Empirical calibration of full scale agent-based models of school choice: computational challenges and model validation

Keywords: agent-based modeling, likelihood-free inference, simulation-based inference, empirical calibration, school segregation

Extended Abstract

Despite a wealth of studies and counteracting policies, many educational systems still consist of substantial levels of school segregation along various lines, such as race, ethnicity, income levels, educational attainment and ability. This means that children with similar characteristics cluster together in different schools, which is widely acknowledged to reproduce, and even exacerbate, inequalities and result in unequal outcomes [2]. A recurring finding in existing school choice research is that interactions within and between the components of the system are important mechanisms underlying school segregation [8]. Paradoxically, often employed methodologies in this field fail to account for such interactions or their consequences. While households might state in interviews/surveys that they interact with others, these methods cannot quantify the consequences of these interactions on the macro-level. Whereas methods such as regression analysis do provide quantitative estimates, but often assume observations are identically and independently distributed (i.e. not interacting). Moreover, households are found to have heterogeneous preferences/constraints with respect to their characteristics (e.g. income, ethnicity), complicating the mechanisms even more.

However, important methodological progress has been made in recent years, for example by using Agent-Based Models (ABMs). ABMs are algorithmic models that consist of many, typically heterogeneous agents that interact (spatially, temporally, in social networks) and can respond to their time-varying environment all the while they pursue a certain objective [1]. Hence they allow for previously described heterogeneous components and their interactions explicitly. They can systematically link household behavior on the micro-level to their consequences for school segregation on the macro-level. Both for science in general [1] and school choice in particular [4], they are considered a very suitable tool to model the observed dynamics and interactions.

For school choice research however, the use of ABM has been limited. Theoretical work [10, 3], with choice processes inspired by empirical studies, finds indeed that interactions are an important component in the observed levels of school segregation. Although insightful, these studies have limited applicability to reality. Directly estimating the model parameters of ABMs using individual-level empirical data is not trivial, as for most ABM the likelihood cannot be written down explicitly nor sampled from. Ruling out commonly used techniques such as Maximum Likelihood Estimation (MLE) and Markov-Chain Monte-Carlo (MCMC) techniques for example. This is why existing empirically calibrated ABM of school choice [7, 11] have used other methodologies to estimate household preferences/constraints, before using these estimates in the ABM. However, this excludes the effect of interactions in the estimation process and using aggregate data, such as [5], misses potentially important heterogeneity and could introduce bias on top of that.

However, likelihood-free inference methods, such as Approximate Bayesian Computation (ABC) and techniques inspired by neural networks, such as Neural Posterior Estimation (NPE) have already successfully been utilized for parameter estimation in small ABM of socio-economic systems [9]. Especially NPE and its variant Neural Ratio Estimation (NRE) have shown promising results with a limited simulation budget [6]. This is important, because representing every household in the model, adding more heterogeneity and more factors in the school choice process, leads to a substantial increase in parameters to estimate. This, in turn, increases the simulation budget needed, as it complicates the search space for the calibration methods. Hence, it remains an open question if accurate parameter inference is still possible in these cases or what the potential limits are.

Therefore, this study extends the ABM of [5] and lays out a methodology to calibrate it directly on household school choice register data from Amsterdam primary schools using NPE. Although the ABM is tailored to a specific case, empirical calibration is a general problem for ABM. Hence, in our view, there are at least two additions to existing literature. Firstly, empirically calibrating full city-scale ABM is not trivial and computationally expensive. For ABM of social systems, this has only been shown to be feasible for small systems without much heterogeneity [6]. Due to privacy concerns the results of the actual data cannot be shown, but we demonstrate that, in our ABM that exhibits a substantial number of components and heterogeneity, NPE is successfully able to recover the true parameters (from synthetic data). Secondly, employing posterior-predictive checks, simulation-based calibration and global sensitivity analysis techniques, robustness and validation of the results is assessed.

References

[1] An, L., Grimm, V., Sullivan, A., Turner II, B., Malleson, N., Heppenstall, A., Vincenot, C., Robinson, D., Ye, X., Liu, J., et al.: Challenges, tasks, and opportunities in modeling agent-based complex systems. Ecological Modelling **457**, 109685 (2021)

[2] Boterman, W., Musterd, S., Pacchi, C., Ranci, C.: School segregation in contemporary cities: Socio-spatial dynamics, institutional context and urban outcomes. Urban Studies **56**(15), 3055–3073 (2019b)

[3] Dignum, E., Athieniti, E., Boterman, W., Flache, A., Lees, M.: Mechanisms for increased school segregation relative to residential segregation: A model-based analysis. Computers, Environment and Urban Systems **93**, 101772 (2022)

[4] Dignum, E., Boterman, W., Flache, A., Lees, M.: Modelling school choice as a complex system: understanding school segregation [working paper] (2022)

[5] Dignum, E., Boterman, W., Flache, A., Lees, M.: Unexpected school segregation? an empirically calibrated agent-based model of primary school choice in amsterdam [working paper] (2023)

[6] Dyer, J., Cannon, P., Farmer, J.D., Schmon, S.: Black-box bayesian inference for economic agent-based models. arXiv preprint arXiv:2202.00625 (2022)

[7] Mutgan, S.: Free to Choose? Studies of Opportunity Constraints and the Dynamics of School Segregation. Ph.D. thesis, Linköping University Electronic Press (2021)

[8] Perry, L.B., Rowe, E., Lubienski, C.: School segregation: theoretical insights and future directions (2022)

[9] Platt, D.: A comparison of economic agent-based model calibration methods. Journal of Economic Dynamics and Control **113**, 103859 (2020)

[10] Stoica, V.I., Flache, A.: From schelling to schools: A comparison of a model of residential segregation with a model of school segregation. Journal of Artificial Societies and Social Simulation **17**(1), 5 (2014)

[11] Ukanwa, K., Jones, A.C., Turner Jr, B.L.: School choice increases racial segregation even when parents do not care about race. Proceedings of the National Academy of Sciences **119**(35), e2117979119 (2022)