

Furtado, Bernardo Alves; Andreão, Gustavo Onofre

**Working Paper**

## Machine learning simulates agent-based model towards optimal policy: A surrogate model for public policy assessment

Discussion Paper, No. 273

**Provided in Cooperation with:**

Institute of Applied Economic Research (ipea), Brasília

*Suggested Citation:* Furtado, Bernardo Alves; Andreão, Gustavo Onofre (2023) : Machine learning simulates agent-based model towards optimal policy: A surrogate model for public policy assessment, Discussion Paper, No. 273, Institute for Applied Economic Research (ipea), Brasília, <https://doi.org/10.38116/dp273>

This Version is available at:

<https://hdl.handle.net/10419/298106>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**MACHINE LEARNING SIMULATES  
AGENT-BASED MODEL TOWARDS  
OPTIMAL POLICY: A SURROGATE  
MODEL FOR PUBLIC  
POLICY ASSESSMENT**

**BERNARDO ALVES FURTADO  
GUSTAVO ONOFRE ANDREÃO**

## **MACHINE LEARNING SIMULATES AGENT-BASED MODEL TOWARDS OPTIMAL POLICY: A SURROGATE MODEL FOR PUBLIC POLICY ASSESSMENT**

**BERNARDO ALVES FURTADO<sup>1</sup>  
GUSTAVO ONOFRE ANDREÃO<sup>2</sup>**

---

1. Researcher at the Department of Sectoral Studies and Policies, of Innovation, Regulation and Infrastructure of the Institute for Applied Economic Research (Diset/Ipea); and National Council for Scientific and Technological Development. E-mail: <bernardo.furtado@ipea.gov.br>.

2. Researcher in the National Development Research Program (PNPD) at Diset/Ipea. PhD student of Economics at the Institute of Economics of the State University of Campinas (IE-Unicamp). E-mail: <gustavo.93.andrea@gmail.com>.

**Federal Government of Brazil**

**Ministry of Planning and Budget**

**Officer** Simone Nassar Tebet

**ipea** Institute for Applied  
Economic Research

A public foundation affiliated to the Ministry of Planning and Budget, Ipea provides technical and institutional support to government actions – enabling the formulation of numerous public policies and programs for Brazilian development – and makes research and studies conducted by its staff available to society.

**President**

**LUCIANA MENDES SANTOS SERVO**

**Director of Institutional Development (substitute)**

**FERNANDO GAIGER SILVEIRA**

**Director of Studies and Policies of the State,  
Institutions and Democracy (substitute)**

**LUSENI MARIA CORDEIRO DE AQUINO**

**Director of Macroeconomic Studies and Policies (substitute)**

**CLÁUDIO ROBERTO AMITRANO**

**Director of Regional, Urban and Environmental  
Studies and Policies (substitute)**

**ARISTIDES MONTEIRO NETO**

**Director of Sectoral Studies and Policies,  
of Innovation, Regulation and Infrastructure (substitute)**

**FERNANDA DE NEGRI**

**Director of Social Studies and Policies (substitute)**

**CARLOS HENRIQUE LEITE CORSEUIL**

**Director of International Studies (substitute)**

**FÁBIO VÉRAS SOARES**

**General Coordinator of Press and Social Communication**

**JOÃO CLÁUDIO GARCIA RODRIGUES LIMA**

**Ombudsman: <http://www.ipea.gov.br/Ouvidoria>**

**URL: <http://www.ipea.gov.br>**

## Discussion Paper

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

© Institute for Applied Economic Research – ipea 2023

Discussion paper / Institute for Applied Economic  
Research.- Brasília : Rio de Janeiro : Ipea, 1990-

ISSN 1415-4765

1. Brazil. 2. Economic Aspects. 3. Social Aspects.  
I. Institute for Applied Economic Research.

CDD 330.90.908

Ipea publications are available for free download in PDF (all) and EPUB (books and periodicals).

Access: <http://www.ipea.gov.br/portal/publicacoes>

The opinions expressed in this publication are of exclusive responsibility of the authors, not necessarily expressing the official views of the Institute for Applied Economic Research and the Ministry of Planning and Budget.

Reproduction of this text and the data contained within is allowed as long as the source is cited. Reproduction for commercial purposes is prohibited.

JEL: C63; H71; R38.

DOI: <http://dx.doi.org/10.38116/dp273>

# CONTENTS

## ABSTRACT

1 COMPUTATIONAL MODELS FOR POLICY .....	6
2 LITERATURE BACKGROUND .....	8
3 METHODS .....	13
4 RESULTS .....	19
5 DISCUSSION .....	28
REFERENCES .....	30
APPENDIX .....	37
1 PERCENTAGE BY POLICY .....	37
REFERENCES .....	43

## ABSTRACT

Public policies are not intrinsically positive or negative. Rather, policies provide varying levels of effects across different recipients. Methodologically, computational modeling enables the application of a combination of multiple influences on empirical data, thus allowing for heterogeneous response to policies. We use a random forest machine learning algorithm to emulate an agent-based model (ABM) and evaluate competing policies across 46 Metropolitan Regions (MRs) in Brazil. In doing so, we use input parameters and output indicators of 11,076 actual simulation runs and one million emulated runs. As a result, we obtain the optimal (and non-optimal) performance of each region over the policies. Optimum is defined as a combination of production and inequality indicators for the full ensemble of MRs. Results suggest that MRs already have embedded structures that favor optimal or non-optimal results, but they also illustrate which policy is more beneficial to each place. In addition to providing MR-specific policies' results, the use of machine learning to simulate an ABM reduces the computational burden, whereas allowing for a much larger variation among model parameters. The coherence of results within the context of larger uncertainty – vis-à-vis those of the original ABM – suggests an additional test of robustness of the model. At the same time the exercise indicates which parameters should policymakers intervene, in order to work towards precise policy optimal instruments.

**Keywords:** agent based model; machine learning; public policies comparison; metropolitan areas, Brazil.

## 1 COMPUTATIONAL MODELS FOR POLICY

Policymakers know that policies, as broad as they may be, lead to different results in different contexts (Stone, 2017; Mitchell and Woodman, 2010; Gawel, Strunz and Lehmann, 2016). Why certain policies impact different regions differently, however, is difficult to pinpoint (Cravo and Resende, 2013; Boschma, 2013; Lundvall, 2007). Interest groups, institutions and context may be relevant factors for how policies are elaborated, implemented, revised and assessed (Mueller, 2020; Nelson, 1994; Hochstetler and Kostka, 2015). In that sense, existing structural factors, such as households attributes, income and location, along with characteristics of businesses and government might play a role (Faber, Valente and Janssen, 2010).

Computational models may act as a coherent set of sequential rules and procedures that help understand, explore, and learn about effects and impacts of policies (Tsfatsion, 2011). Models allow for the anticipation of probable results (Marchi and Page, 2014; Faber, Valente and Janssen, 2010). Additionally, computational models allow the possibility of counterfactual tests and scenarios, thus amplifying the available set of information that policy makers have prior to the elaboration of policies and after their implementation<sup>1</sup> (Lespagnol and Rouchier, 2018; Huang, Zheng and Chia, 2010). Policymakers facing complex scenarios would be in a better position to make decisions when they have subsidies from an array of factual and counterfactual indicators<sup>2</sup> (Brenner and Werker, 2009; Schmidt et al., 2016).

Agent-based models (ABMs) fit a class of computational models in which the emphasis lies on the agents of a system, the minimum decision-making units, and their interactions (Tsfatsion, 2003; Arthur, 2006). It is a bottom-up approach in which agents follow deterministic rules and interact to reproduce known patterns (Axelrod, 1997; Janssen and Ostrom, 2006). Moreover, ABMs are discrete, dynamic systems in which agents of various kind and their environment co-evolve following algorithmic steps (Epstein and Axtell, 1996; Vazquez and Hallack, 2018). ABMs are tools that social scientists, in particular, may use to systematically codify tacit knowledge and behaviors and “animate” trajectories in order to experiment (Galán et al., 2009; Chen, Lux and Marchesi, 2001; Ehrentreich, 2008).

Specifically, PolicySpace2 (PS2) is a spatial-economic ABM that uses census and geographical data from households, businesses and municipalities within Metropolitan Regions (MRs) to

---

1. Such counterfactuals became a must for simulation models, especially models that deal with large systems, e.g. Lamperti et al. (2018).

2. Arthur (1994) understands that inductive reasoning plays an important part on the policymaking activity.

simulate interactions in the housing, labor, goods and services markets. The empirical, intraurban model runs from 2010 to 2020 and is used to simulate three policy tests. Validation is made for Brasilia, the capital of Brazil, and additional results are also presented to four other MRs (Furtado, 2022). We use PS2 as our simulation baseline model in this paper.

The results of PS2 consistently show that rental vouchers (b) and monetary transfers (c) seem to be the most beneficial policy choice producing higher gross domestic product (GDP) with less inequality. These two options also reduced the percentage of households defaulting on rent and going any month without goods and services consumption. The policy choice of house (a) seems to perform worse in comparison to the two alternatives and the no-policy baseline across most of the indicators.

Furtado (2022) ran the ABM model with limited parameter space and counterfactual analyses for different combinations of parameters and rules. Models that use geographical information are normally large and time-consuming to run. Indeed, computational costs impose constraints to the size of simulation tests. In order to broaden the analysis without such impairments, a surrogate model is a possible solution.

We use PS2 open source as our simulation baseline model for this paper. However, whereas PS2 ran manual-hundreds-combinations of parameters for its sensitivity analysis, we design a surrogate model that expands the analyses to one million possible combinations.

A surrogate model is a computational artifice that simultaneously mimics the procedures of a model using its inputs and outputs, whereas saving time and maximizing outputs. "An emulator is thus a model of a model: a statistical model of the simulator, which is itself a mechanistic model of the world".<sup>3</sup> Surrogate models provide a simple and cheap option to emulate the original run of a model to a higher order of magnitude (van der Hoog, 2019; ten Broeke et al., 2021).

Given this context, our research question is: *do the same policy produce the same results in different MRs regarding a pre-established optimal result?* In other words, we would like to know whether each policy affects the different MRs heterogeneously or if there is a one-size-fits-all policy for all places. After answering such research question, we explore whether there are model parameters and rules that would have higher (or lower) values associated to optimal scenarios. In other words, if there is heterogeneity in MR response to different policies, what could be the

---

3. Available at: <<http://bit.ly/3V7LFt>>.



factors for such responses. Thus, which ones would demand (if possible) policymakers' intervention in order to nudge their MR towards optimized scenarios?

In order to answer these questions we apply Machine Learning (ML) procedures to create a surrogate model that emulates an ABM, PS2, and explores results of three alternative policies against the no-policy baseline for an array of 46 Brazilian MRs. As such, we anticipate empirical scenarios for each policy and each MR and contrast these results with the absence of policy. Results expand the test originally made by the ABM – which implied a quasi-manual test of 11 thousand runs on mostly five MRs – and amplifies the combination of parameters values to one million runs.

We chose a relational, comparative optimum ex-post target for our policy analysis. Basically, the optimal status results simultaneously imply that the MR has achieved the highest quartile of GDP and the lowest quartile of Gini coefficient. This choice was made to train the ML and execute the expansion of runs into a larger scope that included all parameters and MRs. Hence, optimal in this case refers to GDP and Gini coefficient indicators comparatively to all other MRs.

The contribution of this paper is to expand the empirical results of a policy-test ABM while bringing more uncertainty and more robustness into the simulation process. More uncertainty as we emulate the ABM with a much larger set of combinations among parameters that enter the model, compared to the original ABM run. More robustness as we explore results over one million outputs instead of the original 11 thousand simulation runs.

Additionally, we produce a rank of MRs intrinsic attributes that measures their response to a specific pair of optimal policy indicator targets: Gini and GDP. The response also illuminates the best policy for each individual MR. That comparison would not be possible out of the context of what represents the optimal, given by the ML environment. Finally, we present a surrogate model that qualify as a valid approach for spatial ABM generalization.

## 2 LITERATURE BACKGROUND

In this section, we discuss the pertinent literature background. First, we review the literature on ABM and their use for policy analysis. Then we review the literature on ML and its intersection with ABM: surrogate models.

## 2.1 ABMs and public policies

ABMs are simulation models that highlight interactions between decision-making units, the agents, and their local environment. ABMs basically consist of a discrete system in which the states of the agents and the environment are updated according to explicit rules (Epstein and Axtell, 1996).

ABMs are useful when the dynamic of events matter, thus, making ABMs adequate to model feedback, phase transitions and reinforced learning, for example. Moreover, ABMs are flexible enough when there is agent heterogeneity of, path-dependence, and space is embedded in the decision-making process (Batty, 2021; Taillandier et al., 2019). Being a bottom-up approach, ABMs start from the decision-units being modeled, and from their initial states and their rules, macro dynamics emerge. The conceptualization of ABMs suggests the need for a “sufficient” number of decision-units that incorporate local knowledge and influence changes in other agents (Polhill et al., 2019).

ABMs have been applied to a number of different applications and disciplines (Lee et al., 2015; Dawid and Gatti, 2018). However, they seem to be especially suitable for the analysis of complex systems and social sciences (Edmonds and Meyer, 2017), including a variety of modeling purposes (Edmonds et al., 2019).

Policy analysis specifically seem to be a perfect use of ABMs. First, public policies constitute complex systems with a number of ever-changing agents (citizens, politicians, businessmen, scientists, institutions) acting and reacting in time and space, with limited information, and unaware of others rationality (Furtado, Sakowski and Tóvolli, 2015; Geyer and Cairney, 2015; Arthur, 1994).

Second, ABMs do not need to work with a single – likely to fail (Mueller, 2020) – policy goal. Quite on the contrary, ABMs are fit to work as a communication tool (Gilbert et al., 2018) that makes the analysis systematical and visual, enabling parties from different backgrounds, experiences and disciplines to converse and exchange, whilst producing probabilistic scenarios.

Third, computational modeling in general and ABMs in particular are suitable for experimenting. Counterfactuals and *what-if* questions are formulated and tested within a safe environment, with costs restricted to the computational burden, and the time-consumed by modelers, programmers and policymakers.

There has been a plethora of models using ABMs in a variety of disciplines, methods and applications (Edmonds and Meyer, 2017; Lee et al., 2015; Heppenstall et al., 2012; Dawid and Gatti, 2018; Scott et al., 2016; Ingham-Dempster, Walker and Corfe, 2017; Furtado, 2022; van der Veen,

Kisjes and Nikolic, 2017). However, a much smaller number has been actually applied and used to aid and change real policy (Carley et al., 2006; Kerr et al., 2021; Gilbert et al., 2018). Mainly, the difficulty comes from the hurdle of bridging “the gap between policy practice, often expressed in qualitative and narrative terms, and the scientific realm of formal models” (Ahrweiler et al., 2015, p. 1). Nevertheless, ABMs have proved to be valuable policy tools (Gilbert et al., 2018), when observing its limitations (Aodha and Edmonds, 2017).

In this paper, we focus mainly on the expansion of the original model’s (Furtado, 2022) capacity to anticipate policy trajectories and thus enable policy decision-making. Hence, the analysis supports comparison and evaluation of alternative policies, or rather, alternative decisions on public money investing, before actually applying the policy.

## 2.1 ML and surrogate models

ML consists of a series of computer methods in different areas within mathematics and statistics that attempt to automate the processes of finding, sorting and weighting different patterns in data sets. Moreover, ML aims at finding “hidden rules” in order to predict and analyze data (Breiman, 2001). This is a relevant characteristic that differentiates ML from typical statistics or econometrics: organizing data from various sources with an emphasis on predicting and forecasting.

ML is typically divided between supervised and unsupervised learning. Supervised learning has an optimal, target or desired output for its models. Unsupervised learning has no target and identifying patterns or clustering is the main goal. In this paper, we apply a supervised learning process in which the target is an inter-city comparative duet of binary indicators (see section *Procedures*).

ML also include semi-supervised and reinforcement learning. In semi-supervised ML, some data contain labels – whereas the complementing data do not –, and it is mostly used for classification. In fact, information on the labeled data may be used to inform the unlabelled ones (van Engelen and Hoos, 2020). Reinforcement learning focuses on a reward function that guides the learning process of agents.

Random forest is a common method within supervised ML. A random forest consists of a selection of subsets that branch randomly into different trees. Those subsets are independent and non-identical decision trees defined by independent random vector parameters and input data. With a large number of trees, the algorithm then searches for the most popular class through equal-weight voting (Breiman, 2001; Ren, Cheng and Han, 2017; Mishina et al., 2015).

A neural network is another method of supervised ML in which several structures called neurons connect themselves from input to output through weights. Those weights (synaptic weights) display how connected, in a single-way, are two neurons. Neurons group themselves in relation to one another in terms of layers: there are the two visible layers (input and output)<sup>4</sup> as well as hidden layers, with just one or two layers being common (Olivas, 2010).

In this paper we tested ML methods to construct a random forest surrogate model that emulates the original simulation model, PS2 (Furtado, 2022). “A surrogate model is an approximation method that mimics the behavior of an expensive computation or process. (...) [a surrogate model] can be trained to represent the system, learning a mapping from simulator inputs to outputs”.<sup>5</sup> Moreover, surrogates are essentially emulators that substitute an actual simulation model for a simpler method in order to produce results through less demanding means, both in relation to time and to processing power. ML algorithms reiterate over an ABM: first the ML collects data from ABM observations, linking input data to outputs, then it reiterates over the available data in order to understand the patterns of the ABM itself.

A surrogate model has advantages that seem to be adequate for our exercise, as described in the following.

- 1) Generalization: a surrogate model enables many more combinations of input parameters. In practice, the surrogate model works as if we had run our original simulation model a much higher number of times, whilst testing a wider range of possible parameters.
- 2) Practicality and speed: running a surrogate 1,000,000 times is way faster (and actually feasible) than running 1,000 regular simulations. It is a different order of magnitude concept. The simulations are emulated, rather than actually run. This is especially valid for models with geographical information and a large number of agents.
- 3) Statistical extrapolation: surrogate models’ larger number of runs brings statistical relevance for the results and provides robustness for the extrapolation.

In that sense, the aim of our surrogate is to extract the most information of the model with as little computational time as possible. As such, we revise surrogate models in relation to the three main steps.

---

4. We analyze only many-to-one input-output relation. For more information on the developments of many-to-many, see Chambers and Yoder (2020).

5. Available at: <<http://bit.ly/3tWmlxy>>.

In the ML surrogate proposed by Lamperti et al. (2018), the authors explore the parameter space through the use of a non-parametric ML surrogate without prior knowledge of the spatial distribution of data. Starting with a sample of initial conditions values (through Sobol sampling), random subsets are drawn, ran through the ABM and then evaluated by user-defined criterion of positive or negative. Then the surrogate is learned through a ML algorithm using a variation of random forest. By repeating the steps, the surrogate predicts the probability of false negatives (non-optimal results that are actually optimal) and is developed focusing on producing less of false negatives. In a nutshell, with every run of the ABM, it processes the data, labels and evaluates the results.

The surrogate of Lamperti, Roventini, and Sani (2018) is directly comparable to surrogate models that use Bayesian emulation or kriging, such as the model of ten Broeke et al. (2021). According to the authors of ten Broeke et al. (2021), their kriging surrogate is built through a prior-distribution of hyper-parameters, then the ABM is ran and results are gathered by the use of an integration function of the generalized least squares estimator of positive-definite symmetric matrices. The modeler then must choose a simulator: a multi-output (MO) emulator or a many to single-output. The authors choose the MO emulator due to convenience and least computational time (ten Broeke et al. 2021). By using the MO emulator, one may use the produced surrogate to make model predictions. Their surrogate is quite similar to that of Conti and O'Hagan (2010), especially in relation to the discussion of multi-output emulators versus other options. Essentially, their surrogate separates data into training and testing classifications, expanding the training data until sufficient performance (decided ad hoc by the modeler) is reached.

The surrogate of Hayashi et al. (2016), focuses on the parameter space of the simulation. The authors use four ML methods, with neural networks and random forests being part of those tested. The authors analyze many inputs to one output, again separating between "positive" and "negative" outcomes. ML feeds back the ABM by revising the inner workings of agents with each round of the simulation restarting with initial conditions given by the ML process. Their models connects the ABM and ML portions by comparing their results, attempting to minimize the error between the outputs of the ABM and the ML surrogate.

The surrogate of Edali and Yücel (2019) relies heavily on random forest and focuses on output predictions. For the training and fitting of the surrogate, they use the latin hypercube sampling and then sequential sampling. It then combines random forests with uncertainty sampling in order to train the surrogate model.

The surrogates we reviewed interact with an ABM model or the modeler itself to increase the quality of the prediction or the performance of the original model. Our proposal is somewhat independent of the original ABM. Our surrogate uses input and output from the original ABM to

learn the inner mechanisms – the mapping – that transforms inputs into outputs. Once trained, a new analysis that takes advantage of a much larger parameter space (unfeasible for the case of the original ABM) is applied with a clear target of policy optimality. The results are then compared providing a counterfactual panorama for policymakers.

In the next section, we describe first the basics of the original PS2 model, then the data generated by PS2 and used as input to our surrogate model. After that, we detail the procedures and steps that we use in running the emulator.

### 3 METHODS

#### 3.1 Baseline ABM: PS2

PS2 is an open source, readily available.<sup>6</sup> ABM that focuses on three alternative policy-schemes for the case of Brazilian MRs. According to the authors, PS2 is defined “as a primarily endogenous computational agent-based model (ABM) that includes mortgage loans, housing construction, tax collection and investments, with firms and households interacting in real estate, goods and services, and labor markets” (Furtado 2022, p. 2). The default run of PS2 is for the MR of Brasilia from 2010 to 2020, using census and business data, along with detailed geographical intraurban (census tract equivalent) boundaries.

The main goal of PS2 is to contrast and compare three competing policy investments alternatives for the case of Brasilia. Given low-income households, municipalities may either i) purchase houses to transfer; ii) provide rental vouchers; or iii) make monetary transfers. These policies are tested against a no-policy baseline and results are compared using macroeconomic indicators and their trajectories. We claim to analyze policies beyond just housing policies since

providing housing alone may not be sufficient if the benefit does not include jobs and services and access to the city, as shown by the policy program implemented through 2009-2019. (...) However, if the municipality is not seeking a housing policy per se, but otherwise a general policy to invest its financial resources, a Monetary aid alternative might be comparatively better (Furtado, 2022, p. 23).

---

6. Available at: <<http://bit.ly/3TThRaJ>>.

PS2, however, is developed and evaluated for the case of Brasilia MR alone. Although the authors also run the model and present comparative results to four other MRs,<sup>7</sup> results were not exhaustively run for all of the 46 available MRs. This is exactly how this surrogate model contributes to the analysis. We use the bulk of runs used for the sensitivity analysis and expand the results to include all 46 MRs, while also expanding the variability contained in the data by better exploring the parameter space.

### 3.2 Data

In order to emulate the original PS2 ABM, we start out with two sets of data produced in the 11,076 runs the authors of PS2 executed (Furtado, 2022). On the one hand, we access parameters' configuration for each single run. On the other hand, we evaluate the outputs – in terms of indicators produced by the model – associated with each configuration of parameters (see the left ABM portion of figure 1).

The configuration of parameters and rules includes information such as which MR is being simulated, the percentage of the population for a given run, the choice of wage-decision the firms used, or which policy is tested, if any. The full list of parameters include: i) all of those in table 3; ii) the Boolean choices of table 4; and iii) each MR, as listed in table 2. Moreover, parameters include a test of tax distribution, interest rate alternatives, availability of construction lots, and simulation starting year.

Additionally, PS2 produces an array of 66 indicators, providing details that refer to all the municipalities of a MR, individual municipalities, but that also contains information on households, workers, businesses, and the bank. All in all, our data maps the exact configuration of parameters for each run to the set of indicators produced by that same run. We then proceed to the learning process that maps inputs to outputs and enables extrapolation using the surrogate ML model.

### 3.3 Concept and design of the optimum

Here comes the “towards policy” part of the paper. Instead of just learning patterns from data (unsupervised ML), we have opted to apply a supervised ML to construct a surrogate model focused on policy itself. A model that distinguishes whether tested policies are socially beneficial or not.

---

7. Belo Horizonte, Campinas, Fortaleza and Porto Alegre.

A large portion of public economic debate surpasses the growth of the economy, and adopts GDP as its main indicator. There has been some attempts to qualify the discussion and include other dimensions, such as a Human Development Index (HDI), a focus on the United Nations Sustainable Development Goals – UN-SDG (Guerrero and Castañeda, 2020; 2022), or even include happiness as a subjective indicator, although with limited reach (Austin, 2016).

Here we decided to focus simultaneously on GDP and Gini coefficient due to their broad dimensions. The algorithm we constructed enables the running of a surrogate with any choice of pair of indicators available in the original model, along with their threshold quantiles. However, we believe that social benefit can be well represented by a concept of relative largest production and smallest inequality as optimum. Hence, the binary classification of optimal and non-optimal encompasses all simulation runs – out of the total sample of 11,076 – simultaneously in the superior quartile (0.75) of GDP indicator and in the inferior quartile (0.25) of the Gini coefficient indicator.

The optimum in this case is designed as a comparative and relational statistic among the set of 46 MR. As such, we may say that optimal refers to a national bundle that includes larger and smaller MR with diverse historical attributes, economic prevalence, population composition, businesses development and accessibility.

The idea here is that the surrogate model enables a counterfactual optimum as an would-be ex-post analysis, before actually implementing any of the policies. Given a photo of the existing attributes of each MR, the surrogate acknowledges which policy – if any – would change the status quo for the better, given the national context.

### 3.4 Procedures

The step-by-step of the implementation of the surrogate is as follows (figure 1).

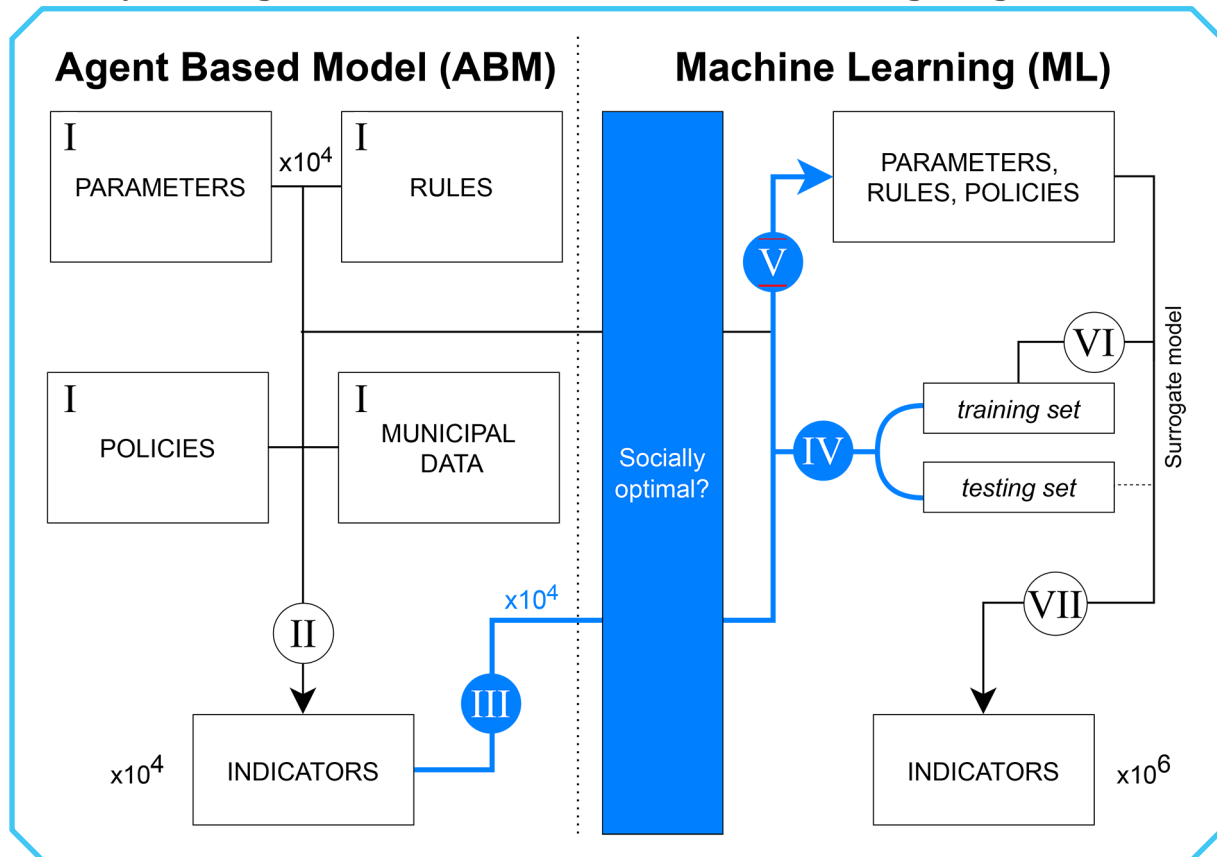
- 1) The PS2 ABM was manually run by the original authors, producing 11,076 individual runs. Each run contains the configuration parameters: i) including exogenous parameters; ii) rules and decision-making choices; iii) each one of tested policies – or the no-policy baseline; iv) the necessary initial municipal input data; and v) the output of the model containing all final indicators of the simulation. Examples of each one include:
  - a) exogenous parameters: workers' productivity, mortgage interests, businesses' markup, or the relevance of neighborhood quality perception in housing prices;



- b) rules and decision-making mechanisms: whether to consider distance to the firm as a job application criteria, how to set workers' wage, or how to distribute metropolitan funds;
  - c) policies: each one of the three tested policy, plus the no-policy baseline; and
  - d) municipal data: local information on households composition, workers qualification, businesses, and the intraurban spatial configuration. Municipal data is constant. However, one of the parameters determines the initial year, and thus the original data, to be either 2010 (default), or 2000.
- 2) We then map the input configuration parameters to the output indicators, constructing the typical ML X matrix. Hence, every configuration of parameters is linked to the output of the simulation depicted by the indicators.
  - 3) Next, we construct the optimal and non-optimal from output indicators, thus producing the typical ML Y vector.
  - 4) We proceed to the separation of original sample between training and test sets  $s$  using the typical python scikit-learn (skelarn) model selection procedure of train-test split, with test size of 0.25 and random state of 10.0. Then, we trained the model and tested it.
  - 5) Afterwards, we generate 1,000,000 new parameters, expanding the original configuration parameters into one million combinations (see subsection *Generation of parameters*).
  - 6) We apply the learning algorithm: ML random forest (see subsection *ML implementation*).
  - 7) Then, we apply the trained ML algorithm to the newly generated parameters to find optimal or non-optimal results.
  - 8) Finally, we analyzed the results.

**FIGURE 1**

Simplified diagram of the connection between ML and ABM regarding the PS2 model



Authors' elaboration.

Obs.: 1. The figure depicts the steps taken to emulate the original ABM, whereas amplifying the possible combination of parameters and steering the results towards optimal policy. The original ABM (left side) produced sample configuration parameters and output indicators (I and II). Indicators were then divided between optimal and non-optimal depending on their relative GDP and inequality indicators (III). A set of 1,000,000 new parameters were generated and trained in the ML surrogate model, producing one million results (IV-VIII).

2. Figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

In a nutshell, the ABM and the ML surrogate model are separated in their inner-workings. First, the ABM itself runs and produces outputs from configuration parameters, with some variation. Secondly, the output indicators enable the optimal classification. Finally, a much larger configuration of parameters uses the trained ML model – the surrogate – to generate new outputs.

We implemented and tested some ML implementations, such as support-vector machine (SVC), multi-layer perceptron classifier (MLPC) and a voting classifier. The SVC and MLPC were unable to predict optimal cases. Their optimal predictions were observed to be in fact non-optimal. These two unadjusted results influenced the Voting Classifier, which is a weighted score of the models tested.

### 3.4.1 Generation of parameters

In this subsection we elaborate on how the surrogate model generates parameters from its sample sets (derived from the ABM). Basically, parameters are generated according to two processes, as follows.

- 1) If they are discrete parameters or rules, they are chosen according to a probability that is simply the inverse of the number of alternatives ( $m^{-1}$ ), being  $m$  the number of possibilities for each dummy analyzed. In the case of policies, for example, each one of the three policies tested, along with the no-policy baseline, had a 0.25% probability.
- 2) If they are continuous parameters, they are drawn from a truncated normal distribution, given the lower and upper bounds (table A.3), the mean, and a standard deviation three times larger the original sample.

### 3.4.2 ML implementation

We implement and test some ML implementations, such as SVC, MLPC and a voting classifier. Nevertheless, the best results are achieved with the Random Forest Classifier method. We used trees, Gini criterion, with a maximum depth of 15 for each tree.

The confusion matrix is presented at table 1. The results obtained led to an accuracy of 0.9917 with a precision of 0.9863. However, the recall was not so good at 0.8727 and an F1 indicator of 0.9260. Considering that these results were obtained after 11,076 time-consuming simulation runs, we understand that the surrogate is adequate for the exercise proposed.

**TABLE 1**  
Confusion matrix for the random forest ML implementation

		Predicted	
		Non-optimal	Optimal
Observed	Non-optimal	2602	2
	Optimal	21	144

Authors' elaboration.

Obs.: The table shows that only 2 cases were predicted optimal when in fact they were not-optimal. Conversely, 21 cases that were optimal, were predicted as non-optimal.

## 4 RESULTS

In this section, we first present the results of each competition policy across the cities in decreasing order. Then, we detail the difference in parameters and rules given by the optimal and non-optimal classification.

### 4.1 Expansion of PS2 model results'

Table 2 depicts how the surrogate model is able to expand the number of results while reducing the difference in simulation runs between MRs. On average, each MR has 20.000 runs on the surrogate (about 5.000 for each policy in force and 5.000 for the no-policy baseline). As such we expand all other MRs to the same level of experiments that Brasilia had on the original PS2 ABM. To be more specific, there was about 5.000 runs performed for Brasilia, whereas most other MR runs' ranged from 33 (Joinville) to 201 (Rio de Janeiro).

**TABLE 2**

**Results from surrogate and ABMs in relation to sample size (in number of simulations), and both optimal and non-optimal cases (in number of simulations that fall under each category) for MRs of the model**

MR	ABM		Surrogate	
	Optimal	Non-optimal	Optimal	Non-optimal
Belo Horizonte	180	42	16698	4970
Fortaleza	7	215	3891	17786
Porto Alegre	0	222	2313	19664
Campinas	73	149	8943	12866
Salvador	13	104	3960	17711
Recife	0	117	4101	17576
São Paulo	117	0	6814	15098
Joinville	0	33	4326	17318
Campo Grande	0	117	4208	17322
Jundiaí	0	117	4099	17706
Feira de Santana	0	117	4089	17560
Ipatinga	0	117	4187	17560
Londrina	0	117	4195	17611
Sorocaba	0	117	4181	17528
João Pessoa	0	117	4249	17515

(Continues)

(Continued)

MR	ABM		Surrogate	
	Optimal	Non-optimal	Optimal	Non-optimal
SJRP <sup>1</sup>	0	117	3898	17788
Maceió	0	117	4247	17351
SJC <sup>2</sup>	0	117	4123	17631
Ilhéus-Itabuna	0	117	4191	17641
São Luis	0	117	4023	17690
Uberlândia	0	117	4235	17560
Maringá	0	117	4295	17286
Vitória	0	117	4227	17785
Cuiabá	0	117	4132	17501
Belém	25	92	5782	16045
NH-SL <sup>3</sup>	0	117	4105	17527
Teresina	0	117	4123	17791
Manaus	0	117	4217	17500
Brasilia	57	5334	0	21898
Aracaju	0	117	4016	18093
Campina Grande	0	117	4173	17462
Campos dos Goytacazes	0	117	3991	17692
Caxias do Sul	0	117	4122	17640
Crajuubar	0	117	4161	17659
Curitiba	56	61	10112	11407
Florianópolis	0	117	4188	17650
Goiânia	0	117	4058	17573
Juiz de Fora	0	117	4052	17520
Macapá	0	117	4098	17675
Natal	0	117	4057	17831
Pelotas-Rio Grande	0	117	4273	17298
Petrolina-Juazeiro	0	117	4179	17573
Ribeirão Preto	0	117	4051	17582
Rio de Janeiro	110	91	6116	15763
Santos	0	117	3907	17792
Volta Redonda	0	117	4119	17480

Authors' elaboration.

Notes: <sup>1</sup> São José do Rio Preto.

<sup>2</sup> São José dos Campos.

<sup>3</sup> Novo Hamburgo/São Leopoldo.

## 4.2 MRs results per policy

Figure 2 shows the performance of all MRs for each policy in terms of percentage of optimal runs, given the optimum definition of best quartile of production and lowest quartile of inequality.<sup>8</sup> On average (table A.1), in 17.87% of the simulations performed, the no-policy baseline was the optimal scenario. Three MRs performed relatively better than the others (Belo Horizonte, São Paulo and Rio de Janeiro), whereas four others performed worse (Belém, Porto Alegre, Campinas and Brasília). All of the remaining 39 were classified as optimal between 19% and 12% of their own runs.

Comparatively and in tune with the results found by (Furtado, 2022), the no-policy baseline and the Purchase policy perform worse than the other alternatives for all cases, although with varying intensity. Simulations in which Purchase policy is in effect fare consistently worse than simulations with no policy at all. To be more specific, a simulation with Purchase policy in force fares 14.25 percentage points (p.p.) worse than its baseline, on average (table A.1). On the other hand, for three cities (Rio de Janeiro, São Paulo and Campinas) the loss is small, around one p.p.

Conversely, both – very different – policy alternatives of Rent vouchers or Monetary aid perform better than the no-policy baseline (figure 2). On average, Monetary aid is slightly better, surpassing the no-policy baseline by 15.05 p.p. Rent vouchers in turn provide a gain of 13.16 p.p. Across the MRs, however, the performance varies with Rent vouchers being the best option for 24 MRs, and Monetary aid for other 19 MRs, with one tie.

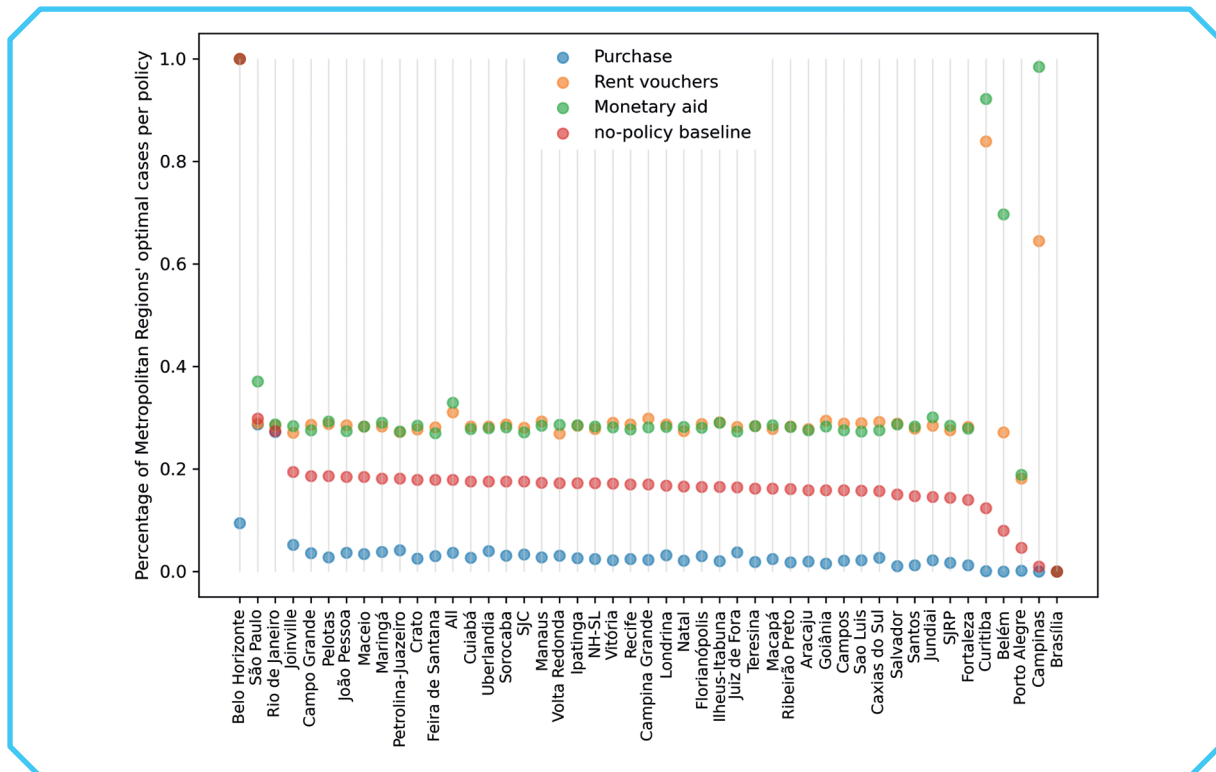
---

8. All of the MRs rejected the null hypothesis of the Welch's t-test of equal means between optimal and non-optimal results.

**FIGURE 2**

Optimal cases for the surrogate results, each of the tested policies and the no-policy baseline

(In %)



Authors' elaboration.

Obs.: 1. The no-policy baseline (in red) clearly separates the mostly negative policy of Purchase (blue) and the alternating better policies of Rent vouchers (orange) and Monetary aid (green).

2. Figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

There are some special cases. Given the design and conception of the optimal and their comparative nature across MRs, two cases are extreme. Brasilia – the federal capital and host to a number of well-paid civil servants – has the highest structural inequality of all MRs. Hence, Brasilia does not have any policy as optimal, although it accounts for 2.19% of the 1,000,000 surrogate runs (table 2). In the manual ABM run, in which Brasilia accounts for 48.67% of the runs, it does reach an optimal case in 1.06% of the time. Our interpretation is that Brasilia should evaluate alternative housing social policies that might together tackle the city's status quo more effectively.

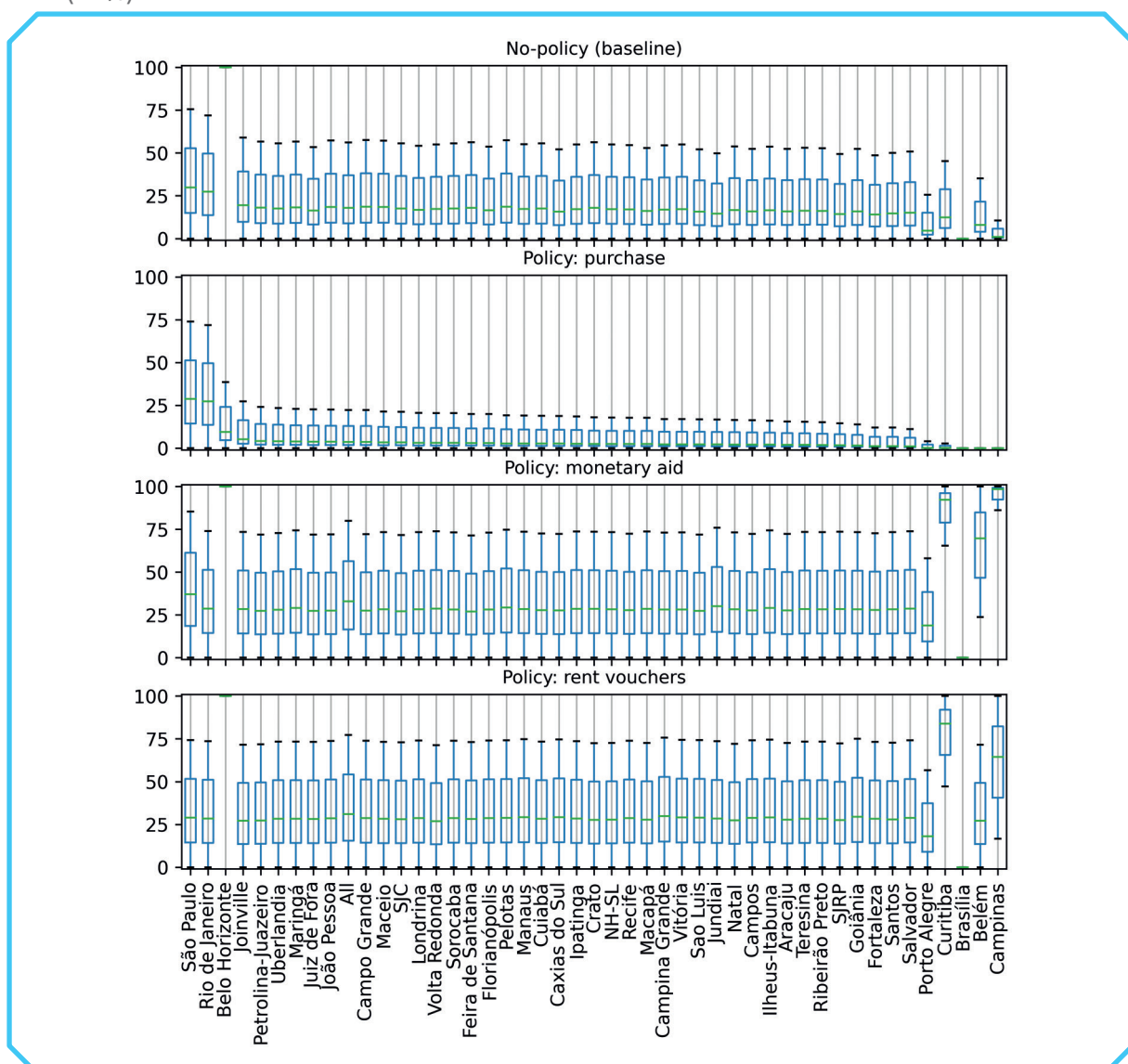
Conversely, Belo Horizonte seems to benefit from a relatively large GDP, without a largely unequal population, thus it reaches the optimal cases in all of the no-policy, rent vouchers and monetary aid surrogate runs. Specifically for the case of Belo Horizonte, the Purchase policy is especially harmful, reducing the optimal in 90.6 p.p.

## DISCUSSION PAPER

Moreover, Curitiba would gain nearly 80 p.p. with monetary aid policy and 71.6 with Rent vouchers. Belém would improve 61.7 with Monetary aid and Campinas would gain more than 95.7 points also with the monetary aid policy (table A.1).

**FIGURE 3**

**One standard deviation above and below mean results per MR per policy**  
(In %)



Authors' elaboration.

- Obs.: 1. Results show that simulations with Purchase policy in force tend to have consistently worse results than all other scenarios, including the no-policy scenario. On the other hand, Monetary aid and Rent voucher policies have a wider breadth of results, i.e., the chance of reaching optimum is present but not certain.
2. Figure whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).



Figure 3 showcases the differences between mean results per MR per policy. It is a box-plot graph depicting one standard deviation above and one below the mean number of simulations that are optimal for each MR under each different housing policy. Since results are binary (zero or one), the values on figure 3 mean that the percentage of simulation runs that are optimal under each criteria (MR and policy). The entry "All" encompass the mean results per policy, i.e., the mean number of simulations that are optimal under a specific housing policy. Focusing on the entry "All", i.e. regardless of the specific MR analyzed, by comparing each policy, it is clear that the standard deviation of the baseline case is higher than the standard deviation of the Purchase policy case, and lower than the measurements for Rent vouchers and monetary aid cases. As such, the smaller the standard deviation, the closer the results are to the mean, i.e., it reflects a probability to get optimal results. Consequently, since Monetary aid consistently provides less optimal results than other options and it has the smaller standard deviation, it is more likely for a simulation to reach a non-optimal situation under Monetary aid than under any other policy option, including the no-policy baseline scenario. Nevertheless, since the better policy options (Rent vouchers and Monetary aid) also possess higher standard deviations: there is more chance for simulations using those housing policies to reach optimal, but there is still chance to reach non-optimal results. In a nutshell, it is easier and more certain to fail with a Purchase policy than to succeed with either of the other ones.

In policy terms, a purchase policy, for most MRs with the exception of São Paulo and Rio de Janeiro, will almost always lead to non-optimal results. Nevertheless, a Purchase or Monetary aid policy may not be as certain in terms of leading that MR into an optimal state, due to the higher standard deviations for those two better policy options. As such, it becomes clear that there are more factors that come into play when determining the outcome of a policy than only the policy itself. However, policymakers may rest assured that the Purchase policy remains as a distant fourth-best option, unsuited for most MRs and consistently worse than even the no-policy baseline scenario.

Another problem with Purchase policy is that it benefits fewer people than other policies. Purchase policy appears as a costly policy that has a great impact on a smaller group of beneficiaries. In that sense, it also exposes the clash between individual and collective rationality that exists in many policies (Al-Suwailem, 2014; Chen, Tai and Chie, 2002): for the beneficiary it is better to get the house, whereas for the population as a whole, other policies may be more well suited for achieving optimum.<sup>9</sup> Carvalho (1997) also discusses such dichotomy from the point-of-view of macroeconomic policies.

### 4.3 Parameters and rules analysis

The ABM model was calibrated and validated towards a reasonable performance along four macroeconomic indicators and the spatial distribution of the real estate market for the city of Brasilia

9. For a theoretical discussion of individual versus collective rationality, we recommend Denis (2002).

in the no-policy baseline scenario (Furtado, 2022). The ABM optimal results were calculated after its completion, as an ex-post indicator. Conversely, the ML surrogate was designed to separate optimal from non-optimal policy results. Moreover, the running of the ML surrogate model enabled a much larger combination of possible parameters, rules, policies and MRs (see generation of parameters).

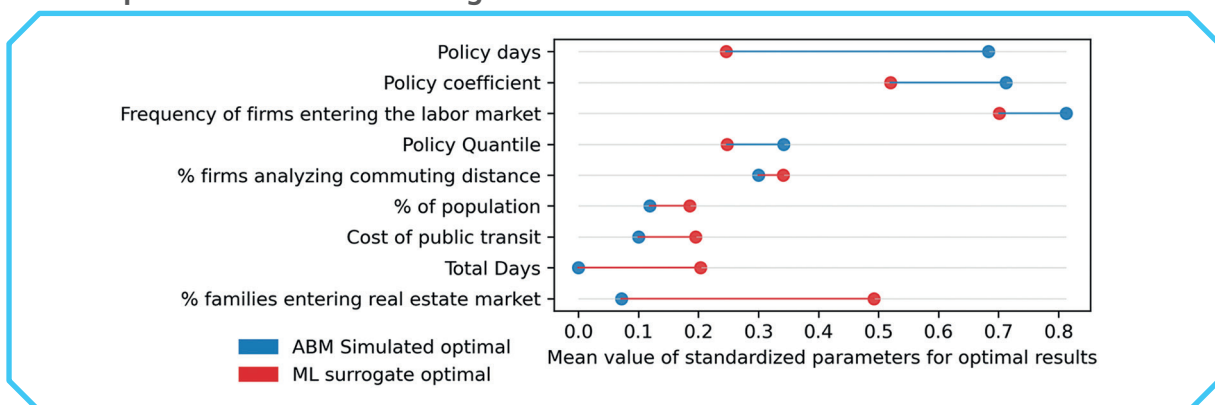
We standardize values of parameters regarding optimal runs from the ABM and ML surrogate models. Doing so, we may compare which parameters and rules (when possible) would need to be modified were policymakers to nudge their MRs towards better results.

There are higher, similar and smaller values of the standard score for the ABM when compared to the ML surrogate optimal (figure 4). Our interpretation is that the values of the ML surrogate – considering the much wider uncertainty included and the number of runs executed – would be the ideal ones when considering policies towards the optimal. Hence, when possible, policymakers should work towards changing actual parameters into best parameters.

We could also not reject the null hypothesis that the means were the same for some of the parameters. That indicates that either those parameters were already at an optimal level at the ABM calibration, or that they are not much relevant to generate optimal results.

#### FIGURE 4

##### Comparative standard score of parameters for optimal results in relation to the full sample for ABM and ML surrogate runs



Authors' elaboration.

Obs.: 1. We included the parameters for which we could reject the null hypothesis that the means were statistically the same between ABM and ML and also between optimal and non-optimal ML surrogate. The difference from the validated ABM simulation run and the ML surrogate optimal case suggests the counterfactual changes needed in the parameters in order to increase the number of optimal results. The figure is sorted by the difference between ABM Simulated model and ML surrogate optimal.

2. Figure displayed in low resolution and whose layout and texts could not be formatted due to the technical characteristics of the original files (Publisher's note).

The three parameters that refer to the design of the families' sample size that are to receive policy aid should have smaller values according to the ML surrogate optimal results. Policy days refers to the number of days the municipalities look back in time to check the financial conditions of the families and decide to include or not in the beneficiary pool. The default value is one year (360 days), but the ABM simulation also tested it for 180 days. ML results suggest then that using a smaller number of backward months in order to include households as policy beneficiaries would produce better overall performance.

Policy coefficient specifies the municipalities' budget share to be invested in the policy. Policy quantile is the income threshold for families to be included in the policy program. Whereas lowering policy coefficient would diminish policy investments, a smaller threshold to include households would probably help focus the policy towards the poorest, most in-need families.

Together, these parameter optimal values suggest that the choice of families to be included could be more precise. Conversely, note that the purchase policy was by design a policy that helps a comparatively much smaller number of families, given that the amount of financial resources to purchase houses is considerably higher than that of renting (Furtado, 2022). However, in the case of purchase policy, the number of aided households is too small, which is also not the best solution.

The frequency of firms entering the labor market value is slightly lower in the ML surrogate when compared to the ABM model. This suggests that the optimal is reached more frequently when the labor market is less volatile.

The parameter of percentage of firms that take commute distance as a selection criteria should be slightly higher to enable more optimal results, according to the ML surrogate model results. This parameter refers to the spatial arrangement of the MRs and varies between hiring exclusively via best candidate qualification or via residence proximity to the hiring firm. As discussed in (Furtado, 2022), there is a midway arrangement between qualification and commuting distance that best distributes jobs and workers bringing more prosperity for the MR as a whole.

Percentage of population refers to the size of the sample of simulated inhabitants. Higher ML surrogate optimal values depict the gain that moderately larger MRs have when compared to smaller ones.<sup>10</sup> Total days also indicate that longer simulations may generate more positive feedback effects and help MRs.

---

10. In a similar process to what causes the success of Belo Horizonte in most simulations.

The percentage of families that engage into the real estate market is much larger ML surrogate optimal cases, compared to that of ABM surrogate optimal cases. As expected, a more dynamic real estate market indeed brings economic gains to the MRs. It is relevant to highlight, however, that whereas the ABM was calibrated considering reasonable inflation, along with other macroeconomic indicators, the ML surrogate is concerned with production output and inequality alone. Yet, the parameter value suggests that dynamic real estate markets may bring positive overall results.

**TABLE 3**

**Comparison of parameters' absolute values for ABM optimal and ML surrogate optimal results**

Parameters	Standardized values	
	ABM optimal	ML surrogate optimal
% of firms analyze commute distance	0.300	0.341
% of construction firms	0.200	0.446
% of population	0.118	0.185
% that enters the estate market	0.071	0.492
Cost of private transit	0.500	0.498
Cost of public transit	0.100	0.195
Frequency of firms entering the labor market	0.812	0.701
Hiring sample size	0.655	0.510
Loan/permanent income ratio	0.625	0.528
Markup	0.500	0.505
Maximum Loan-to-Value	0.625	0.532
Municipal efficiency management	0.474	0.444
Neighborhood effect	0.571	0.579
Perceived market size	0.474	0.476
Policy quantile	0.342	0.248
Policy coefficient	0.712	0.520
Policy daus	0.683	0.246
Productivity: divisor	0.314	0.566
Productivity: exponent	0.667	0.579
Sticky Prices	0.500	0.489
Supply-demand effect on real estate prices	0.500	0.483
Tax over estate transactions	0.667	0.514
Total days	0.000	0.204

Authors' elaboration.

Finally, ML surrogate optimal cases suggest that cost of public transit should be a bit larger, when compared to the ABM. As the cost of public transit is considered by the candidate as a criteria when deciding the firm to work with, the results may indicate that a better spatial match between workers and firms – mediated via considering more heavily commuting costs – might bring more positive results for the full set of the MR.

**TABLE 4**

**Differences between surrogate and ABMs in relation to sample size,<sup>1</sup> and both optimal and non-optimal cases<sup>2</sup> for selected dummy parameters of the model**

(In %)

Rules and policies' choices	Size		Optimal		Non-optimal	
	Surrogate	ABM	Surrogate	ABM	Surrogate	ABM
Policy: Purchase	25.06	12.19	3.62	2.59	96.38	97.41
No-policy baseline	24.93	63.43	17.87	3.91	82.13	96.09
Policy: Rent vouchers	25.01	12.19	31.03	10.37	68.97	89.63
Policy: Monetary aid	25.00	12.19	32.92	13.93	67.08	86.07

Authors' elaboration.

Notes: <sup>1</sup> In % of total simulations.

<sup>2</sup> In % of cases out of the total that fall under each category.

## 5 DISCUSSION

In practical terms, the exercise proposed generalizes earlier results and pinpoints which policy is best for which city. Moreover, the robustness of the results confirm that the Property acquisition policy performs worse in terms of GDP and Gini coefficient within a varied input of MRs. Rent vouchers and Monetary aid provide similar results, but may be different for specific MRs.

The federal government (and policymakers) may learn from the analysis performed. The results provide quantitative and empirical data that support alternative emphasis on policy instruments. Typically, Brazilians and Organization for Economic Cooperation and Development (OECD) members prefer to own their houses (Causa, Woloszko and Leite, 2019). Indeed the Property acquisition program in Brazil was called My House, My Life (PMCMV). However, the program has been criticized by the lack of city integration (distant locations and without infrastructure) – Amore, Shimbo and Rufino (2015) –, and has virtually stopped due to difficulties in funding. Nationally, no other housing policy has replaced PMCMV. A few studies have recommended rent vouchers as alternatives (Dias and Santos, 2021; Rogar and Vieira, 2018). Monetary aid, in turn, has steadily

gained support in Brazil (Fonseca et al., 2018). Our robust results support the novel comparison among housing and social welfare policy instruments making it explicit for policymakers how they compare and what benefits are reached, considering GDP and Gini coefficient as policy goals.

The findings of the surrogate contributes to policy analysis literature by showing that other alternatives (to the Property acquisition policy) not only are available, but also might be more beneficial. PMCMV – the large Property acquisition program in Brazil – was decided upon without observing the planning legislation that was in place, nor the opinion of urban policy experts (Ferreira et al., 2019).

The findings of the comparison among the ABM and the ML surrogate optimal results make sense and quantitatively inform policymakers about the direction and intensity of necessary adjustments to replicate optimal MRs performance. All in all, the comparisons suggest: i) a more focused set of low-income definition of households to assist; ii) with a smaller number of months (of income observation) to include the households; iii) a less volatile (safer) labor market; iv) with a consideration of transport costs to improve the quality of spatial match between firms and employees; and v) and that larger metropolises fair better with increased gains from agglomeration.

## 6 CONCLUDING REMARKS

We present a ML surrogate model that builds from an ABM model, encompassing a wider number of MRs and policy tests. In the process, we include a much larger, more varied sample of possible input parameters and rules. We also steer the ML surrogate model towards a policy optimum that combines larger production output and smaller inequality within the set of all 46 modeled MRs. Besides the actual findings of best-policy for each MR, the exercise also serves as a test of robustness of the original results.

Considering the research questions, we find that, indeed, different MRs do best in different tested policies. Although on average, the Monetary aid policy performs better, the Rent vouchers is the best policy for a higher number of MRs (24). We further confirm results from Furtado (2022) and achieve worse results for the Purchase policy, vis-à-vis those of the no-policy baseline.

Results have also shown that some parameters, for simulations that are optimal, have statistically significant results in the ML surrogate model than in the ABM model. We interpret these values as the values the parameters would have to be transformed into, were the MRs to aspire to more socially optimal positions.

The paper presents an exercise to expand the combination of parameters, test the robustness of the results, and evaluate possible pathways for policymakers in terms of which parameters would need changing, were they interested in reaching higher GDP with lower Gini coefficient. However, results are limited by the available data used to produce the surrogate model and the intrinsic validated of the original model itself. Conceptually, we show that a reasonably fast policy evaluation of ABM results – “ex post” – is possible, and one that adds information to policymakers.

In terms of future research, we are interested to learn whether intrinsic MR' attributes and structural characteristics translates into a higher probability of getting comparatively optimal results. That is, are household attributes and location, number of firms, population qualification and age composition, municipalities boundaries and their geographical interconnections sufficient to determine relative position of one MR in relation to the others?

In order to answer such question, we would need to search for the determinants of the optimum, probably via regression techniques.

Moreover, we envision a study that uses dimensionality reduction techniques to summarize a larger number of output indicators to better reflect policymakers preferences. Along with a more comprehensive output indicator, future work should take advantage of the possibilities of mixing simultaneous policy instruments, and those from different domains, such as housing policy and social welfare.

## REFERENCES

AHRWEILER, P. et al. Modelling research policy: ex-ante evaluation of complex policy instruments. **Journal of Artificial Societies and Social Simulation**, v. 18, n. 4, p. 1-5, 2015.

AL-SUWAILEM, S. Complexity and endogenous instability. **Research in International Business and Finance**, v. 30, p. 393-410, Jan. 2014. Retrieved from: <<http://bit.ly/3UXBVBX>>.

AMORE, C. S.; SHIMBO, L. Z.; RUFINO, M. B. C. (Org.). **Minha casa... e a cidade?** Avaliação do Programa Minha Casa Minha Vida em seis estados brasileiros. Rio de Janeiro: Letra Capital, 2015.

AODHA, L. N.; EDMONDS, B. Some pitfalls to beware when applying models to issues of policy relevance. In: EDMONDS, B.; MEYER, R. (Ed.). **Simulating social complexity: a handbook**. 2nd ed. Cham: Springer International Publishing, 2017. p. 801-822. (Understanding Complex Systems Series).

ARTHUR, W. B. Out-of-equilibrium economics and agent-based modeling. In: TESFATSION, L.; JUDD, K. L. (Ed.). **Handbook of computational economics**. Amsterdam: North-Holland, 2006. v. 2, p. 1551-1564. Retrieved from: <<http://bit.ly/3i3Pr0K>>.

\_\_\_\_\_. Inductive reasoning and bounded rationality. **The American Economic Review**, v. 84, n. 2, p. 406-411, May 1994.

AUSTIN, A. On well-being and public policy: are we capable of questioning the hegemony of happiness? **Social Indicators Research**, v. 127, n. 1, p. 123-138, 2016. Retrieved from: <<http://bit.ly/3tUbl9b>>.

AXELROD, R. M. (Ed.). **The complexity of cooperation**: agent-based models of competition and collaboration – Princeton studies in complexity. Princeton: Princeton University Press, 1997.

BATTY, M. Defining urban science. In: SHI, W. et al. (Ed.). **Urban informatics**. Singapore: Springer, 2021. p. 15-28. (The Urban Book Series). Retrieved from: <<http://bit.ly/3V7c5DV>>.

BOSCHMA, R. **Constructing regional advantage and smart specialization**: comparison of two European policy concepts. Utrecht: Utrecht University, 2013. (Papers in Evolutionary Economic Geography, n. 13.22). Retrieved from: <<http://bit.ly/3V7cMNx>>.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001.

BRENNER, T.; WERKER, C. Policy advice derived from simulation models. **Journal of Artificial Societies and Social Simulation**, v. 12, n. 4, p. 2, 2009.

CARLEY, K. M. et al. BioWar: scalable agent-based model of bioattacks. **IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans**, v. 36, n. 2, p. 252-265, 2006. Retrieved from: <<http://bit.ly/3GJNhxP>>.

CARVALHO, F. J. C. de. Economic policies for monetary economies: Keynes's economic policy proposals for an unemployment-free economy. **Revista de Economia Política**, v. 17, n. 4, p. 31-51, Oct.-Dec. 1997.

CAUSA, O.; WOLOSZKO, N.; LEITE, D. **Housing, wealth accumulation and wealth distribution**: evidence and stylized facts. Paris: OECD Publishing, Dec. 2019. (Economics Department Working Papers, n. 1588).

CHAMBERS, R. D.; YODER, N. C. FilterNet: a many-to-many deep learning architecture for time series classification. **Sensors**, v. 20, n. 9, p. 2498, 2020. Retrieved from: <<http://bit.ly/3ET8VOv>>.

CHEN, S.-H.; LUX, T.; MARCHESI, M. Testing for non-linear structure in an artificial financial market. **Journal of Economic Behavior & Organization**, v. 46, n. 3, p. 327-342, Nov. 2001.



CHEN, S.-H.; TAI, C.-C.; CHIE, B.-T. Individual rationality as a partial impediment to market efficiency. In: CHEN, S.-H. (Ed.). **Genetic algorithms and genetic programming in computational finance**. Boston: Springer US, 2002. p. 357-377. Retrieved from: <<http://bit.ly/3VeO2CU>>.

CONTI, S.; O'HAGAN, A. Bayesian Emulation of complex multi-output and dynamic computer models. **Journal of Statistical Planning and Inference**, v. 140, n. 3, p. 640-651, Mar. 2010. Retrieved from: <<http://bit.ly/3B2xczt>>.

CRAVO, T. A.; RESENDE, G. M. Economic growth in Brazil: a spatial filtering approach. **The Annals of Regional Science**, v. 50, n. 2, p. 555-575, 2013. Retrieved from: <<http://bit.ly/3GHClci>>.

DAWID, H.; GATTI, D. D. Agent-based macroeconomics. In: HOMMES, C.; LEBARON, B. (Ed.). **Handbook on computational economics**. Amsterdam: Elsevier, 2018. v. 4, p. 63-156.

DENIS, A. Collective and individual rationality: Maynard Keynes' methodological standpoint and policy prescription. In: ZAREMBKA, P. (Ed.). **Confronting 9-11, ideologies of race, and eminent economists**. Bingley: Emerald (MCB UP), 2002. p. 187-215. (Research in Political Economy, v. 20). Retrieved from: <<http://bit.ly/3V5ymC4>>.

DIAS, D. M. dos S.; SANTOS, J. C. dos. Um estudo acerca do aluguel social como instrumento de acesso ao direito à moradia digna diante do déficit de habitação no Brasil. **Revista de Direito da Cidade**, v. 13, n. 3, p. 1631-1659, 2021.

EDALI, M.; YÜCEL, G. Exploring the behavior space of agent-based simulation models using random forest metamodels and sequential sampling. **Simulation Modelling Practice and Theory**, v. 92, p. 62-81, Apr. 2019. Retrieved from: <<http://bit.ly/3hW0VTJ>>.

EDMONDS, B. et al. Different modelling purposes. **Journal of Artificial Societies and Social Simulation**, v. 22, n. 3, June 2019. Retrieved from: <<http://bit.ly/3EyCPpP>>.

EDMONDS, B.; MEYER, R. (Ed.). **Simulating social complexity: a handbook**. 2nd ed. Amsterdam: Springer International Publishing, 2017. (Understanding Complex Systems Series). Retrieved from: <<http://bit.ly/3XptTf9>>.

EHRENTREICH, N. The original Santa Fe institute artificial stock market. In: EHRENTREICH, N. (Ed.). **Agent-based modeling: the Santa Fe institute artificial stock market model revisited**. Berlin: Springer, 2008. p. 91-112.

EPSTEIN, J. M.; AXTELL, R. **Growing artificial societies: social science from the bottom up**. Cambridge: MIT Press, 1996.

FABER, A.; VALENTE, M.; JANSSEN, P. Exploring domestic micro-cogeneration in the Netherlands: an agent-based demand model for technology diffusion. **Energy Policy**, v. 38, n. 6, p. 2763-2775, June 2010. Retrieved from: <<http://bit.ly/3TYdo6J>>.

FURTADO, B. A. PolicySpace2: modeling markets and endogenous public policies. **Journal of Artificial Societies and Social Simulation**, v. 25, n. 1, p. 1-30, 2022. Retrieved from: <<http://bit.ly/3XCh9IB>>.

FURTADO, B. A.; SAKOWSKI, P. A. M.; TÓVOLI, M. H. (Ed.). **Modeling complex systems for public policies**. Brasilia: Ipea, 2015.

GALÁN, J. M. et al. Errors and artefacts in agent-based modelling. **Journal of Artificial Societies and Social Simulation**, v. 12, n. 1, 2009.

GAWEL, E.; STRUNZ, S.; LEHMANN, P. **Support policies for renewables instrument choice and instrument change from a public choice perspective**. Helsinki: UNU-WIDER, 2016. Retrieved from: <<http://bit.ly/3V0wRoB>>.

GEYER, R.; CAIRNEY, P. (Ed.). **Handbook on complexity and public policy**. Cheltenham: Edward Elgar Publishing, 2015. (Handbooks of Research on Public Policy Series).

GILBERT, N. et al. Computational modelling of public policy: reflections on practice. **Journal of Artificial Societies and Social Simulation**, v. 21, n. 1, p. 14, 2018.

GUERRERO, O. A.; CASTAÑEDA, G. Policy priority inference: a computational framework to analyze the allocation of resources for the sustainable development goals. **Data & Policy**, v. 2, p. 1-31, 2020. Retrieved from: <<http://bit.ly/3tSnZ8I>>.

\_\_\_\_\_. How does government expenditure impact sustainable development? Studying the multidimensional link between budgets and development gaps. **Sustainability Science**, v. 17, n. 3, p. 987-1007, Feb. 2022. Retrieved from: <<http://bit.ly/3Ey30gj>>.

HAYASHI, S. et al. Improving behavior prediction accuracy by using machine learning for agent-based simulation. In: NGUYEN, N. T. et al. (Ed.). **Intelligent information and database systems**. Berlin: Springer, 2016. p. 280-289. Retrieved from: <<http://bit.ly/3tW0JGW>>.

HEPPENSTALL, A. J. et al. (Ed.). **Agent-based models of geographical systems**. Dordrecht: Springer, 2012.

HOCHSTETLER, K.; KOSTKA, G. Wind and solar power in Brazil and China: interests, state-business relations, and policy outcomes. **Global Environmental Politics**, v. 15, n. 3, p. 74-94, 2015. Retrieved from: <<http://bit.ly/3VktsRw>>.

HUANG, W.; ZHENG, H.; CHIA, W.-M. Financial crises and interacting heterogeneous agents. **Journal of Economic Dynamics and Control**, v. 34, n. 6, p. 1105-1122, June 2010. Retrieved from: <<http://bit.ly/3GDOTJ9>>.

INGHAM-DEMPSTER, T.; WALKER, D. C.; CORFE, B. M. An agent-based model of anoinis in the colon crypt displays novel emergent behaviour consistent with biological observations. **Royal Society Open Science**, v. 4, n. 4, p. 1-13, Apr. 2017. Retrieved from: <<http://bit.ly/3EUNjBo>>.

JANSSEN, M.; OSTROM, E. Empirically based, agent-based models. **Ecology and Society**, v. 11, n. 2, p. 1-13, 2006.

KERR, C. C. et al. Covasim: an agent-based model of covid-19 dynamics and interventions. **PLOS Computational Biology**, v. 17, n. 7, p. 1-32, July 2021. Retrieved from: <<http://bit.ly/3OtpcWN>>.

LAMPERTI, F. et al. Faraway, so close: coupled climate and economic dynamics in an agent-based integrated assessment model. **Ecological Economics**, v. 150, p. 315-339, Aug. 2018. Retrieved from: <<http://bit.ly/3EFuxwC>>.

LAMPERTI, F.; ROVENTINI, A.; SANI, A. Agent-based model calibration using machine learning surrogates. **Journal of Economic Dynamics and Control**, v. 90, p. 366-389, May 2018. Retrieved from: <<http://bit.ly/3OujMSe>>.

LEE, J.-S. et al. The complexities of agent-based modeling output analysis. **Journal of Artificial Societies and Social Simulation**, v. 18, n. 4, p. 4, 2015.

LESPAGNOL, V.; ROUCHIER, J. Trading volume and price distortion: an agent-based model with heterogeneous knowledge of fundamentals. **Computational Economics**, v. 51, n. 4, p. 991-1020, 2018. Retrieved from: <<http://bit.ly/3GGwZoV>>.

LUNDEVALL, B. National innovation systems: analytical concept and development tool. **Industry & Innovation**, v. 14, n. 1, p. 95-119, Feb. 2007. Retrieved from: <<http://bit.ly/3VpupYS>>.

MARCHI, S. de; PAGE, S. E. Agent-based models. **Annual Review of Political Science**, v. 17, n. 1, p. 1-20, 2014. Retrieved from: <<http://bit.ly/3gwddSw>>.

MISHINA, Y. et al. Boosted random forest. **IEICE Transactions on Information and Systems**, v. E98.D, n. 9, p. 1630-1636, 2015. Retrieved from: <<http://bit.ly/3ABcRAH>>.

MITCHELL, C.; WOODMAN, B. Regulation and sustainable energy systems. In: BALDWIN, R.; CAVE, M.; LODGE, M. (Ed.). **The Oxford handbook of regulation**. Oxford: Oxford University Press, 2010. p. 573-590.

MUELLER, B. Why public policies fail: policymaking under complexity. **Economía**, v. 21, n. 2, p. 311-323, May-Aug. 2020. Retrieved from: <<http://bit.ly/3V4H2bK>>.

NELSON, R. R. The co-evolution of technology, industrial structure, and supporting institutions. **Industrial and Corporate Change**, v. 3, n. 1, p. 47-69, 1994.

OLIVAS, E. S. (Ed.). **Handbook of research on machine learning applications and trends: algorithms, methods, and techniques**. Hershey: Information Science Reference, 2010.

POLHILL, J. et al. Crossing the chasm: a 'tube-map' for agent-based social simulation of policy scenarios in spatially-distributed systems. **GeoInformatica**, v. 23, n. 2, p. 169-199, 2019. Retrieved from: <<http://bit.ly/3i8LojJ>>.

REN, Q.; CHENG, H.; HAN, H. Research on machine learning framework based on random forest algorithm. **AIP Conference Proceedings**, n. 1820, 2017. Retrieved from: <<http://bit.ly/3EVYmdj>>.

SCHMIDT, T. S. et al. Do deployment policies pick technologies by (not) picking applications? A simulation of investment decisions in technologies with multiple applications. **Research Policy**, v. 45, n. 10, p. 1965-1983, Dec. 2016. Retrieved from: <<http://bit.ly/3i3TASe>>.

SCOTT, N. et al. SimDrink: an agent-based netlogo model of young, heavy drinkers for conducting alcohol policy experiments. **Journal of Artificial Societies and Social Simulation**, v. 19, n. 1, 2016.

STONE, D. Understanding the transfer of policy failure: bricolage, experimentalism and translation. **Policy & Politics**, v. 45, n. 1, p. 55-70, 2017. Retrieved from: <<http://bit.ly/3i9Nyzm>>.

TAILLANDIER, P. et al. Building, composing and experimenting complex spatial models with the GAMA platform. **GeoInformatica**, v. 23, n. 2, p. 299-322, 2019. Retrieved from: <<http://bit.ly/3tUxXGF>>.

TEN BROEKE, G. et al. The use of surrogate models to analyse agent-based models. **Journal of Artificial Societies and Social Simulation**, v. 24, v. 2, p. 1-25, 2021. Retrieved from: <<http://bit.ly/3gpfIWV>>.

TESFATSION, L. Agent-based computational economics: modeling economies as complex adaptive systems. **Information Sciences**, v. 149, n. 4, p. 262-268, Feb. 2003. Retrieved from: <<http://bit.ly/3V9GXE0>>.

\_\_\_\_\_. Agent-based modeling and institutional design. **Eastern Economic Journal**, v. 37, n. 1, p. 13-19, 2011. Retrieved from: <<http://bit.ly/3AEFmh0>>.

VAN DER HOOG, S. Surrogate modelling in (and of) agent-based models: a prospectus. **Computational Economics**, v. 53, n. 3, p. 1245-1263, 2019. Retrieved from: <<http://bit.ly/3U43ikG>>.

VAN DER VEEN, R. A. C.; KISJES, K. H.; NIKOLIC, I. Exploring policy impacts for servicing in product-based markets: a generic agent-based model. **Journal of Cleaner Production**, v. 145, p. 1-13, Mar. 2017. Retrieved from: <<http://bit.ly/3ExxK18>>.

VAN ENGELEN, J. E.; HOOS, H. H. A survey on semi-supervised learning. **Machine Learning**, v. 109, n. 2, p. 373-440, 2020. Retrieved from: <<https://bit.ly/3YtaVEG>>.

VAZQUEZ, M.; HALLACK, M. The role of regulatory learning in energy transition: the case of solar PV in Brazil. **Energy Policy**, v. 114, p. 465-481, Mar. 2018. Retrieved from: <<http://bit.ly/3VOT46c>>.

### COMPLEMENTARY BIBLIOGRAPHY

ADEPETU, A.; KESHAV, S.; ARYA, V. An agent-based electric vehicle ecosystem model: San Francisco case study. **Transport Policy**, v. 46, p. 109-122, Feb. 2016.

CALDER, M. et al. Computational modelling for decision-making: where, why, what, who and how. **Royal Society Open Science**, v. 5, n. 6, p. 1-15, June 2018.

CHANG, C.-C.; LIN, C.-J. LIBSVM: a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, v. 2, n. 3, p. 1-27, Apr. 2011.

## DISCUSSION PAPER

## APPENDIX

## 1 PERCENTAGE BY POLICY

See table A.1 for the percentage points (p.p.) difference of policies relative to the no-policy baseline.

**TABLE A.1**

**Mean performance of each policy relative to the no-policy baseline for Brazilian Metropolitan Regions (MRs)**

MR/Policy	%	Difference to the 'no-policy' scenario (p.p.)		
	No-policy (baseline)	Purchase	Rent voucher	Monetary aid
Belo Horizonte	100.00	-90.58	0.00	0.00
São Paulo	29.84	-1.14	-0.90	7.23
Rio de Janeiro	27.38	-0.12	1.11	1.33
Joinville	19.43	-14.26	7.67	8.91
Campo Grande	18.63	-15.02	10.02	8.91
Pelotas	18.58	-15.81	10.24	10.71
João Pessoa	18.47	-14.78	10.11	8.97
Maceio	18.43	-15.06	9.84	9.83
Maringá	18.14	-14.33	10.16	10.86
Petrolina-Juazeiro	18.11	-13.97	9.16	9.25
Crato	17.92	-15.45	9.81	10.55
Feira de Santana	17.88	-14.90	10.29	9.08
All	17.87	-14.25	13.16	15.05
Cuiabá	17.59	-14.91	10.70	10.21
Uberlândia	17.57	-13.62	10.73	10.39
Sorocaba	17.56	-14.45	11.14	10.60
SJC <sup>1</sup>	17.54	-14.21	10.55	9.59
Manaus	17.30	-14.58	11.94	11.14
Volta Redonda	17.23	-14.12	9.72	11.38
Ipatinga	17.22	-14.61	11.26	11.34
NH-SL <sup>2</sup>	17.21	-14.78	10.61	11.06

(Continues)

(Continued)

MR/Policy	%	Difference to the 'no-policy' scenario (p.p.)		
	No-policy (baseline)	Purchase	Rent voucher	Monetary aid
Vitória	17.18	-14.97	11.86	10.99
Recife	17.02	-14.62	11.65	10.69
Campina Grande	16.96	-14.73	12.92	11.15
Londrina	16.77	-13.63	11.96	11.48
Natal	16.62	-14.53	10.81	11.59
Florianópolis	16.50	-13.53	12.25	11.57
Ilheus-Itabuna	16.50	-14.49	12.57	12.52
Juiz de Fora	16.40	-12.68	11.81	10.90
Teresina	16.21	-14.34	12.15	12.15
Macapá	16.15	-13.75	11.67	12.36
Ribeirão Preto	16.08	-14.28	12.20	12.16
Aracaju	15.87	-13.97	11.93	11.73
Goiânia	15.86	-14.32	13.58	12.45
Campos	15.83	-13.76	13.06	11.73
São Luis	15.73	-13.54	13.22	11.61
Caxias do Sul	15.71	-13.06	13.46	11.89
Salvador	15.04	-14.01	13.83	13.62
Santos	14.67	-13.48	13.20	13.62
Jundiaí	14.56	-12.39	13.91	15.50
SJRP <sup>3</sup>	14.36	-12.68	13.23	14.07
Fortaleza	13.99	-12.79	14.26	13.88
Curitiba	12.37	-12.30	71.55	79.85
Belém	8.00	-8.00	19.12	61.69
Porto Alegre	4.62	-4.47	13.50	14.24
Campinas	0.94	-0.94	63.56	97.53
Brasília	0.00	0.00	0.00	0.00

Authors' elaboration.

Obs.: Results for Purchase, Rent vouchers and Monetary aid denote performance superior (positive) and inferior (negative) in terms of p.p., compared to that same MR for the no-policy baseline. All MR perform worse when policy Purchase is applied. Conversely, Monetary aid result in positive performance for all MR, with varying levels of improvement.

Notes: <sup>1</sup> São José dos Campos.

<sup>2</sup> Novo Hamburgo/São Leopoldo.

<sup>3</sup> São José do Rio Preto.

## DISCUSSION PAPER

## 1.1 Standard-deviation

See table A.2 for the standard-deviation of the results for each MR.

**TABLE A.2**

**Standard deviation of the no-policy baseline compared to policy performance for Brazilian MR**

MR/Policy	%	Difference to the 'no-policy' scenario (p.p.)		
	No-policy (baseline)	Purchase	Rent voucher	Monetary aid
São Paulo	45.75	-0.51	-0.40	2.55
Rio de Janeiro	44.59	-0.06	0.55	0.65
Joinville	39.56	-17.41	4.89	5.51
Campo Grande	38.94	-20.30	6.27	5.73
Pelotas	38.90	-22.48	6.39	6.61
João Pessoa	38.80	-19.95	6.38	5.82
Maceió	38.78	-20.74	6.25	6.25
Maringá	38.54	-19.39	6.51	6.84
Petrolina-Juazeiro	38.51	-18.59	6.02	6.07
Crato	38.35	-22.82	6.42	6.78
Feira de Santana	38.32	-21.31	6.66	6.05
All	38.31	-19.63	7.95	8.68
Cuiabá	38.08	-21.93	6.96	6.72
Uberlândia	38.06	-18.59	6.99	6.82
Sorocaba	38.05	-20.70	7.19	6.93
SJC <sup>1</sup>	38.03	-20.09	6.91	6.43
Manaus	37.82	-21.56	7.67	7.29
Volta Redonda	37.77	-20.41	6.60	7.42
Ipatinga	37.75	-21.81	7.38	7.42
NH-SL <sup>2</sup>	37.75	-22.36	7.06	7.28
Vitória	37.72	-23.02	7.68	7.26
Recife	37.58	-22.28	7.64	7.18
Campina Grande	37.53	-22.75	8.24	7.42
Londrina	37.36	-19.91	7.89	7.66
Natal	37.23	-22.91	7.39	7.77

(Continues)



(Continued)

MR/Policy	%	Difference to the 'no-policy' scenario (p.p.)		
	No-policy (baseline)	Purchase	Rent voucher	Monetary aid
Florianópolis	37.12	-20.15	8.14	7.81
Ilhéus-Itabuna	37.11	-23.08	8.30	8.28
Juiz de Fora	37.03	-18.12	7.97	7.52
Teresina	36.85	-23.30	8.22	8.23
Macapá	36.80	-21.49	8.01	8.35
Ribeirão Preto	36.74	-23.44	8.29	8.28
Aracaju	36.54	-22.87	8.26	8.16
Goiânia	36.53	-24.20	9.05	8.52
Campos	36.50	-22.25	8.82	8.18
Sao Luis	36.41	-21.76	8.94	8.16
Caxias do Sul	36.39	-20.32	9.06	8.31
Salvador	35.75	-25.64	9.57	9.47
Santos	35.38	-24.54	9.45	9.66
Jundiaí	35.27	-20.71	9.86	10.58
SJRP <sup>3</sup>	35.07	-22.23	9.63	10.04
Fortaleza	34.69	-23.80	10.33	10.15
Curitiba	32.92	-30.22	3.81	-6.13
Belém	27.12	-27.12	17.34	18.84
Porto Alegre	20.99	-17.11	17.53	18.13
Campinas	9.63	-9.63	38.22	2.64
Brasília	0.00	0.00	0.00	0.00
Belo Horizonte	0.00	29.21	0.00	0.00

Authors' elaboration.

Notes: <sup>1</sup> São José dos Campos.

<sup>2</sup> Novo Hamburgo/São Leopoldo.

<sup>3</sup> São José do Rio Preto.

## 1.2 Parameters boundaries configuration

See table A.3 for the minimum, maximum for each model parameter.

### 1.3 Parameters configuration boundaries

We shall briefly explain some parameters, nevertheless Furtado (2022) provides further explanation. Capped low and top values for the bank refer to the total amount of resources available for the bank.

Policy quantile refers to the size of portion of the poorest families to be included as policy recipients.

Decay factor refers to the loss of property value with time, following equation:

$$value = (1 - Discount) * e^{Decay*t} + Discount,$$

in which, *Discount* refers to the maximum offer discount, *t* refers to the time that such property has been on the market, and *Decay* is the decay factor of property value.

Monthly revenue installments division refers only to construction firms, and it refers to the number of months firms take to incorporate revenue after sales. This is relevant only to determine employees' wages.

Available lots per Neighbourhood, in the model "T\_LICENSES\_PER\_REGION" is a monthly designation of licenses for urbanized lots to become available for construction.

Unemployment refers to whether firms observe unemployment or not when setting wages.

Finally, Alternative0 and FPM distribution refer to a previously policy test made by Furtado (2018).

**TABLE A.3**

**Maximum and minimum values or alternatives for simulation parameters**

Parameter	Max	Min
Productivity: exponent	1	0
Productivity: divisor	20	1
Municipal efficiency management	0.001	0.0001
Markup	0.5	0
Sticky Prices	1	0
Perceived market size	100	1
Frequency of firms entering the labor market	1	0
% firms analyzing commuting distance	1	0
Hiring sample size	100	1

(Continues)

(Continued)

Parameter	Max	Min
Tax: consumption	0.6	0.1
Tax: labor	0.6	0.01
Tax over estate transactions	0.01	0.0001
Tax: firm	0.6	0.01
Tax: property	0.01	0.0001
Policy coefficient	0.4	0
Policy days	3600	0
Policy Quantile	1	0
Age cap for borrower at end of contract	100	50
Loan/permanent income ratio	1	0
Maximum Loan-to-Value	1	0
Bank resources: maximum	1	0
Value cap for banks: top	2	1
Value cap for banks: bottom	1	0
Supply-demand effect on real estate prices	5	0
Decay factor for properties	0	-0.1
Maximum offer discount	1	0.4
% families entering real estate market	0.01	0
Neighborhood effect	5	0
Rental Share	1	0
Initial rental price	0.01	0
% of construction firms	0.2	0
Monthly revenue installments division (in months)	100	1
Cost of lots (% of construction)	0.7	0
Cost of private transit	0.5	0
Cost of public transit	0.5	0
% of population	1	0
Total Days	14610	1826
Available lots per neighbourhood	True, false	
Starting day	2010-01-01, 2000-01-01	
Interest	nominal, real, fixed	
Wage is unrelated to unemployment	True, false	
Alternative0	True, false	
FPM distribution	True, false	
Policies	No-policy, Purchase, Monetary aid, rent vouchers	

Authors' elaboration.

## REFERENCES

FURTADO, B. A. PolicySpace: a modeling platform. **Journal on Policy and Complex Systems**, v. 4, n. 2, p. 17-30, 2018. Retrieved from: <<http://bit.ly/3uY3WWS>>.

\_\_\_\_\_. PolicySpace2: modeling markets and endogenous public policies. **Journal of Artificial Societies and Social Simulation**, v. 25, n. 1, p. 1-30, 2022. Retrieved from: <<http://bit.ly/3XCh9IB>>.

# **Ipea – Institute for Applied Economic Research**

## **PUBLISHING DEPARTMENT**

### **Head of the Publishing Department**

Aeromilson Trajano de Mesquita

### **Assistants to the Head of the Department**

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

### **Supervision**

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

### **Typesetting**

Anderson Silva Reis

Cristiano Ferreira de Araújo

Danielle de Oliveira Ayres

Danilo Leite de Macedo Tavares

Leonardo Hideki Higa

### **Cover design**

Aline Cristine Torres da Silva Martins

### **Graphic design**

Aline Cristine Torres da Silva Martins

*The manuscripts in languages other than Portuguese published herein have not been proofread.*

### **Ipea – Brasilia**

Setor de Edifícios Públicos Sul 702/902, Bloco C

Centro Empresarial Brasilia 50, Torre B

CEP: 70390-025, Asa Sul, Brasilia-DF

**Ipea's mission**  
Enhance public policies that are essential to Brazilian development  
by producing and disseminating knowledge and by advising  
the state in its strategic decisions.



**ipea** Instituto de Pesquisa  
Econômica Aplicada

MINISTÉRIO DO  
PLANEJAMENTO  
E ORÇAMENTO

