Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structural support constraints prevent invalid surrogate predictions in branched agent-based simulations

Alaa Aldin Hajjar

a.hajjar@innopolis.university

Innopolis University

Armen Beklaryan

National Research University Higher School of Economics

Article

Keywords:

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9808186/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Structural support constraints prevent invalid surrogate predictions in branched agent-based simulations

Alaa Aldin Hajjar[1,*] **and Armen Beklaryan**[2]

[1]Lab of Machine Learning and Knowledge Representation, Research Center of the Artificial Intelligence Institute, Innopolis University, Innopolis, Russia
[2]Faculty of Computer Science, HSE University, Moscow, Russia
[*]Correspondence: a.hajjar@innopolis.university

ABSTRACT

Branched simulators can impose deterministic output support: some outputs are possible only under specific simulator branches. Standard surrogate regressors may ignore this logic and predict values that the simulator itself could never generate. We evaluate a structural-support validation workflow for branched agent-based simulations, combining branch-aware representation, simulator-derived support masking, grouped validation by simulator invocation, capacity control, and model-family comparison. The testbed is an affordability-constrained residential segregation simulator in which blockage outcomes are structurally zero whenever the economic-feasibility branch is inactive. The dataset contains 400 simulator invocations expanded into 1600 condition-specific rows. The hard mask enforces the known no-economy zero-support rule, reducing invalid blockage predictions to zero while preserving useful economy-active accuracy. In grouped five-fold validation, the selected 137-parameter masked multilayer perceptron achieves root-mean-square errors of 0.0494 for blocked-mobility burden and 0.0746 for final affordability-blocked share, with masked tree and two-stage baselines providing close comparisons. Targeted verification of 40 response-map points with 120 direct simulator replicates gives mean regime agreement of 0.7417. Dense diagnostic mapping is estimate-efficient: surrogate inference takes 0.0566 s, while direct simulation is estimated at 14,209–361,228 s. The response maps are simulator-internal diagnostics, not real-world causal or policy evidence.

Introduction

Many computational simulators contain branch logic that determines whether an output can exist at all. A simulator may activate a constraint, intervention, event type, or feasibility filter only under selected conditions. In those settings, the valid output set depends on the input configuration. A surrogate model that approximates the simulator can therefore be numerically close on average while still producing outputs that contradict the simulator's own logic.

Surrogate modeling is widely used to make expensive simulation studies more tractable. Learned emulators can support dense scenario exploration, sensitivity analysis, response-surface inspection, and approximate optimization at much lower cost than repeated direct simulation[1–6]. Agent-based models are a challenging case because stochasticity, discrete updates, local interactions, and parameter-dependent runtime can make systematic exploration expensive even when the model rules are simple[8–10,12]. The same discreteness that makes agent-based models useful can also create deterministic branches that ordinary tabular regressors may smooth across.

This paper studies that problem in an affordability-constrained residential segregation simulator. The simulator expands each design point into four condition rows: baseline without the economic constraint, baseline with the economic constraint, influencer variant without the economic constraint, and influencer variant with the economic constraint. The two primary outcomes are `blocked_move_burden` and `final_affordability_blocked_share`. Both are defined by affordability blocking and are structurally zero when `use_economic_constraint = 0`. A positive no-economy blockage prediction is therefore not merely inaccurate; it lies outside the simulator-implied support.

We define structural-support validity as the requirement that a surrogate prediction belongs to the simulator-implied valid output set,

$$\hat{y}_j(x) \in \mathscr{S}_j(x), \tag{1}$$

where x is the simulator input and $\mathscr{S}_j(x)$ is the valid support of outcome j. For the two blockage outcomes in this case study,

the support rule is

$$\mathscr{S}_j(x) = \begin{cases} \{0\}, & \text{if } e = 0, \\ [0,1], & \text{if } e = 1, \end{cases} \tag{2}$$

where $e \in \{0,1\}$ is the observed economic-constraint indicator, with $e = 0$ denoting an inactive economic constraint and $e = 1$ denoting the economy-active branch. The corresponding post-hoc support mask is

$$g_j(x) = e\,\mathrm{clip}(f_j(x), 0, 1), \tag{3}$$

where $f_j(x)$ is the unconstrained surrogate prediction and $g_j(x)$ is the structurally masked prediction.

The contribution is methodological rather than architectural. The workflow can be layered on top of linear, tree-based, and neural surrogate regressors to detect and prevent structurally invalid predictions in branched simulators with observed deterministic branch indicators. The empirical question is not whether a known branch indicator can mechanically force zeros, but whether ordinary surrogate workflows violate simulator-implied output support, whether branch-aware representation alone is sufficient, whether support enforcement preserves economy-active predictive accuracy, and whether this result is stable across the evaluated surrogate families under grouped validation.

We make five contributions. First, we formulate a structural-support validation workflow for branched simulator surrogates. Second, we show that reduced feature representations can collapse branch regimes. Third, we evaluate a hard simulator-derived support mask under grouped validation by simulator invocation. Fourth, we compare capacity-controlled neural, linear, tree-based, and two-stage structural baselines. Fifth, we use the resulting structurally valid surrogate for simulator-internal diagnostic mapping and targeted direct-simulation checks, while explicitly avoiding real-world causal or policy claims.

Results

The main experimental dataset contains 400 unique simulator invocations and 1600 condition-specific rows. Each invocation expands to four rows, one for each condition. Grouped five-fold cross-validation keeps all four rows from the same invocation in the same fold, so the validation unit is the simulator invocation rather than the condition row. Thus, a validation fold holds out simulator invocations, not individual condition rows, and all branch rows from a held-out invocation are evaluated only in the validation fold. The two primary outcomes are blockage outcomes; no-economy rows are structural zeros.

Branch integrity reveals a structural-support failure mode

The expanded dataset satisfies the simulator-implied support rule. Across the 800 no-economy rows, blocked-mobility burden is zero in 800/800 rows and final affordability-blocked share is zero in 800/800 rows. The reduced representation consisting only of lattice size, seed, affordability margin, and vacancy fraction collapses all four condition rows for each simulator invocation. All 400 invocation groups collapse under that reduced key, confirming that branch-defining variables are required for a valid surrogate representation.

The representation ablation shows the consequence of omitting branch information. Under the reduced representation, the unmasked MLP-8 produced structurally invalid no-economy predictions for all 800 no-economy held-out predictions for each primary outcome. Grouped-CV RMSE was 0.1344 for blocked-mobility burden and 0.1348 for final affordability-blocked share. Restoring branch-aware variables reduced grouped-CV RMSE to 0.0496 and 0.0735, respectively. However, representation alone did not guarantee structural validity for the unmasked MLP-8: positive no-economy blockage predictions remained possible. Figure 2 summarizes the transition from reduced representation to branch-aware representation and then to hard support masking.

Hard support masking preserves useful active-branch prediction

The hard post-hoc mask eliminates invalid no-economy predictions by construction because it applies the known simulator-implied support rule. The empirical question is therefore whether support enforcement damages the predictive task in the economy-active branch. In the selected masked MLP-8 model, invalid no-economy predictions were reduced to zero for both primary outcomes, while grouped-CV RMSE remained 0.0494 for `blocked_move_burden` and 0.0746 for `final_affordability_blocked_share`. Economy-active RMSE values were 0.0698 and 0.1055, respectively.

The result is not evidence that the hard mask learns the structural rule. The rule is provided by the simulator logic. The evidence is that ordinary surrogate workflows can violate that rule, branch-aware representation reduces but does not guarantee support validity for all evaluated regressors, and explicit support enforcement can be applied without losing useful active-branch accuracy. Figure 4 shows grouped out-of-fold predictions for the selected masked MLP-8 model. The visible compression at the upper end is consistent with a low-capacity model trained on a sparse stochastic design: high-blockage regimes are rarer, bounded to $[0,1]$, and adjacent to transition regions where small input changes can alter feasibility outcomes.

The conclusion is stable across evaluated surrogate families

The model-family comparison shows that the structural-support result is not specific to a single neural architecture. Masked ridge regression, masked tree ensembles, masked gradient boosting, masked MLPs, and two-stage active-regime baselines all eliminated no-economy support violations for the two primary blockage outcomes. Their grouped-CV errors were tightly clustered among the strongest non-linear models.

The selected masked MLP-8 achieved grouped-CV RMSE 0.0494 ± 0.0037 for `blocked_move_burden` and 0.0746 ± 0.0042 for `final_affordability_blocked_share`. The masked Random Forest, masked ExtraTrees, masked MLP-16, and two-stage MLP-8 remained close comparison models. The selected MLP-8 is therefore retained as the primary model because it is structurally valid, low-complexity, reproducible, and within the one-standard-error region of the strongest alternatives, not because it is uniquely superior. The similarity between the masked MLP-8 and the two-stage active-regime baseline indicates that post-hoc support masking does not materially worsen accuracy compared with training a regressor only on the economy-active branch. Figure 3 and Table 2 report the compact model-family evidence.

Diagnostic response maps remain simulator-internal

The structurally valid surrogate was used to construct dense diagnostic response maps over affordability margin and vacancy fraction. Figure 5 shows the predicted response surface for blocked-mobility burden at $L = 40$ in the baseline economy branch. These maps recover the expected structural-zero no-economy regime and show that predicted blockage is highest when the affordability margin is small and vacancies are relatively high. The response surface therefore provides a compact simulator-internal summary of where the implemented affordability filter is most restrictive.

Targeted direct simulation was used to check selected response-map regions. Forty points were sampled from low-, medium-, high-, and near-boundary or high-gradient regions. Three simulator seeds were run at each point, producing 120 completed replicate rows. Mean absolute error was 0.0551 for `blocked_move_burden` and 0.0687 for `final_affordability_blocked_share`; mean regime agreement by point was 0.7417. Given only three stochastic direct-simulation seeds per selected point, this value is interpreted as moderate targeted agreement rather than definitive phase-boundary recovery. The maps should be interpreted as surrogate-assisted summaries of the implemented simulator, not as empirical claims about real residential mobility or housing policy.

Discussion

This study shows that surrogate models for branched simulators should be evaluated not only by predictive error but also by whether their outputs lie in the simulator-implied support. In the affordability-constrained segregation simulator, blockage outcomes are structurally zero whenever the economic-feasibility branch is inactive. A positive no-economy blockage prediction contradicts the simulator even if its contribution to aggregate RMSE is small.

The hard support mask used here is intentionally simple. It is appropriate because the branch indicator is observed, the support rule is deterministic, and the blockage outcomes are bounded. Under these conditions, post-hoc masking is a transparent validity layer rather than a hidden modeling assumption. The same idea would not automatically apply when branch activation is latent, when the support rule is uncertain, when an output is undefined rather than zero, or when support depends on unobserved internal simulator state. In those settings, a hard mask could hide rather than solve the modeling problem. Future work should extend the workflow to latent or hidden branches, where branch membership must be inferred or jointly modeled rather than supplied as an observed simulator input.

The model-family comparison is central to the interpretation. The study does not depend on a new neural-network architecture. Masked tree models and two-stage active-regime models perform similarly to the selected masked MLP-8, and all structurally masked or two-stage variants eliminate no-economy violations. The MLP-8 is selected because it is accurate, stable, low-complexity, and reproducible, not because it is uniquely superior.

The response maps demonstrate the practical value of a checked surrogate. They make dense simulator-internal diagnostic mapping feasible and identify regions for future direct simulation. The runtime analysis supports this use case only as an estimate-based comparison: surrogate inference time is measured, but the corresponding dense direct simulator sweep is estimated from observed and modeled simulator runtimes. The dense sweep itself was not executed.

The workflow is applicable to simulators with observed branch indicators and known deterministic support constraints. Examples include intervention-gated epidemiological simulations, feasibility-filtered engineering simulations, policy-switch models, and ecological models with regime-dependent outputs. This broader relevance is conditional on the branch and support assumptions. The present study evaluates the workflow in one affordability-constrained ABM testbed; it does not demonstrate performance across all such domains.

The main limitations follow from the controlled scope of the study. The evidence comes from one branched ABM and two primary blockage outcomes. The expanded dataset contains 400 unique simulator invocations, and grouped cross-validation is limited by that number even though the condition-specific table contains 1600 rows. The outputs are simulator-defined quantities,

not externally validated housing-market measurements. Targeted direct simulation strengthens selected response-map regions, but it does not establish full phase-boundary recovery. Generalization to other simulators, hidden branches, high-dimensional outputs, and more complex support structures requires additional studies.

Methods

Case-study simulator and condition expansion

The case study is an affordability-constrained residential segregation ABM. Each unique simulator invocation is defined by lattice size `size_L`, simulator seed, affordability margin, and vacancy fraction. Each invocation expands into four condition-specific rows crossing the simulator variant with the economic-constraint branch: `baseline_noecon`, `baseline_econ`, `with_influencer_noecon`, and `with_influencer_econ`. The expanded dataset contains 400 unique invocations and 1600 condition-specific rows.

The main design uses two lattice sizes, 25 seeds, four affordability margins, and two vacancy fractions, yielding 400 unique simulator invocations. The exact parameter values, generation command, and resumability details are recorded in the released design and run-configuration artifacts. No row-wise random split is used for the reported evaluations.

Outcomes and structural support

The primary outcomes are `blocked_move_burden` and `final_affordability_blocked_share`. Both are run-level simulator summaries of affordability blockage. Let $\mathcal{T}_i$ denote the recorded step set for condition row i, $D_{i,t}$ the number of dissatisfied agents at step t, $B_{i,t}$ the number of blocked moves at step t, and $A_{i,t}$ the number of affordability-blocked agents at step t. The blocked-mobility burden is the mean recorded blocked-move rate,

$$y_{b,i} = \frac{1}{|\mathcal{T}_i|} \sum_{t \in \mathcal{T}_i} \begin{cases} B_{i,t}/D_{i,t}, & D_{i,t} > 0, \\ 0, & D_{i,t} = 0. \end{cases} \tag{4}$$

The final affordability-blocked share is the final-step affordability-blocked count divided by the final-step dissatisfied-agent count,

$$y_{a,i} = \begin{cases} A_{i,T_i}/D_{i,T_i}, & D_{i,T_i} > 0, \\ 0, & D_{i,T_i} = 0, \end{cases} \tag{5}$$

where T_i is the final recorded step for row i.

For both primary outcomes, the support rule is deterministic. If `use_economic_constraint` $= 0$, the outcome support is $\{0\}$. If `use_economic_constraint` $= 1$, the support is bounded to $[0, 1]$. Branch-integrity checks verify that both blockage outcomes are zero in all 800 no-economy rows.

Feature representations

The reduced representation contains `size_L`, seed, affordability margin, and vacancy fraction. The minimal branch-aware representation adds variant and `use_economic_constraint`. The full branch-aware representation also includes the named condition. The final selected predictive representation removes seed and condition, using `size_L`, affordability margin, vacancy fraction, variant, and `use_economic_constraint`. Seed remains part of the simulator invocation identity and grouped split construction.

Seed-removed and condition-removed sensitivity checks were evaluated with the same grouped folds. Removing seed changed best grouped-CV RMSE by less than 0.01 for both primary outcomes. Adding condition did not materially improve performance once variant and `use_economic_constraint` were included. Condition is therefore retained for row identity and reporting, not required as a primary predictive feature.

Surrogate families and hard support mask

The evaluated surrogate families include a mean baseline, ridge regression, Random Forest, ExtraTrees, Gradient Boosting, MLP-8, MLP-16, post-hoc masked variants, and two-stage active-regime baselines. The final selected MLP-8 has two hidden layers of width 8, ReLU activations, Adam optimization, learning rate 0.01, weight decay 0.0001, 300 epochs, and mean-squared error on standardized targets. The selected configuration uses random seed 20260514 and fold training seeds 20260515–20260519.

For each blockage outcome, the hard post-hoc support mask is given by equation (3). If the economic-constraint branch is inactive, prediction is forced to zero. Otherwise, the raw prediction is clipped to $[0, 1]$. The two-stage structural baseline is also structurally valid by design: it assigns zero to no-economy rows and fits its regressor only on economy-active rows. Its close performance to the post-hoc masked MLP-8 is therefore a comparison between two support-valid designs, showing that the hard mask does not degrade accuracy relative to a cleanly separated active-regime model.

Grouped validation and metrics

All main evaluations use grouped five-fold cross-validation by simulator invocation. The group identity contains lattice size, seed, affordability margin, and vacancy fraction, ensuring that all four condition rows from one simulator invocation remain in the same fold. With 400 invocation groups, each fold contains 80 held-out groups and 320 condition-specific rows. This grouping prevents the four branch rows generated from the same invocation from being split between training and validation folds.

Metrics include RMSE, mean absolute error, Spearman correlation, high-regime F1, structural invalidity count, out-of-range prediction count, economy-active RMSE and MAE, and near-boundary MAE. Near-boundary rows are defined as rows whose observed outcome lies within 0.05 of the relevant high-regime threshold; the width was chosen as a 5% margin on the unit-bounded outcome scale to mark transition-adjacent cases without treating the entire active regime as boundary. Fold-level uncertainty is reported as grouped-fold standard deviation unless otherwise specified.

Model selection prioritizes zero structural-support violations, economy-active predictive accuracy, grouped-CV stability, simplicity, and reproducibility. For each primary outcome, the one-standard-error threshold is $\bar{e}_{\min} + s_{\min}/\sqrt{5}$, where $\bar{e}_{\min}$ and $s_{\min}$ are the mean and grouped-fold standard deviation of the best structurally valid model.

Diagnostic response maps and targeted verification

The selected structurally valid surrogate is evaluated over a dense diagnostic grid of affordability margin and vacancy fraction. The maps are constructed after applying the hard support mask, so no-economy branches remain structurally zero. Forty verification points were sampled as a fixed diagnostic budget: ten points from each of four map regions—low, medium, high, and near-boundary or high-gradient predictions. Each point was evaluated with three direct simulator seeds, producing 120 replicate rows. This design was intended to check representative response-map regimes under a bounded simulation budget, not to estimate a dense phase boundary.

Runtime-estimate comparison

The runtime comparison is estimate-based. Let N_{grid} be the number of unique simulator invocations in the diagnostic grid, $\bar{T}_{\text{obs}}$ the observed mean runtime per completed expanded-dataset invocation, $\hat{T}(x)$ the conservative runtime-model estimate, and T_{sur} the measured surrogate inference time. The lower-bound dense direct runtime estimate is

$$T_{\text{direct,lb}} = N_{\text{grid}} \bar{T}_{\text{obs}}, \tag{6}$$

and the conservative runtime-model estimate is

$$T_{\text{direct,cons}} = \sum_{x \in \mathscr{D}_{\text{grid}}} \hat{T}(x). \tag{7}$$

The diagnostic grid has 2542 unique invocations and 10,168 condition rows. The expanded campaign mean runtime gives $T_{\text{direct,lb}} = 14{,}209.0174$ s; the conservative runtime model gives $T_{\text{direct,cons}} = 361{,}228.3647$ s; measured surrogate inference time is $T_{\text{sur}} = 0.056635$ s. The corresponding estimate-based speedup range is 2.51×10^5 to 6.38×10^6. The dense direct sweep was not executed.

Ethics and reproducibility

This study uses simulator-generated data only. It does not involve human participants, human tissue, live vertebrates, or higher invertebrates.

The release package contains the expanded dataset, data dictionary, schema, grouped split files, simulator source and build instructions, structural-mask implementation, grouped-CV scripts, model-family comparison scripts, table and figure generation scripts, runtime-estimate scripts, and environment files. Large raw outputs and frame archives are excluded from the GitHub repository, and simulator reruns are optional.

During manuscript preparation, the authors used OpenAI Codex/GPT-based tooling for code scaffolding, report drafting, LaTeX editing, and consistency checking. The authors reviewed, corrected, and validated the resulting materials and take full responsibility for the final content. No AI tool was used as an author.

Data availability

The expanded condition-specific dataset, data dictionary/schema, grouped split files, generated table files, generated figure inputs, and lightweight reproduction artifacts are available in the public GitHub repository: https://github.com/AlaaAldinInno/Structural-Support-Validity-for-Surrogate-Emulation-of-Branched-Agent-Based-Simulations. The version used for this manuscript is commit `f8e2818992677940e5a8fd5ccab25e989658d231`. Large raw simulation outputs and frame archives are excluded from the repository because they are not required to reproduce the reported surrogate evidence from the released processed data.

Code availability

The simulator source code, build instructions, grouped-validation scripts, structural-mask implementation, model-family comparison scripts, diagnostic-map scripts, runtime-estimate scripts, and figure/table generation scripts are available in the same public GitHub repository: https://github.com/AlaaAldinInno/Structural-Support-Validity-for-Surrogate-Emulation-of-Branched-Agent-Based-Simulations. The manuscript version corresponds to commit `f8e2818992677940e5a8fd5ccab25e989658d231`. The repository includes environment files and lightweight reproduction commands. No proprietary software or restricted data are required.

References

1. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and Analysis of Computer Experiments. *Statistical Science* **1989**, *4*. https://doi.org/10.1214/ss/1177012413.
2. Kennedy, M.C.; O'Hagan, A. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **2001**, *63*, 425–464. https://doi.org/10.1111/1467-9868.00294.
3. Forrester, A.I.; Keane, A.J. Recent Advances in Surrogate-Based Optimization. *Progress in Aerospace Sciences* **2009**, *45*, 50–79. https://doi.org/10.1016/j.paerosci.2008.11.001.
4. Bhosekar, A.; Ierapetritou, M. Advances in Surrogate Based Modeling, Feasibility Analysis, and Optimization: A Review. *Computers & Chemical Engineering* **2018**, *108*, 250–267. https://doi.org/10.1016/j.compchemeng.2017.09.017.
5. Santner, T.J.; Williams, B.J.; Notz, W.I. *The Design and Analysis of Computer Experiments*, 2 ed.; Springer: New York, NY, USA, 2018. https://doi.org/10.1007/978-1-4939-8847-1.
6. Gramacy, R.B. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2020. https://doi.org/10.1201/9780367815493.
7. Razavi, S.; Tolson, B.A.; Burn, D.H. Review of Surrogate Modeling in Water Resources. *Water Resources Research* **2012**, *48*, W07401. https://doi.org/10.1029/2011WR011527.
8. Bonabeau, E. Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 7280–7287. https://doi.org/10.1073/pnas.082080899.
9. Angione, C.; Silverman, E.; Yaneske, E. Using Machine Learning as a Surrogate Model for Agent-Based Simulations. *PLOS ONE* **2022**, *17*, e0263150. https://doi.org/10.1371/journal.pone.0263150.
10. Lamperti, F.; Roventini, A.; Sani, A. Agent-Based Model Calibration Using Machine Learning Surrogates. *Journal of Economic Dynamics and Control* **2018**, *90*, 366–389. https://doi.org/10.1016/j.jedc.2018.03.011.
11. Natterer, E.S.; Rao, S.R.; Tejada Lapuerta, A.; Engelhardt, R.; Hörl, S.; Bogenberger, K. Machine Learning Surrogates for Agent-Based Models in Transportation Policy Analysis. *Transportation Research Part C: Emerging Technologies* **2025**, *180*, 105360. https://doi.org/10.1016/j.trc.2025.105360.
12. Patsatzis, D.G.; Russo, L.; Kevrekidis, I.G.; Siettos, C. Data-Driven Control of Agent-Based Models: An Equation/Variable-Free Machine Learning Approach. *Journal of Computational Physics* **2023**, *478*, 111953. https://doi.org/10.1016/j.jcp.2023.111953.
13. Schelling, T.C. Dynamic Models of Segregation. *The Journal of Mathematical Sociology* **1971**, *1*, 143–186. https://doi.org/10.1080/0022250X.1971.9989794.
14. Ubarevičienė, R.; van Ham, M.; Tammaru, T. Fifty Years after Schelling's Models of Segregation: Bibliometric Analysis of the Legacy of Schelling and the Future Directions of Segregation Research. *Cities* **2024**, *147*, 104838. https://doi.org/10.1016/j.cities.2024.104838.
15. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering* **2017**, *29*, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.
16. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nature Reviews Physics* **2021**, *3*, 422–440. https://doi.org/10.1038/s42254-021-00314-5.
17. Marra, G.; Dumančić, S.; Manhaeve, R.; De Raedt, L. From Statistical Relational to Neurosymbolic Artificial Intelligence: A Survey. *Artificial Intelligence* **2024**, *328*, 104062. https://doi.org/10.1016/j.artint.2023.104062.

18. Brosowsky, M.; Keck, F.; Dünkel, O.; Zöllner, J.M. Sample-Specific Output Constraints for Neural Networks. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2021, Vol. 35, pp. 6812–6821. https://doi.org/10.1609/aaai.v35i8.16841.

19. Amos, B.; Kolter, J.Z. OptNet: Differentiable Optimization as a Layer in Neural Networks. In Proceedings of the Proceedings of the 34th International Conference on Machine Learning, 2017, Vol. 70, *Proceedings of Machine Learning Research*, pp. 136–145.

20. Donti, P.L.; Rolnick, D.; Kolter, J.Z. DC3: A Learning Method for Optimization with Hard Constraints. In Proceedings of the International Conference on Learning Representations, 2021.

21. Özese, D.; Birbil, Ş.İ.; Baydoğan, M.G. Output-Constrained Decision Trees. *arXiv preprint arXiv:2405.15314* **2024**.

22. Cragg, J.G. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica* **1971**, *39*, 829–844.

23. Mullahy, J. Specification and Testing of Some Modified Count Data Models. *Journal of Econometrics* **1986**, *33*, 341–365. https://doi.org/10.1016/0304-4076(86)90002-3.

24. Lambert, D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* **1992**, *34*, 1–14. https://doi.org/10.1080/00401706.1992.10485228.

25. Min, Y.; Agresti, A. Modeling Nonnegative Data with Clumping at Zero: A Survey. *Journal of the Iranian Statistical Society* **2002**, *1*, 7–33.

26. Jacobs, R.A.; Jordan, M.I.; Nowlan, S.J.; Hinton, G.E. Adaptive Mixtures of Local Experts. *Neural Computation* **1991**, *3*, 79–87.

27. Jordan, M.I.; Jacobs, R.A. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation* **1994**, *6*, 181–214. https://doi.org/10.1162/neco.1994.6.2.181.

28. Gramacy, R.B.; Lee, H.K.H. Bayesian Treed Gaussian Process Models with an Application to Computer Modeling. *Journal of the American Statistical Association* **2008**, *103*, 1119–1130. https://doi.org/10.1198/016214508000000689.

29. Pya, N.; Wood, S.N. Shape Constrained Additive Models. *Statistics and Computing* **2015**, *25*, 543–559. https://doi.org/10.1007/s11222-013-9448-7.

30. Gupta, M.; Cotter, A.; Pfeifer, J.; Voevodski, K.; Canini, K.; Mangylov, A.; Moczydlowski, W.; van Esbroeck, A. Monotonic Calibrated Interpolated Look-Up Tables. *Journal of Machine Learning Research* **2016**, *17*, 1–47.

31. Amos, B.; Xu, L.; Kolter, J.Z. Input Convex Neural Networks. In Proceedings of the Proceedings of the 34th International Conference on Machine Learning, 2017, Vol. 70, *Proceedings of Machine Learning Research*, pp. 146–155.

32. Cranmer, K.; Brehmer, J.; Louppe, G. The Frontier of Simulation-Based Inference. *Proceedings of the National Academy of Sciences* **2020**, *117*, 30055–30062. https://doi.org/10.1073/pnas.1912789117.

33. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics* **2010**, *26*, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

34. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems 30, 2017.

35. ten Broeke, G.; van Voorn, G.; Ligtenberg, A. Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation* **2016**, *19*, 5. https://doi.org/10.18564/jasss.2857.

36. Borgonovo, E.; Pangallo, M.; Rivkin, J.; Rizzo, L.; Siggelkow, N. Sensitivity Analysis of Agent-Based Models: A New Protocol. *Computational and Mathematical Organization Theory* **2022**, *28*, 52–94. https://doi.org/10.1007/s10588-021-09358-5.

Acknowledgements

This work was supported by the Ministry of Economic Development of the Russian Federation, agreement No. 139-10-2025-034 dated 19.06.2025, IGK 000000C313925P4D0002.

Author contributions statement

A.A.H. and A.B. conceptualized the study and designed the methodology. A.A.H. implemented the software, conducted validation, performed the formal analysis, curated the data, generated the visualizations, and prepared the original manuscript draft. A.A.H. and A.B. reviewed and edited the manuscript. A.B. supervised the project. Both authors read and approved the final manuscript.

Funding:

This research did receive funding. Alaa Aldin Hajjar received funding from Ministry of Economic Development of the Russian Federation; Grant ID No. 139-10-2025-034 dated 19.06.2025, IGK 000000C313925P4D0002.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to A.A.H.

Figure legends

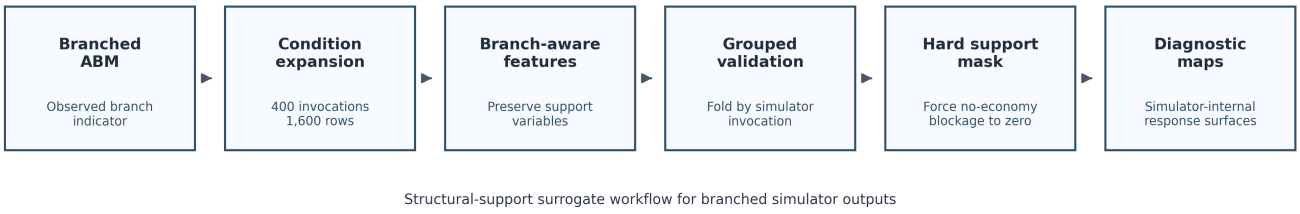


Figure 1. Structural-support validation workflow for surrogate emulation of the branched simulator. The workflow starts from the branched ABM, expands each simulator invocation into condition-specific rows, preserves branch-aware inputs, applies simulator-derived hard support masking, evaluates models with grouped cross-validation by simulator invocation, and uses the checked surrogate for diagnostic response mapping.

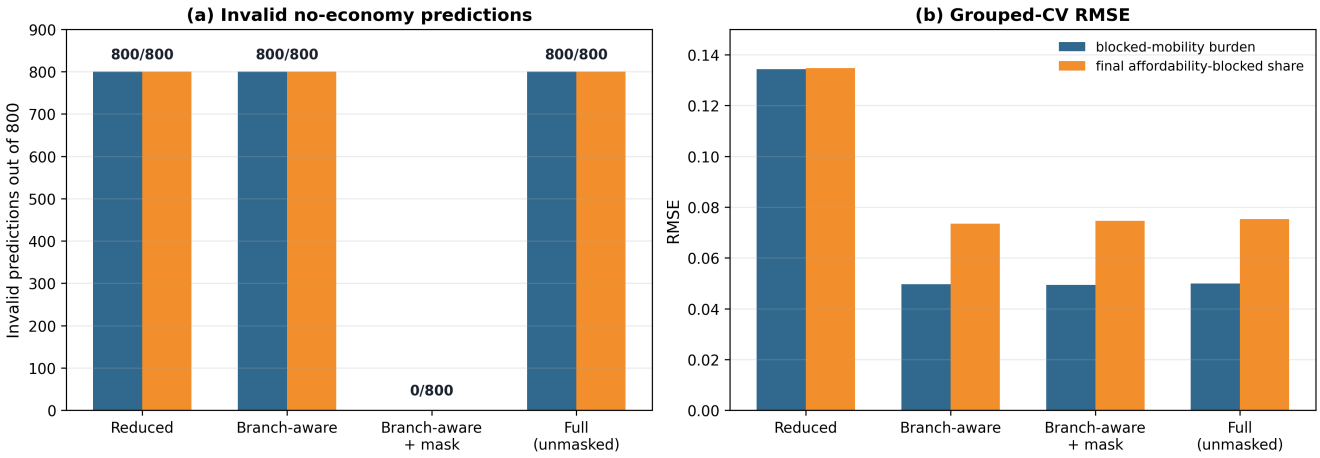


Figure 2. Representation and hard-mask ablation for the two primary blockage outcomes. Panel (a) reports structurally invalid no-economy predictions among 800 no-economy rows. Reduced and unmasked branch-aware representations can produce positive blockage predictions in branches where blockage is impossible. The full unmasked representation also remains structurally invalid because branch variables reduce prediction error but do not constrain the output support. The hard support mask sets these predictions to zero by construction. Panel (b) reports grouped-CV RMSE. The empirical question is not whether the mask mechanically enforces zeros, but whether support enforcement preserves useful predictive accuracy in the economy-active branch.

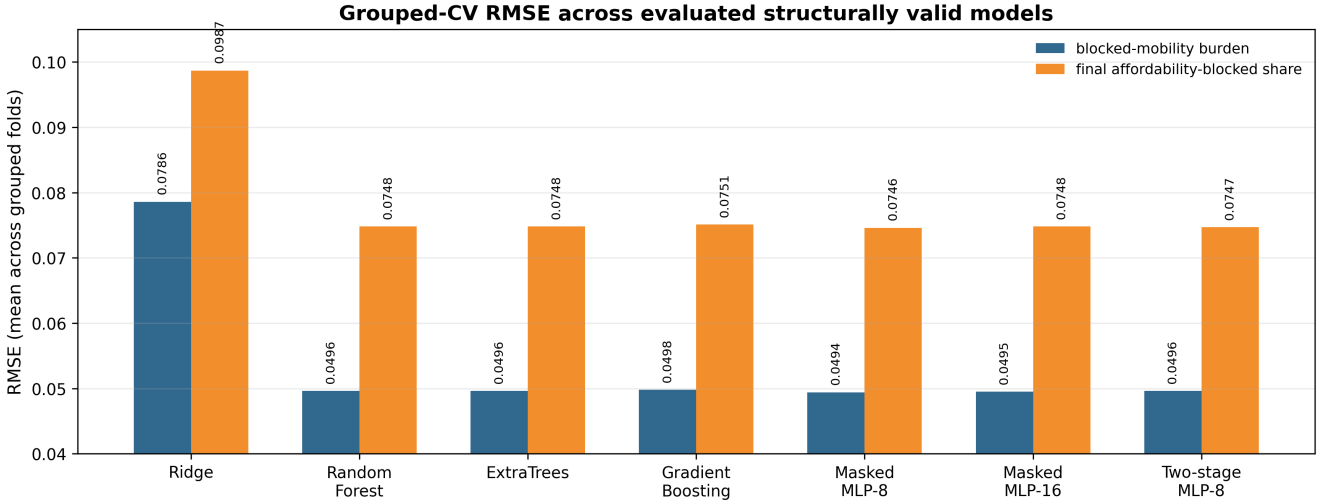


Figure 3. Grouped-CV RMSE across evaluated structurally valid models for the two primary blockage outcomes. Bars report mean grouped-fold RMSE. The selected masked MLP-8 remains close to the strongest structurally valid alternatives while using a low-complexity architecture. Exact fold means and fold standard deviations are reported in Table 2.

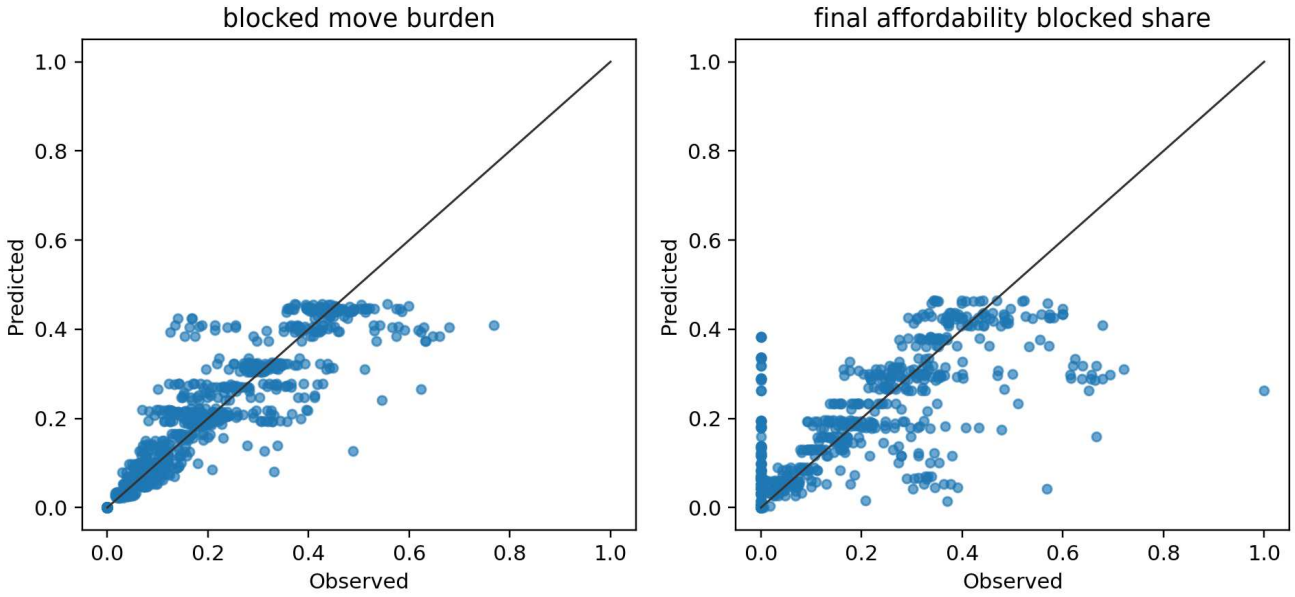


Figure 4. Grouped out-of-fold predicted versus observed values for the selected masked MLP-8 model. The diagonal line indicates perfect prediction. The plots show useful agreement under grouped validation, with visible compression at higher observed values and weaker behavior near transition regions. Structural-zero no-economy predictions are masked; observed zeros may also occur in economy-active rows.

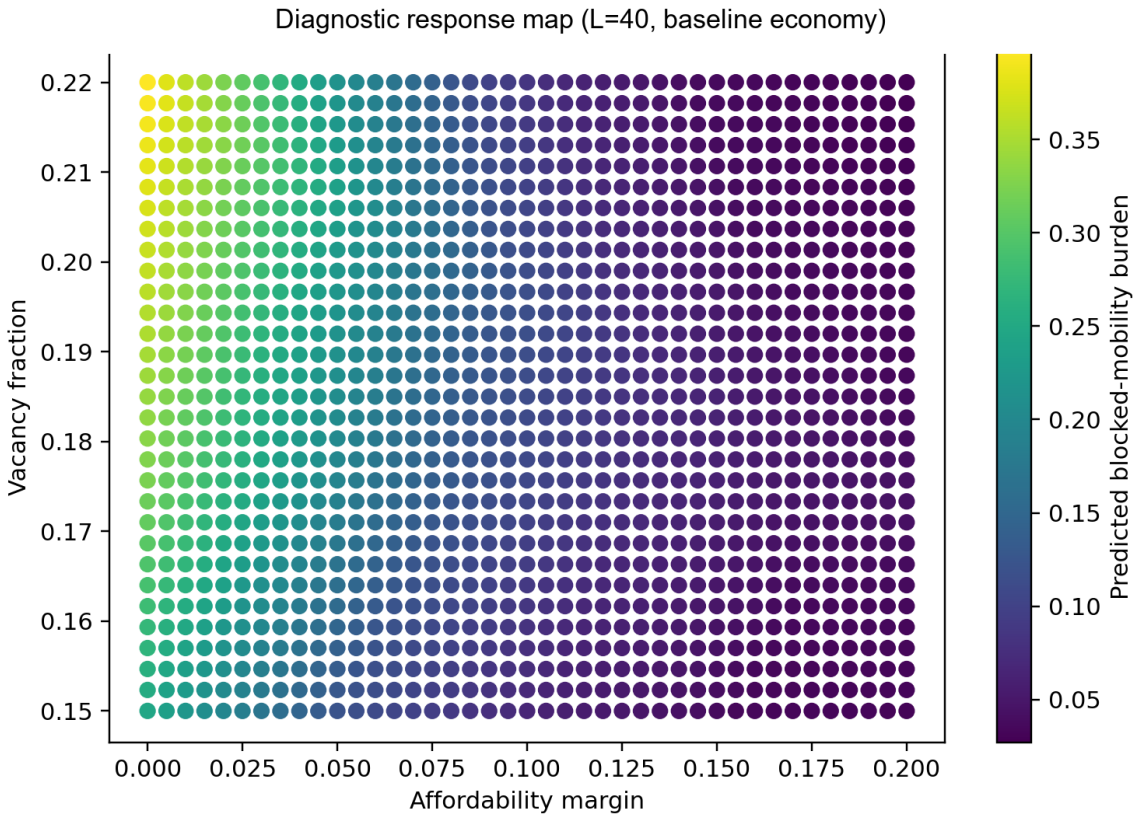


Figure 5. Surrogate-predicted diagnostic response map for blocked-mobility burden at $L = 40$ in the baseline economy branch. The map varies affordability margin and vacancy fraction while applying the hard support mask. It should be interpreted as a simulator-internal diagnostic response surface, not as a validated empirical phase boundary or real-world policy prediction. Targeted direct-simulation verification of selected map regions is summarized in Table 3.

Tables

Table 1. Dataset design, grouping unit, and structural-support rule.

Item	Value / role
Unique simulator invocations	400 grouped invocations.
Condition-specific rows	1600 rows after four-condition expansion.
Conditions	`baseline_noecon`, `baseline_econ`, `with_influencer_noecon`, and `with_influencer_econ`.
Grouped-validation unit	Simulator invocation, defined by lattice size, seed, affordability margin, and vacancy fraction.
Rows per grouped fold	80 held-out invocations and 320 condition-specific rows.
Primary outcomes	`blocked_move_burden` and `final_affordability_blocked_share`.
No-economy support rule	Both primary blockage outcomes have support $\{0\}$ when `use_economic_constraint` $= 0$.
Economy-active support rule	Both primary blockage outcomes are bounded to $[0, 1]$ when `use_economic_constraint` $= 1$.
Branch-integrity check	Both primary outcomes are zero in all 800 no-economy rows.

Table 2. Grouped-CV structural-masking comparison across evaluated surrogate families.

Model family	Mode	Blocked RMSE	Afford. RMSE	Invalid (b,a)	Selection role
Ridge regression	Masked	0.0786 ± 0.0063	0.0987 ± 0.0073	0, 0	Simple linear baseline; structurally valid but less accurate.
Random Forest	Masked	0.0496 ± 0.0035	0.0748 ± 0.0040	0, 0	Strong non-neural comparison baseline.
ExtraTrees	Masked	0.0496 ± 0.0036	0.0748 ± 0.0041	0, 0	Tree-ensemble sensitivity baseline.
Gradient Boosting	Masked	0.0498 ± 0.0036	0.0751 ± 0.0044	0, 0	Boosted-tree sensitivity baseline.
MLP-8	Masked	$\mathbf{0.0494 \pm 0.0037}$	$\mathbf{0.0746 \pm 0.0042}$	0, 0	Primary model; 137 parameters.
MLP-16	Masked	0.0495 ± 0.0036	0.0748 ± 0.0041	0, 0	Neural sensitivity model; 401 parameters.
Two-stage MLP-8	Structural	0.0496 ± 0.0036	0.0747 ± 0.0042	0, 0	Active-regime structural comparison baseline.

Values are mean $\pm$ grouped-fold standard deviation. "Invalid (b,a)" reports structurally invalid no-economy predictions for blocked-mobility burden and final affordability-blocked share. The hard support mask makes the zero-violation result guaranteed after masking; the empirical question is whether useful economy-active prediction remains.

Table 3. Targeted diagnostic verification of selected response-map regions.

Item	Value	Interpretation
Selected response-map points	40	Sampled from low-, medium-, high-, and near-boundary or high-gradient predicted regions.
Direct simulator replicate rows	120/120	Three direct simulator seeds completed for each selected point.
Blocked-mobility burden MAE	0.0551	Mean absolute error against direct simulator observations.
Affordability-share MAE	0.0687	Mean absolute error against direct simulator observations.
Mean regime agreement by point	0.7417	Diagnostic agreement between predicted and observed blockage-regime labels.

This is targeted diagnostic verification of selected response-map regions. It strengthens the diagnostic interpretation of the response maps but does not constitute dense phase-boundary validation. Across the 120 replicate rows, the standard deviations of absolute error were 0.0583 for blocked-mobility burden and 0.0815 for final affordability-blocked share; regime agreement by selected point ranged from 0.0000 to 1.0000.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- scientificreportsstructuralsupportSUPPLEMENTUPLOAD.zip