower distance metric (Bektas and Schumann 2019a). However, *k*-medoids is an unsupervised algorithm; thus, we need to find ex-ante the optimal number of clusters. There are the goodness-of-fit metrics in the literature [e.g., Average Silhouette Width (ASW), Calinski and Harabasz Index (CH) and Pearson version of Hubert's Γ (PH) (Campello and Hruschka 2006)], which can provide quantitative measurement scores regarding the quality of clustering with the different number of clusters. The ASW is one of the most widely used approaches that measures how well an instance is matched with its own cluster (Maulik and Bandyopadhyay 2002; Bektas and Schumann 2019a). As a goodness-of-fit measure, it reflects how well intra-cluster homogeneity and inter-cluster dissimilarity are maximized (Rousseeuw 1987). The idea for pre-specifying the optimal number of clusters is to try different *k*-values in an interval and appoint one of them, which has the highest ASW value, as the optimal number of clusters. For each *k* number, the ASW value of the clusters is calculated according to Eq. (3), which depicts the Silhouette value of instance *i*. The feature a_i represents average dissimilarity of *i* to all other objects in the cluster *a* (the smaller the value, the better the assignment). Another feature b_i reflects the minimum dissimilarity of the instance *i* to all objects in any other cluster (the closest cluster to i except its own cluster). Equation (3) returns values between −1 and 1. Values close to 1 indicate that instance *i* is assigned to the proper cluster. Average Sil values of all instances (ASW) give an idea about the quality of the clustering (Rousseeuw 1987).

---

[2] We provide to readers an example latent space, which we generated for the experiment in Online Appendix A.

[3] Following the dominant language conventions in the distance metrics and cluster domains, columns mean attributes throughout the article. Rows are cases (agents) which are clusterised.

Algorithm 1 Pseudo-code of the k-Medoids Algorithm (Bektas and Schumann, 2019)

Input: m (a dissimilarity matrix), k (the number of clusters)
Output: k clusters
    Initialization: select arbitrary k of the n cases as the medoids
  1: **while** The Mean Square Error (MSE) decreases **do**
  2:    Distribute n cases to k medoids
  3:    Update medoids from the cases added to the cluster
  4:    Calculate MSE for each update
  5:    **if** MSE does not decrease anymore **then**
  6:      **break**
  7:    **end if**
  8: **end while**
  9: **return** *k clusters*

After the instances are placed in a latent space, and the optimal number of clusters are found, the *k*-medoids algorithm (see Algorithm 1) partitions the instances into *k* (the optimal number) clusters. To understand how well clusters from real and artificial data overlap, we compare the quantity of artificial and real individuals in the clusters according to the indicator (4). In the formulation of the indicator (4), *R* represents the number of real instances, *A* represents the number of artificial instances, and *N* is the optimal number of clusters. The indicator finds the dissimilarity in the balance of artificial and real instances for each cluster. Finally, it returns a normalized score in a spectrum between zero and one. The indicator uses the L1 norm (i.e., least absolute deviation) similarly to the Manhattan distance, since it gives equal importance to all clusters that might have different dissimilarities (i.e., balance differences).[4] Besides, the L1 form is more preferable for high-dimensional data applications (Aggarwal et al. 2001):

$$\frac{\sum_{k=1}^{N} = \frac{|R_k - A_k|}{R_k + A_k}}{N}. \tag{4}$$

If an artificial agent behaves observationally equivalent to the real individual with whom it has the same characteristic, they are placed in the same position in the latent space; thus, they are supposed to be in the same cluster. If all artificial agents behave observationally equivalent with the real individuals with whom they have the same characteristics, it is expected that the clusters would have 50% artificial fifty percent real instances (as in the simple experiment in Online Appendix A). In this case, the indicator's outcome (4) becomes zero, which indicates a perfect match. In other words, a zero score demonstrates that the behavior patterns in real data are perfectly overlapping with the ones from the artificial data. Conversely, if an artificial agent produces different ex-post behavior than his real counterpart, they are placed in different positions in the latent space. Thus, they are supposed to be in different clusters. That leads to unbalanced clusters and, consequently, a weak validation score according to the indicator in (4).

---

[4] The L2 norm puts more emphasis on the clusters with large balance discrepancies.

The overall concept is completed with the determination of the place of the score in the distribution of all possible scores it could theoretically take, which allows us to interpret it. To determine a meaningful threshold, we obtain all possible scores it can have and their frequency in the exhaustive list of all possible cases, which is the state space. The state space contains all possible alternative ways in which a total can be distributed,[5] with Page (2012) demonstrating a Java algorithm to obtain them in a broad variety of restrictions. In the case at hand, we study the scores that the indicator (4) generates in all possible subdivisions of the total number of artificial agents and of the total number of real individuals in the clusters. Accordingly, we obtain the distribution of possible scores, which allows us to judge the specific score—that a model achieves in the previous steps—concerning all other possible scores.

Overall, this procedure builds on the idea that a validated model should produce "indistinguishable" results from real data. Going beyond the inter-personal qualitative procedure proposed in Piana (2013), we deliver a method having a quantitative indicator of "goodness-of-validation," taking values from zero to one. The method can be used for the AB models having micro-data as input and produce results accordingly. We discuss such models and provide examples in "Discussion". The method provides these models two advantages: avoiding over-fitting of the micro-level validation and having more detailed validation than macro-level, as recommended in Fagiolo et al. (2019). In the next section, we apply the method to a specific model in the personal mobility domain, implementing a certain simulation platform.

Case study

This section consists of three subsections. In the first, we describe an AB simulation platform called Behavior Driven Demand Model (BedDeM) that we configured to build a specific model. The platform building process is discussed in Nguyen and Schumann (2019).[6] and a use case is addressed in Bektas et al. (2018) and Bektas and Schumann (2019b). In the second subsection, we discuss the model building process by configuring the generic platform with empirical (real) data. The proposed validation method is applied to the built model, and the results are discussed in the next section. In the third subsection, we discuss the specific variables that constitute individuals' ex-ante characteristics and ex-post behavior in the built model. These variables are used to place real and artificial individuals in a multidimensional latent space to compare meso-level patterns.

---

[5] We provided a numerical example in Online Appendix B regarding how the total can be distributed among clusters. We provide also example state spaces for different clustering configurations.

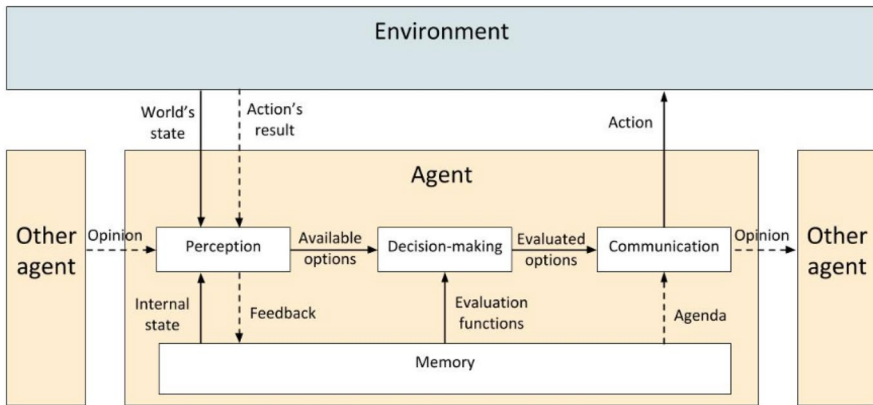[6] The source code of the platform is available online at BedDeM (2020).

Fig. 3 Overview of agent's design (Nguyen and Schumann 2019)

The simulation platform

General features

The BedDeM platform has been developed as a generic tool that can be configured to address specific issues from different research domains (e.g., household consumption, mobility, tourism, etc.). It comprises of the key theoretical tenets of the multi-agent cognitive system, in which heterogeneous and autonomous agents are capable of making choices (decisions). It enables modeling the micro-behavior of each individual (agent) separately.

The core element of the BedDeM platform is an agent-based simulator, written in Java based on the RePast library (Nguyen and Schumann 2019), complemented by key concepts from Triandis' Theory of Interpersonal Behavior (TIB) (Triandis 1979), described in "Agent's decision-making mechanism". TIB explains the origin of individual behavior, which is utilized as a decision-making framework (component) in the platform. Hence, agents make their choices (decisions) according to the contained determinants in the TIB.

Overview of agent's design

BedDeM consists of autonomous agents that have (not necessarily) heterogeneous characteristics and preferences. Agents are assigned tasks and are supposed to choose an option to perform their tasks according to the ex-ante defined behavioral rules (e.g., decision-making mechanism). Tasks and options are specified according to the application domain. For instance, agents might choose a tourist place to visit or choose a mobility mode to perform their trips, according to configuration.

When an agent performs a task, he first collects information. The perception module (see Fig. 3) gets information about the present state of the environment and
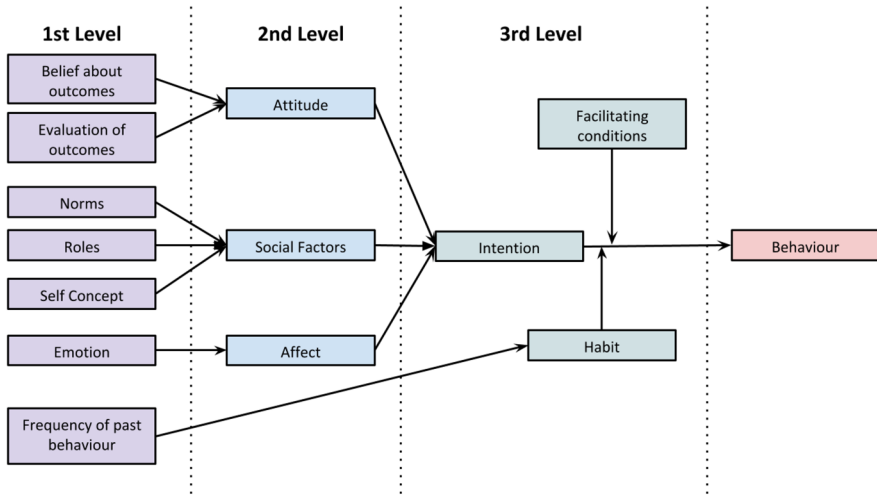
Fig. 4 Triandis TIB model (Triandis 1979)

combines it with other agents' opinions. It then brings the information to the deci-
sion-making module for reasoning. The obtained information is combined with het-
erogeneous preferences and also with the past decisions in the memory. As agents
maintain their local state (the individual level memory, see Fig. 3), the decision-
making becomes time-inseparable. In the end, the agent lists all available options
to perform the task and choose the most preferable one according to his individual
reasoning. After the choice, he informs the other agents about his choice that can be
used as information in others' decision-making modules.

Agent's decision-making mechanism

Triandis' theory of inter-personal behavior (*TIB*) Since micro-behavior is the main
output of the mDGP, it should be well specified to obtain precise emergent proper-
ties with the original system. While the BedDeM platform was being constructed,
first, the origin of individual behavior was addressed. The idea was obtaining a
standard theory that depicts the origin of individual behavior and using this theory
as agents' decision-making mechanism. In cognitive science, there exist such theo-
ries, e.g., Ajzen's theory of planned behavior (TBP) (Ajzen et al. 1991) and Ajzen
and Fischbein's theory of reasoned action (TRA) (Chang 1998). These theories state
that individual's intention to act is the key determinant of behavior (Bektas et al.
2018). There are several AB models and platforms they attempted to incorporate
these theories (Nguyen and Schumann 2019).

Triandis extended these theories in his TIB model (see Fig. 4). He added two
new components over them, habits and facilitating conditions. According to TIB,
the frequency of past behavior forms a habit that partly impacts current behavior.
Hence, the current behavior is determined by the current status of the environment
(e.g., economic parameters) and the previous decisions in the individual memory.
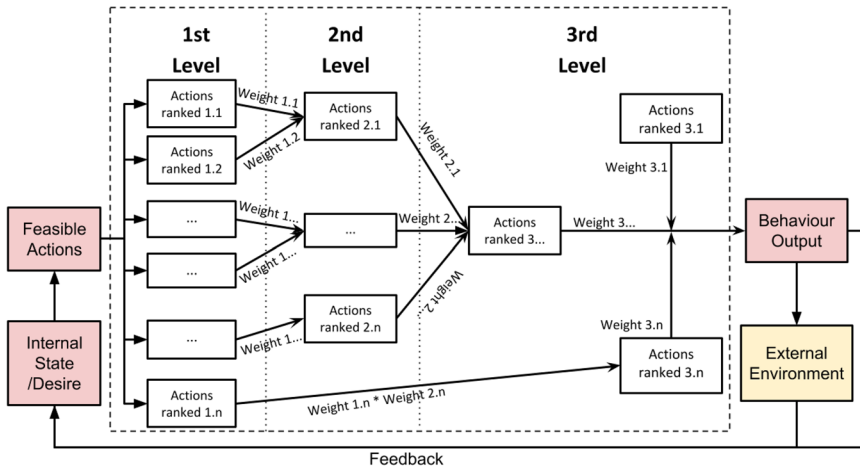
Fig. 5 Agent's decision-making mechanism with mapping to TIB (Bektas et al. 2018)

The theory as well as other empirical research state that intention is moderated by habit that leads to non-deliberate decision-making (Verplanken et al. 1994; Bamberg et al. 2003).

As Nguyen and Schumann (2019) state, TIB includes all aspects of TRA and TPB, as well as additional components such as habits that potentially improve its predictive power and descriptive accuracy. Although there is no proof which theory is more suited to build an AB platform, TIB was chosen for the BedDeM platform, since it provides a more comprehensive understanding of the origins of individual behavior.

Implementation of TIB as agent decision-making mechanism The full implementation of the TIB model as an agent decision-making mechanism is illustrated in Fig. 5. When a task is assigned to an agent, he first gets information from the environment to have available options to perform the task. Then, according to each determinant (d) (i.e., box in Fig. 5) in the first layer, the agents sorts available options (opt) in a list according to their score (see Eq. (5)). The score is calculated by comparing the property of an option with other's ($R_d(\text{opt})$). To calculate the scores in the first level, either a real numerical system (for quantitative determinants such as price) or a ranking function (for the determinants such as emotions) is utilized. Both numerical values and rankings can come from empirical data or be calibrated through experts' assessment (Nguyen and Schumann 2019):

$$R_d(\text{opt}) = \sum_{c=1}^{C} \left(R_c(\text{opt}) / \left(\sum_{o=1}^{O} R_c(o) \right) * w_c \right)$$

•R_d(opt) is the score of an option (opt) at determinant d.

•C is the set of the children of d(i.e., determinants connects with d (5)

in the previous level).

•O is the set of all available options.

•w_c is the weight of child determinant c.

Once all options are ranked in lists according to each determinant in the first layer, the lists are merged and normalized with associated weights (w_d) to pass in the next layer (see Eq. (5)). The score of each option according to each determinant is multiplied with the associated weight, which becomes the new score of the option. The weights in decision-making represent the importance of determinant. For instance, if it is desired to have time-separable decision-making, the weight of habit can set to zero, which means that the memory (i.e., past decisions) does not impact the current behavior. Once all decision-making steps are merged, the agent ends up with a sorted list of options according to their scores. According to the configuration, he can choose the first best option deterministically in the list, or certain probabilities can be created over the scores; thus, he can choose an option stochastically. More detailed information regarding the platform and its decision-making mechanism can be found in Nguyen and Schumann (2019).

Model building

We are currently applying the platform in the mobility domain. The BedDeM platform becomes an AB mobility mode-choice model through the configuration, which aims to generate heterogeneous mobility demands at the household level. The model allows for mode-choices for mobility trips based on price and non-price signals through its decision-making mechanism. It has the ability to generate yearly data that can be interpreted at the granularity of historical evolution of mobility, which largely hinges on aggregate kilometers traveled and emissions produced by mode, including possible decarbonization trajectories (Bektas et al. 2018).

Input of the model We utilize the "Mobility and Transport Micro-census (MTMC)" (ARE/BfS 2017) data of the Swiss statistical office to build the model. The data are at the micro-level and can be easily mapped to the agents. The data contain information regarding Swiss households' socio-economic characteristics (e.g., location, income level, car/travelcard ownership, etc.) and daily mobility activities. We map the real respondents one-to-one agents; thus, each agent represents a real Swiss household by having all his characteristics, including mobility activities. Besides, all the exogenous variables that are used to shape the environment, such as fuel prices, reflect the Swiss system.

Output of the model Each agent in the model is assigned a task list (i.e., trips of the real households) to perform. Agents evaluate existing options (e.g., car, public

transportation, soft mobility, etc.) according to the decision-making mechanism introduced in the previous section and choose a mode for each of their trips. The model simulates each agent's micro-behavior separately and generates micro-level heterogeneous mobility mode-choices as the core output. The output can be aggregated to obtain macro-patterns (i.e., modal-split) over which the model has already been calibrated (Nguyen and Schumann 2019), including using data from the Swiss Household Energy Demand Survey (Weber et al. 2017). Each agent has a weight-to-universe value, which is used as upscale factor to get macro-patterns. Through the aggregation, various sorts of outputs can be derived, e.g., total emissions and kilometers traveled per mode; thus, the model can be used to test climate change policies in-silico, for instance.

Variable selection for the case study

To apply the validation method, the variables identifying the ex-post behavior and the ex-ante characteristics of individuals should be chosen and given to the validation method as input (see Fig. 2). As ex-post behavior, mode-choice should be in the chosen variables (the last variable in the list below). As for the variables constituting ex-ante characteristics, we identified the ones influencing mode-choice in our previous research (Bektas and Schumann 2019a) and use them in this case study (the first four variables in the list below). The full list of the used variables is demonstrated below:

– Number of cars in the household
– Number of daily trips
– Having a half-fare travelcard
– Daily distance
– Mode-choice.

Results

We apply the overall concept step-by-step and discuss the results of each step sequentially. We commenced the overall concept with a merged data set containing 3000 artificial and 3000 real (MTMC) individuals with the chosen variables. Before clustering the individuals, we obtained the optimal number of clusters. We utilized the Average Silhouette Width (ASW) score that gives statistics to determine the optimal number of clusters. As illustrated in Fig. 6, we clustered individuals in the final data set into a different number of clusters within an interval ([2:15]) for each, we calculated the ASW score. The results show that we get the highest cluster quality when we cluster the individuals into six clusters. In other words, we obtain the optimal intra-cluster homogeneity and inter-cluster heterogeneity by dividing individuals into six clusters. We utilized the score as a statistical ground and used the obtained optimal number of clusters to proceed the method.
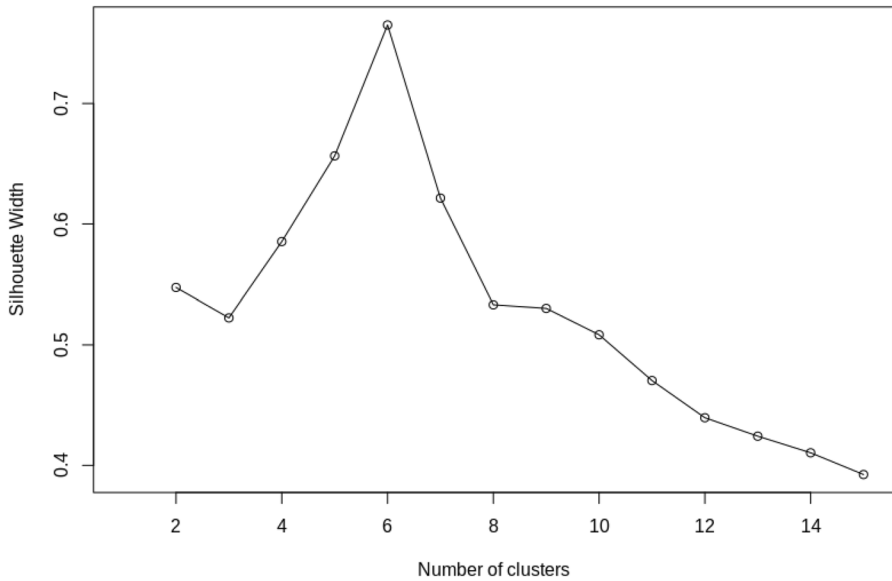
Fig. 6 ASW values with different cluster numbers

Table 1 Quantity of artificial agents and real instances in the clusters and corresponding scores

Clusters	Number of artificial agents	Number of real individuals	Score
Cluster 1	944	586	0.2339
Cluster 2	1329	990	0.1461
Cluster 3	203	439	0.3676
Cluster 4	175	382	0.3716
Cluster 5	164	322	0.3251
Cluster 6	185	281	0.2060
Overall	3000	3000	0.2750

We placed the individuals into a multidimensional latent space based on the chosen variables, divided them into six clusters according to their positions in the latent space, and analyzed such clusters' composition. For each cluster, we got the quantities of artificial and real individuals to compute the indicator. The obtained quantities are demonstrated in Table 1.

After we obtained the quantity of artificial and real instances in the clusters, we applied the indicator (4) as the overall concept indicated and got the value `0.2750`. By construction, it is between zero and one; the lower, the stronger the validation. But how to judge this specific value in general (e.g., independently on the number of clusters)? As anticipated in "Methods", we iterate the computation of this indicator (4) for all possible cases (i.e., balance combinations), which is the matrix product of two identical state spaces. An example of such a case is the situation where the 3000 real agents are all in one cluster. That can be matched by the situation in
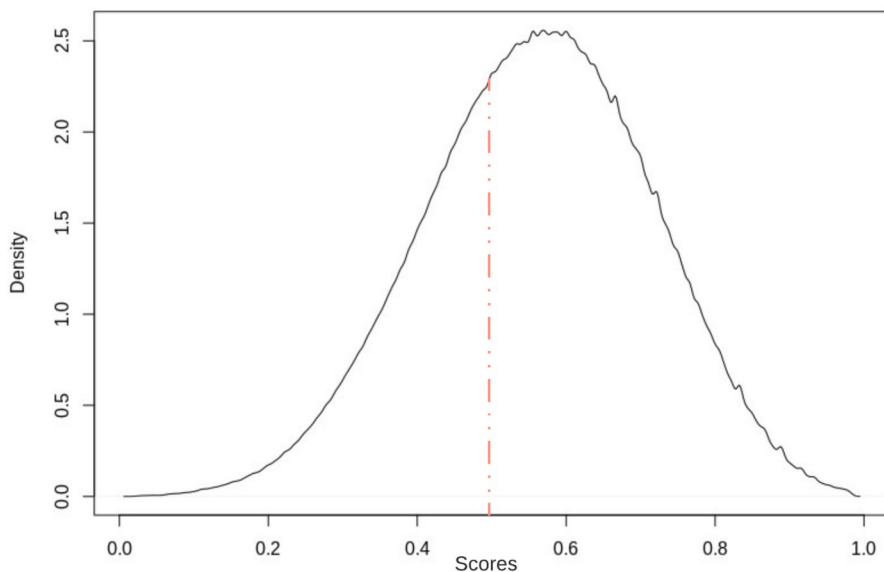
Fig. 7 Density distribution of all possible validation scores

which all 3000 artificial agents are in the same cluster (good) or in another cluster (bad). Alternatively, 2750 artificial agents are in that cluster or in another. Examples like this are many thousands, but Page (2012) provides a computational method to elicit all of them. It computes not only how many but also enlist which ones they are. Mathematically speaking, it generates the weak composition of 3000 in 6. Since the full number is way too high to be computed in a reasonable time, we first quantize and then fit the results with a continuous function. We quantize the 3000 in 20 groups of 150 units each (in a procedure that is similar to bootstrap). We perform Page's algorithm in what Piana et al. (2020) would call shapes (20, 6): a state space enlisting the ways in which 20 units (in our case groups of agents) can be separated into six classes (in our case: clusters). The code to compute this state space is distributed as complementary material to Piana et al. (2020), drawing on Page (2012), McGhee (2008) and McGhee (2006).

For each of its rows, the outcome of the indicator can be computed. By having 20 groups (i.e., 150 quantum size of 3000 individuals) and 6 clusters, we obtained two states (for artificial and real individuals), each with around 50,000 combinations. For all possible number of artificial and real individual distributions in the clusters, we applied the indicator. Then, we obtained the density distribution of all the possible scores (i.e., outcomes of the indicator), which is illustrated in Fig. 7. Thus, we defined the space to see where our model's validation score is, which enables us to judge the score.

Thanks to the computed all possible scores, we could easily judge the specific score that the model achieves. We summed up the number of cases that have a better score than 0.2750. We divided it into the total number of cases to obtain the percentage. By this means, we calculate the area under the curve (integral) in Fig. 7.

The results show us that approximately 4.2% of cases would produce a score equal to or lower than the score of the model, `0.2750`. We interpret these results in this way that the model is validated at the conventional threshold of 5%.

Discussion

The findings show that the built model for the case study satisfactorily represents the original system at the meso-level for the given variables. The artificial agents in the model behave observationally similar to the real individuals, who have the same ex-ante given characteristics. It can be interpreted in this way that the mDGP mimics the rwDGP by producing observationally similar data (behavior) with the given input data.

To judge the validation score that we obtained, we created the density distribution of all possible scores, as explained in "The overall concept of the meso-level validation method". We utilized a simplification by setting the quantum size 150 to reduce high computational time. We tested whether the quantum size is relevant for the density distribution by applying another quantum size (300) (see Fig. 8 in Online Appendix C). We compared the functional forms of the density distributions to see whether different quantum sizes lead to different curves. It was observed that both 150 and 300 quantum sizes produce almost identical curves. We report it to demonstrate that the result is robust to the changes in simplified assumptions. Additionally, in Online Appendix B, we provide two state spaces without quantization for the cases having a lower number of agents.

Discussion of the potential application of the meso-level approach to validation upon further models

In this subsection, we discuss certain insights that we gained during the implementation in the case study that may turn out to be useful to other researchers who want to assess the empirical validation of a certain AB model drawing on micro-data.

Many AB models do not draw upon real data, and for that group, the method cannot be applied. However, if the modeler's golden rule laid down in Piana (2004) is followed, and agents are given rules that can be directly embedded in questionnaires to real people, then by actually carrying out such surveys, the modeler can have at her/his disposal micro-data with which he can initialize the artificial agents. Indeed, this is often the case: AB models are frequently built and initialized with micro-data, since they aim to model heterogeneous behavior of each individual separately and in a highly realistic way (Dawid et al. 2012). Such micro-data contain information at the individual level and can be mapped to artificial agents. Thus, artificial agents get their ex-ante characteristics from real individuals, and they are supposed to generate ex-post behavior according to the behavioral rules of the mDGP. As long as the ex-post behavior of real individuals is known, the proposed empirical validation method can be applied to an AB model drawing on micro-data for which a meso-level can be computed, at which agents can be clustered and compared. For instance, in the

model that we built for the case study, agents represent households and it produces mobility mode-choice behavior. Since we also know real households' behavior with the same characteristics, we could apply the method in a rather straightforward way. In another AB model, agents might represent real firms, and the micro-data containing information regarding real firms may come from accounting systems and declarations to the statistical offices. Real and artificial firms' behavior is clustered with their characteristics as the proposed method suggests and can be compared at the meso-level. Conversely, if in a macroeconomic AB model, there is a wide range of types of agents (firms, households, financial institutions, public institutions, etc.), our procedure might become too cumbersome if applied to all such types. In other words, the method is not dependent on the domain (scope), but its applicability is restricted to AB models for which a meso-level can be computed from available micro-data, possibly of only one type.

In the procedure of clustering, one needs to select the variables upon which clustering occurs and determine the optimal number of clusters. After that, the application of the indicator in (4) can be carried out. The variables should be available for both the artificial and the real agents; they should be relevant for the main behavior that the model is called to describe. In our use case, we used the variables that a previous analysis demonstrated having a large impact on the behavior. However, if one cannot proceed with such an analysis, one might take the neutral stance of taking all common variables across real and artificial agents. The optimal number of clusters can be obtained as we did (by taking the number of clusters for which ASW is maximal), but any method that would single out a non-arbitrary number of clusters might be used, if appropriate. Finally, one needs to compute the probability of the goodness-of-validation to be higher than a certain threshold, much alike the p value. This probability is to be computed using the procedure indicated before.[7]

An overview of AB models that might be validated with our method

Keeping into account its general requirements, our methodology can be applied to many AB models such as the ones introduced in Axtell et al. (2014), Nelson et al. (2015), de Koning and Filatova (2020), and Klein et al. (2020). One should not expect that the authors did utilize our novel methodology to validate their models, and thus, their current validation method is inevitably different from those we are proposing. However, the description of the data they utilized for their AB model suggests applicability. Moreover, in their text, they commit to a certain vision we share: "we seek two classes of data to feed analysis and modeling: micro-data and event data. This fine resolution is necessary if we assume heterogeneous decision-making, a hallmark of agent-based modeling. Aggregate statistics are insufficient. We need to have realistic household socio-demographic variables and resource endowments." (Nelson et al. 2015, p. 1). "We live in the era of 1-to-1 computational instantiations of many complex systems, and agent-based computing is a way for

---

[7] One can download the state space from App. 1 of Piana et al. (2020) or compute it by executing the Java code, downloadable from Online Appendix B of the same paper.

economics to join this zeitgeist of digital synthesis" (Axtell et al. 2014, p. 3). Axtell et al. (2014) provides the methodological explanation of an AB model of a metropolitan housing market, which has been extended to the national level by Geanakoplos et al. (2012), which in turn has been considered the best model to cover such issue by Carstensen (2015).

Klein et al. (2020) describe an AB model of the diffusion of electric vehicles. The initialization of its agents comes from micro-data collected by their original experiment (a conjoint-analysis) conducted with 552 people, representing the German population. "Parametrization and initialization of the characteristics and behavior of consumer agents was done using empirical data from our own study. Using these data, each consumer agent of the ABS was then initialized based on the corresponding characteristics of one real participant from our empirical study. Note, we also simulated larger populations in our sensitivity analysis. However, owing to relatively stable results, we decided to use 552 exactly matching consumers, which significantly reduced the time of each simulation run. Additionally, this allowed the direct initialization of each consumer agent using the responses of exact one consumer from our empirical study" (Klein et al. 2020, p. 12).

de Koning and Filatova (2020) do not only describe an AB model to explore how urban housing markets evolve in the presence of climate-driven floods and behavioral biases on the agent level for which an ad-hoc survey of 600 respondents has been utilized to initialize the agents, but it explicitly calls for multi-scale validation. It falls short of singling out the meso-level as particularly appropriate for validation, which is our novel claim. Moreover, in recognizing that "there is no definite answer as to how much empirical validation is enough in order to make a model useful for its purpose" (p. 139), it implicitly valorizes our attempt to provide a metrics and a quantitative test with a threshold that can give a satisfactory interruption of a potentially never-ending cycle of reparametrizations ("Validation can be a continuous iterative process", as this paper puts it at p. 139). Indeed, it is important to remark that after calibration and validation, and finally, our models need to produce results. For instance, after validating the model that we built by configuring the BedDeM platform, we have been generating 320 alternative scenarios of mobility evolution 2015–2050 for Switzerland (currently delivered in an internal document for the funding agency).

Conclusion

Summary

The present work proposes an unsupervised machine learning algorithm—cluster analysis—as a meso-level empirical validation method for AB models drawing on micro-data. The model aims to cluster the ex-post behavior of real and artificial individuals with the same ex-ante given characteristics. It produces a validation score in [0, 1] by comparing the similarity among clusters. The clusters do not only contain the ex-post behavior of real and artificial individuals but also their ex-ante given characteristics that influence the behavior. Hence, comparing clusters enables us

to compare behavioral patterns in model-generated data and real data. To provide an instance of application of the method, we referred to an AB model that aims to model heterogeneous mobility mode-choice behavior. The specific model obtained a satisfactory validation score that shows that, in this case, the mDGP can mimic the rwDGP successfully for the given variables. More, in general, the proposed empirical validation method has certain advantages. First, it fully leverages the specificity of agent-based models covering highly heterogeneous agents and their potential multi-level aggregation. An agent-based model can be initialized at the micro-level, be calibrated at the macro-level, and be validated at the meso-level with the same data set and for the same time frame. A procedure that is often used in time-series to calibrate the model for a first segment of time periods and then validate in out-of-sample successive time suffers from the necessity of assuming that there are no structural breaks over time. This assumption may not be particularly suitable for models looking for emerging properties, high non-linearities, and, indeed, structural breaks. The second advantage is that with the meso-level validation, we can compare the behavioral patterns that the mDGP and the rwDGP generate, respectively. It is not easy with macro-level validation, because it compares only the aggregates. Therefore, the relationship between the ex-ante given characteristics and ex-post generated behavior cannot be easily compared. In short, we offer to the community of researchers devising and using agent-based models a method to empirically validate them, which is a crucial intermediate step in the overall useful application of this highly promising approach.

Future work

We envisage different dimensions in the frame of future works to take the present work forward. First, as discussed in the related work section, the simulated minimum distance (SMD) methods, including the method of simulated moments (MSM), can be complementary to our method. A model's parameters can be estimated by an SMD method at the macro-level and its output can be validated by our method at the meso-level. We plan to research about the coupling of an SMD method and our method to use them together on the same model. Second, as our method aims to validate AB models drawing on micro-data, we aim to introduce a new technique generating synthetic micro-data from macro-aggregates for the modelers having limited access to micro-data. Third, we aim to assess the impact of the number of agents and the optimal number of clusters on the method's results in detail. We plan to apply the method on AB models having micro-data from different original systems and domains. Finally, we aim to explore the situations in which the output of an AB model is observationally similar with the real data at the macro-level but not at the meso-level.

Availability of data and material The supporting information regarding the data source can be found in Bektas and Schumann (2019a) and ARE/BfS (2017).

Code availability The source code of the agent-based simulation platform is available online at BedDeM (2020). The Java code that is used to create state spaces (morphospaces) and the density distribution can be found in Page (2012).

Declarations

Conflict of interest The authors declare no conflicts of interest/competing interests.

References

Aggarwal CC, Hinneburg A, Keim DA (2001) On the surprising behavior of distance metrics in high dimensional space. In: International conference on database theory. Springer, pp 420–434

Ajzen I et al (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211

ARE/BfS (2017) Verkehrsverhalten der Bevölkerung Ergebnisse des Mikrozensus Mobilität und Verkehr 2015. Federal Office for Spatial Development and Swiss Federal Statistical Office

Arthur WB (1994) Increasing returns and path dependence in the economy. University of Michigan Press

Arthur WB (2006) Out-of-equilibrium economics and agent-based modeling. Handb Comput Econ 2:1551–1564

Axtell R, Farmer D, Geanakoplos J, Howitt P, Carrella E, Conlee B, Goldstein J, Hendrey M, Kalikman P, Masad D et al (2014) An agent-based model of the housing market bubble in metropolitan Washington, DC. In: Whitepaper for Deutsche Bundesbank's spring conference on "Housing markets and the macroeconomy: challenges for monetary policy and financial stability

Balci O (1994) Validation, verification, and testing techniques throughout the life cycle of a simulation study. Ann Oper Res 53(1):121–173

Bamberg S, Ajzen I, Schmidt P (2003) Choice of travel mode in the theory of planned behavior: the roles of past behavior, habit, and reasoned action. Basic Appl Soc Psychol 25(3):175–187

Barde S (2016) Direct comparison of agent-based models of herding in financial markets. J Econ Dyn Control 73:329–353

Barde S (2020) Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion. J Econ Dyn Control 111:103795

Barde S, Van Der Hoog S (2017) An empirical validation protocol for large-scale agent-based models. Bielefeld Working Papers in Economics and Management

BedDeM (2020) Github—silab-group/beddem_simulator. https://github.com/SiLab-group/beddem_simulator. Accessed 16 July 2020

Beisbart C, Saam NJ (2019) Computer simulation validation. Springer

Bektas A, Schumann R (2019a) How to optimize Gower distance weights for the k-medoids clustering algorithm to obtain mobility profiles of the swiss population. In: 2019 6th Swiss conference on data science (SDS). IEEE, pp 51–56, ISBN:978-1-7281-3105-4

Bektas A, Schumann R (2019b) Using mobility profiles for synthetic population generation. In: Proceedings of the social simulation conference 2019, Mainz, Germany

Bektas A, Nguyen K, Piana V, Schumann R (2018) People-centric policies for decarbonization: testing psycho-socio-economic approaches by an agent-based model of heterogeneous mobility demand. In: Computing in economics and finance (CEF) conference

Bianchi C, Cirillo P, Gallegati M, Vagliasindi PA (2007) Validating and calibrating agent-based models: a case study. Comput Econ 30(3):245–264

Campello RJ, Hruschka ER (2006) A fuzzy extension of the silhouette width criterion for cluster analysis. Fuzzy Sets Syst 157(21):2858–2875

Carstensen CL (2015) An agent-based model of the housing market. Steps toward a computational tool for policy analysis. University of Copenhagen, MSc-szakdolgozat

Chang MK (1998) Predicting unethical behavior: a comparison of the theory of reasoned action and the theory of planned behavior. J Bus Ethics 17(16):1825–1834

Colander D, Howitt P, Kirman A, Leijonhufvud A, Mehrling P (2008) Beyond DSGE models: toward an empirically based macroeconomics. Am Econ Rev 98(2):236–40

Dawid H, Gemkow S, Harting P, Van der Hoog S, Neugart M (2012) The eurace@ unibi model: an agent-based macroeconomic model for economic policy analysis. Bielefeld working papers in economics and management

Dawid H, Gemkow S, Harting P, van der Hoog S, Neugart M (2014) Agent-based macroeconomic modeling and policy analysis: the eurace@ unibi model. Bielefeld Working Papers in Economics and Management

Duffy J (2006) Chapter 19 agent-based models and human subject experiments. Volume 2 of handbook of computational economics

Fagiolo G, Moneta A, Windrum P (2007) A critical guide to empirical validation of agent-based models in economics: methodologies, procedures, and open problems. Comput Econ 30(3):195–226

Fagiolo G, Guerini M, Lamperti F, Moneta A, Roventini A (2019) Validation of agent-based models in economics and finance. In: Computer simulation validation. Springer, pp 763–787

Farmer JD, Foley D (2009) The economy needs agent-based modelling. Nature 460(7256):685

Geanakoplos J, Axtell R, Farmer JD, Howitt P, Conlee B, Goldstein J, Hendrey M, Palmer NM, Yang CY (2012) Getting at systemic risk via an agent-based model of the housing market. Am Econ Rev 102(3):53–58

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 857–871

Grazzini J, Richiardi M (2015) Estimation of ergodic agent-based models by simulated minimum distance. J Econ Dyn Control 51:148–165

Grünwald PD, Grunwald A (2007) The minimum description length principle. MIT Press

Guerini M, Moneta A (2017) A method for agent-based models validation. J Econ Dyn Control 82:125–141

Hamill L, Gilbert GN (2016) Agent-based modelling in economics. Wiley Online Library

Heckbert S, Baynes T, Reeson A (2010) Agent-based modeling in ecological economics. Ann N Y Acad Sci 1185(1):39–53

van der Hoog S (2019) Surrogate modelling in (and of) agent-based models: a prospectus. Comput Econ 53(3):1245–1263

Kassambara A (2017) Practical guide to cluster analysis in R: unsupervised machine learning, vol 1. Sthda

Klein M, Lüpke L, Günther M (2020) Home charging and electric vehicle diffusion: agent-based simulation using choice-based conjoint data. Transp Res Part D: Transp Environ 88:102475

Klügl F (2008) A validation methodology for agent-based simulations. In: Proceedings of the 2008 ACM symposium on applied computing. ACM, pp 39–43

de Koning K, Filatova T (2020) Multi-scale validation of an agent-based housing market model. In: Advances in social simulation. Springer, pp 135–140

Lamperti F (2018a) Empirical validation of simulated models through the GSL-div: an illustrative application. J Econ Interact Coord 13(1):143–171

Lamperti F (2018b) An information theoretic criterion for empirical validation of simulation models. Econom Stat 5:83–106

Macal C, North M (2014) Introductory tutorial: agent-based modeling and simulation. In: Proceedings of the winter simulation conference 2014. IEEE, pp 6–20

Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. IEEE Trans Pattern Anal Mach Intell 24(12):1650–1654

McGhee G (2008) Convergent evolution: a periodic table of life, pp 17–31

McGhee GR (2006) The geometry of evolution: adaptive landscapes and theoretical morphospaces. Cambridge University Press

Murray-Smith DJ (2015) Testing and validation of computer simulation models. Springer, Cham. https://doi.org/10:978-3

Nelson JB, Kennedy WG, Greenberg AM (2015) Agents and decision trees from microdata. In: Proceedings of the 24th conference on behavior representation in modeling and simulation (BRIMS)

Nguyen K, Schumann R (2019) On developing a more comprehensive decision-making architecture for empirical social research. In: The 20th international workshop on multi-agent-based simulation (MABS 2019), 13 May 2019

Page DR (2012) Generalized algorithm for restricted weak composition generation. J Math Model Algorithms Oper Res 12(4):345–372

Piana V (2004) Consumer decision rules for agent-based models. Economics Web Institute

Piana V (2013) Routines. Economics Web Institute EWI concepts series

Piana V, Bektas A, Khoa N (2020) Temporal morphogenesis. Economics Web Institute EWI Essay series

Pyka A, Fagiolo G (2007) 29 Agent-based modelling: a methodology for neo-Schumpeterian economics'. Elgar companion to neo-schumpeterian economics 467

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65

Russell S, Norvig P (2002) Artificial intelligence: a modern approach, 3rd edn

Tesfatsion L (2002) Agent-based computational economics: growing economies from the bottom up. Artif Life 8(1):55–82

Triandis HC (1979) Values, attitudes, and interpersonal behavior. In: Nebraska symposium on motivation, University of Nebraska Press

Verplanken B, Aarts H, Van Knippenberg A, van Knippenberg C (1994) Attitude versus general habit: antecedents of travel mode choice 1. J Appl Soc Psychol 24(4):285–300

Weber S, Burger P, Farsi M, Martinez-Cruz AL, Puntiroli M, Schubert I, Volland B (2017) Swiss household energy demand survey (SHEDS): objectives, design, and implementation. Technical report, IRENE Working Paper, SCCER CREST

Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: alternatives and prospects. J Artif Soc Soc Simul 10(2):8

Xiang X, Kennedy R, Madey G, Cabaniss S (2005) Verification and validation of agent-based scientific simulation models. In: Agent-directed simulation conference, vol 47, p 55

Authors and Affiliations

Alperen Bektas[1] [ID] · **Valentino Piana[1]** · **René Schuman[1]**

Valentino Piana
valentino.piana@hevs.ch

René Schuman
rene.schumann@hevs.ch

[1]   HES-SO Valais/Wallis, Rue de Technopole 3, 3960 Sierre, Switzerland