Reviewer 1

|  |  |
| --- | --- |
| **Reviewer Comments** | **Author responses** |
| The topic of the paper is agent-based modelling of policy diffusion between countries, which is a relevant topic for JASSS and important for the ABS community. In particular, this studies COVID-19 policy diffusion. | Thanks very much for your appreciative comment. |
| A very simple model is presented, based on only three aspects of a country: national income, degree of democracy, and geographical location. A country’s adoption of COVID-19 interventions, like lock-down, are based only on these three data about the country. Consequently, the predictive accuracy whether a particular country introduce lock-down or not at a particular point in time is not impressive. In fact, during the first half of the studied time period (March 1-15) the mean accuracy of the base model is about the same as guessing that no country has introduced lock-down (although more than 40% of the countries are in lock-down on March 15). | Thank you very much for your comment.  We believe our model’s accuracy is sufficient for the study at hand since   1. We did not aim to construct the perfect model but to have an interesting case study from the social sciences to combine with data assimilation. 2. And, for the first of its kind, we believe the accuracy of the model is sufficiently high to work with. In fact, predicting the exact sequence of when particular countries adopt a lockdown would be an extremely ambitious goal. 3. We believe our model is, in terms of accuracy, in line with many other diffusion models that achieve “stylized facts” in the sense that they reproduce general shapes of diffusion curves and elucidate mechanisms but hardly ever match empirical data one to one. To take just one example from the policy diffusion literature, Rapaport, Levi-Faur, and Miodownik (2009) elaborate on the diffusion mechanisms of central bank policies, but they do not match their diffusion curve and sequence to any empirical data. In that sense, our study, we believe, is substantially more ambitious already, and to some degree also more accurate since the ensemble actually matches the empirical data on aggregate and to an intermediate degree even by particular countries.   References:  Rapaport, O., Levi‐Faur, D., & Miodownik, D. (2009). The Puzzle of the Diffusion of Central‐Bank Independence Reforms: Insights from an Agent‐Based Simulation. *Policy Studies Journal*, *37*(4), 695-716. |
| That the base model is not good enough is one of the main weaknesses of the paper, as I see it. Why just include the three aspects in modelling the countries? Important but ignored aspects may include the country’s welfare system, its government (e.g. left-wing or right-wing), etc. There are also important international aspects that could be taken into account, like WHO recommendations. | Thanks very much for your comment. We believe the model is good enough for its purposes, which are to deliver a simple base model of international policy diffusion in the context of the pandemic, and to be a suitable case study for the application of data assimilation as outlined in detail above.  We included only these three characteristics to have a model that is as simple as possible but no simpler and to be in line with the policy diffusion literature. Policy diffusion literature shows that geographical and ideological proximity are important factors for countries’ governments to determine policy diffusion.  See, the work by Rapaport, Levi-Faur, and Miodownik (2009) on the adoption of central bank reforms, who capture geographical and ideological proximity in their “zone of influence variable”. Or compare the recent study by Linsenmeier, M., Mohommad, A., & Schwerhoff, G. (2023) on the diffusion of carbon pricing around the world.  Moreover, Polillo, S., & Guillén, M. F. (2005). discussed and showed that more specific macroeconomic and political indicators, such as the number or kind of political parties in the country, fail to explain the adoption of central bank reforms.  However, we agree with the reviewer that the WHO recommendation is a variable that likely played a role and can be considered in extensions of the model. We included a corresponding sentence in the discussion section  “  *It might be useful to extend the model with more circumstantial parameters relevant for the pandemic, such as World Health Organization (WHO) recommendations on virus counter measures, which in all likelihood also a had a substantial influence on countries' decisions.*  “  Lastly, we argue that countries of similar economic status often maintain strong economic ties and coordination such as for instance the club of the G7, the G20 or the OECD.  References  Polillo, S., & Guillén, M. F. (2005). Globalization pressures and the state: The worldwide spread of central bank independence. *American journal of sociology*, *110*(6), 1764-1802.  Linsenmeier, M., Mohommad, A., & Schwerhoff, G. (2023). Global benefits of the international diffusion of carbon pricing policies. *Nature Climate Change*, 1-6. |
| To improve the base model, data assimilation is used, i.e. data about the current situation are feed to the model continuously. Obviously, this improves the accuracy. But why doing it in this way rather than just rerun the simulation with the latest data as the starting point? Please explain the advantage of data assimilation compared to rerunning the simulations. | The advantage of data assimilation is that, usually, data and observations are uncertain too and to fully rely on the new observations would be as risky as to fully rely on a model prediction. Therefore, data assimilation is a set of algorithms that combines observations with model predictions. And therefore, with respect to the particle filter method we apply, the diversity of particles, i.e. the simulation ensemble, is important to attain a more accurate guess of the actual system state.  In our case, the observations may be regarded as 100% certain, however, the goal is to prepare methods for the *general case* in which data as well be uncertain. This could for instance occur in an extension of the model that considers subnational data and the diffusion of actual population response to policy measures, where often it is not really clear how much “in effect” policies actually where.  We added the following paragraph to our Outlook/Discussion section to clarify “*For instance one such natural "complexification" would be to model subnational regions and apply data assimilation not only to the diffusion of policy guidelines but also actual behaviour of people - which would be particularly interesting since the data is likely to be more uncertain. For instance, rural regions may behaved differently than urban ones in terms of actual behavioral changes. People perhaps still mixed socially in urban and rural areas, despite officially stringent guidelines, and sometimes the awareness of countermeasures and behavioural guidelines may not have been developed completely. For instance, in rural India, up to 20\% of people may have not been aware of behavioural guidelines to prevent spread \parencite{ali2023health}.”* |
| Regional differences within a country are ignored and the strictest measure of any region are assumed for the whole country. As you write, this is a major simplification and limitation. Why not use your model also within a country (letting the regions correspond to countries)? | We use data by Hale, T., Angrist, N., Goldszmidt, R. et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 5, 529–538 (2021). <https://doi.org/10.1038/s41562-021-01079-8>  This data has been further curated for consistency by ourworldindata.org and subnational data with international consistency is not available.  In advanced extensions of the model however we also suggest that it would be interesting to model and predict subnational regions, especially when the data becomes more uncertain (see comment and response above). |

Reviewer 2

|  |  |
| --- | --- |
| **Reviewer Comments** | **Author responses** |
| The submission has multiple facets. One of the facets is about studying international policy diffusion via agent-based models. The other facet of the work is about making a methodological contribution to agent-based modelling via empirical data assimilation. Both aspects are promising as valuable contributions. However, I have some comments and suggestions mainly related to clarifying these contributions. For the sake of clarity, I will focus on these two aspects separately as the submission feels even like the union of two independent articles. | Thank you very much for your positive synopsis. |
| While building up the discussion on the motivation, “lack of agent-based studies of …. the formation of lockdown policies” is raised as a notable gap in the literature. It is not clear what is the lost opportunity by not having an ABM in such a specific issue. The authors discuss potential difficulties about building such a model, but one expects to read about potential gains with ABM that address open questions in the domain that cannot be answered with established approaches. | Thank you very much for your comment. Agent-based modelling in general is a proven tool to study complex systems in which the interaction between agents are of high interest. Statistical/econometric approaches can investigate whether certain variables drive the adoption of lockdown policies but cannot assess whether the interaction between countries is a sufficient mechanism to generate the observed diffusion curve.  We added the following sentence to our introduction: *“A significant advantage an ABM has over previous econometric approaches is to emphasize the role of country-interaction and study whether this interaction is a sufficient mechanism for the emergence of diffusion.”* |
| A short sentence that plays a crucial role in the build of the submitted work constitutes a big IF: “…if one of the most important drivers of lockdown policy adoption is peer mimicry” Supporting this if statement with the current policy diffusion literature would contribute significantly to the validity of the modelling work presented. Another supporting evidence for such an if would be showing the generative sufficiency of the mimicry in the observed dynamics via the simulation model presented. However, model results also raise some concerns regarding the behavioral validity (an issue to be discussed below). Therefore, we are also lacking that supporting evidence. I think a stronger theoretical and/or empirical discussion is needed to support the assumption that is fundamental for the discussed modelling work; i.e. peer mimicry plays the central role in policy diffusion. | Thank you very much for your comment.  We are sorry that we have not been sufficiently clear. However, we believe we have described and discussed several works that show that peer mimicry is a driver of international policy diffusion.  In section 2.1 we had included the following paragraphs and works, and added the clarifications in red color.  *“Importantly for our purposes, some policy diffusion research has specifically focused on the COVID-19 emergency in early 2020 and the observed rapid policy diffusion.*  *\textcite{lundgren2020emergency} for example have investigated the diffusion of ``state of emergency (SOE) declarations''. They show that the declarations in early 2020 follow a typical diffusion curve but maxed out at roughly 50\% of all countries world-wide. They also ascertain the drivers of SOE declarations. According to their results regional clustering occurs. Countries opted for SOE, if neighboring countries did as well and the probability to adopt SOE depends on the degree of democracy and pandemic preparedness - supporting our model’s focus on peer mimicry.*  *As aforementioned, \textcite{Sebhatu2020} found that there are some internal drivers of lockdown policies (e.g. population density and level of democracy) but none of those alone suffices to explain the archetypal diffusion curve observed in March 2020. Instead they argue that peer mimicry must have driven the process. This hypothesis is corroborated by other studies. For example \textcite{mistur2020contagious} employ fixed effect models on panel data throughout 2020 and demonstrate that mimicry of geographical neighbors and political peers in addition to having a language in common are the principal drivers for countries to introduce, or abandon again, social distancing measures.”*  In section 2.1 We also discuss the general conceptualization of international policy diffusion as: *“...when government policy decisions in a given country are systematically conditioned by prior policy choices made in other countries... (Beth A. Simmons, Dobbin, and Garrett 2006).”* Which we believe further validates that countries look towards other countries when making their own decision.  Lastly, since we believe this evidence is strong enough to build a model based on peer mimicry, but we admit that the phrasing was confusing, we changed the sentence to: *“However, since, according to literature, one of the most important drivers of lockdown policy adoption is peer mimicry, a model of international lockdown policy may not need to consider the vast complexity and variety of national decision-making structures.”* |
| Related to the former point; the authors also propose the idea that lock-down decisions of the countries are independent of the actual covid case numbers. Coupled with the former assumption about the role of mimicry, these two assumptions are very important for the validity of the model, and in their current form they stand more like speculation (I didn’t use hypothesis as the authors do not have the intention of testing these claims in their work) | Thanks for the comment. But the literature clearly shows that the population-normalized case numbers and death rates do not predict the homogenous adoption of lockdown policies (see for instance section “control variables” in Sebhatu, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020)). As per peer mimicry we discussed above already. Hence, we strongly believe these two assumptions are reasonable.  References  Sebhatu, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences*, *117*(35), 21201-21208. |
| The authors introduce the technical foundations of the model in section 3.1. Some of the points need clarification and further discussion in my opinion. I assume the distance between a pair of countries play an important role in the model dynamics. It is assumed that the distance is an equal-weighted average of there indicators; income similarity, political similarity and geographical proximity. Conceptually the notion of distance is clear, but it would be helpful if the authors can present a set of examples about the resulting proximity between countries. Who are the closest countries to, for example Colombia, Mexico, Turkey, Russia, UK and North Korea according to this metric? As these closest country sets are claimed to be the sets that are influential in policy adoption decisions, knowing a couple of sample sets may help the reader to comprehend what is going on in the model. | Thanks very much for the helpful idea. We have now computed a 164x164 matrix for all countries in the model that stores the distances between all countries with each other and also share it on our corresponding open-access data repository. We also have included several of the examples the reviewer suggests in the text and moreover included a histogram of the distribution of all distances in the supplementary material.  The text added in section 3.1 is:  *“For better illustration what this metric produces, we want to give a few examples. With respect to the United Kingdom, for instance, Austria is at a distance of 0.05, Germany of 0.04, Argentina at a distance of 0.31 and Somalia of 0.5. With respect to New Zealand, for example, many countries are at a distance above 0.5 or even 0.6 just because of its distant position on the globe. Albania has the lowest average distance to all countries (0.2) which is explained through a geographically centric position in Eastern Europe, where plenty of countries in Europe, Asia and even Africa are relatively nearby, as well as intermediate rankings in terms of GDP per capita and the Democracy index. Indeed Luxembourg has the highest average distance to all countries, with \~0.5, explained through an unusually high GDP per capita as well as a strong democracy, hence the economic and political distance to many countries is high. Of course Luxembourg is still relatively close to its European neighbors for instance \~0.2 to the UK, Germany and so forth.”* |
| Every country checks 18 other countries in the model. This number of 18 seems too magical, too specific. The authors mention that it is determined during calibration. I would suggest adding some evidence and discussion on the sensitivity of model output to this peer set size parameter. 18 gives the best results, that is ok. But what should be expected if this number is 20 or 10? | Thanks very much for the comment. We agree with the reviewer. We had indeed added already a sensitivity analysis in a supplementary material.  Please check the supplementary section 2.3. In general, when decreasing the peer group size, the diffusion curve is steeper, meaning the country X is more sensitive to adoption based only on a few peers, and increasing the peer group size flattens the diffusion curve. |
| The peer-mimicry process sounds like it is biased towards adoption. Each country checks similar countries already in lock-down, and if their number amounts to a certain level, the country also goe to a lock down. While doing so, the country ignores the other similar countries no in lockdown. For example, a country X may have 50 similar countries. If 10 of them in lock-down, it may cause X to go into lockdown despite 40 similar countries not in lockdown. Is this understood correct? IF not, the text may require some revision for clarity. Otherwise, the structural bias needs further elaboration and theoretical grounding, perhaps. | Thanks for your comment. The understanding of the reviewer is correct. We would argue however that, since the model produces plausible diffusion pattern, this “bias” does not constitute a larger problem. There are several reasons for this:   1. In the given crisis situation, it makes sense for countries to pay more attention to the other countries who a take a new defense-policy against the virus. Business-as-usual is not as great as a stimulus as a radical emergency policy. 2. There are several parameters that control the speed and extent of diffusion, and even are able to entirely stop the diffusion process (compare supplementary material section 2). So, the model setup chosen, including the parameter values that enable plausible diffusion patterns, is not so much a bias but rather just standard calibration. With distinct parameter values, the diffusion might not happen. 3. Yes, the diffusion direction is one-directional in our current version of the model, but this is similar to the model on central bank policies by Rapaport, O., Levi‐Faur, D., & Miodownik, D. (2009) in which countries also cannot reverse their decision. An extended model that looks at longer time horizons might correct for this.   Rapaport, O., Levi‐Faur, D., & Miodownik, D. (2009). The Puzzle of the Diffusion of Central‐Bank Independence Reforms: Insights from an Agent‐Based Simulation. *Policy Studies Journal*, *37*(4), 695-716. |
| Equation 4 and 5 may need revision. They are presented as the a-social adoption rule. However, the presented information does not clearly describe a rule-based decision. E4 summarizes a fact, where if X is a uniform random, than probability of X being less than a continuous number k in (0,1) interval will be k. This is no rule. Authors may revise the way it is communicated. | Thanks for your comment.  We have clarified equation 4 by stating that it is the Probability of adoption a lockdown by initiative and reworded the text accordingly to:  “*In precise terms, the probability for a-social adoption then is:* ” |
| At the end of section 3.2 the authors clarify that they consider school closures as lockdown. For the sake of clarity, it would be better if this can be told from the beginning. I understand that diffusion of school closures does not sound as good as diffusion of lockdowns, but aligning the scope with the title and text will be better in my opinion. | Thanks for your comment.  We have clarified this in the introduction section 2nd paragraph: “*On the first of March 2020 only around 8 \% of countries had implemented such stringent measures, by the end of March 2020 more than 90\% had implemented school closures and cancellation of all public events (with several more policies such as workplace closures etc. at intermediate degree*)”.  And in the 3rd last paragraph of the introduction with:  *“We focus on March 2020 and initialize the model based on real-world data from 1st March, using a measure of school closures as a reasonable proxy for overall lockdowns as elaborated on in section 3.2.”*  However, the measure chosen does capture more than just school closures. As can be seen in Figure 2b public event-cancellation and school-place closure correlate near perfectly, which alone already constituted a large part of what was experienced as lockdown. Moreover, as we elaborate in section based on the data: “*Public event cancellations correlates very closely with school closures. Other measures do not follow school closures that closely but still align to a high degree. Partly this is because we only display the highest stringency level and for some metrics intermediate steps are more significant. For instance workplace closure correlates more closely with school closures and event cancellations if the second highest level “required for some” is also considered. In sum, all policies correlate substantially over time (Pearson’s ρ > 0.9 for all policies).”* |
| The model results are raising serious behavioral validity concerns which also make one question the explanatory sufficiency of the mechanisms presented in the article. By design, the diffusion will start from 0 level. Considering the a-social adoption and mimicry processes, the diffusion is expected to converge to 100% level inevitable. I guess there is no scenario where the diffusion process halts at a significantly lower level (eg. 65%). Therefore, we expect agreement across model replications and empirical data in the beginning and at the end. The critical phase that would show the generative sufficiency of the model would be the middle-zone where the inflection of the diffusion process is observed.  However, the model scores poorly in this phase. Apart from that, the error reported at the micro level is even poorer. So even when we see a good fit at the aggregate level, it is mainly due to wrong reason as the error at the individual country level is significant. To put it differently, a deterministic compartmental (Potential Adopters - Adopters) might demonstrate a much better behavioral validity at the aggregate level. Then the presented model could have been defended with its micro level dynamics, but since reported results are poor on that level the potential superiority of such an ABM does not realize in my opinion. In the light of these observations, I have doubts about the contribution of the presented model to the domain of policy diffusion due to its low behavioral validity. The authors may wish to discuss the added value of the presented model perhaps by demonstrating some key insights that could have been impossible to develop without such a model, or even with macro level models. | Thanks for your comment.  We believe the model is good enough for its purposes, which are to deliver a simple base model of international policy diffusion in the context of the pandemic, and to be a suitable case study for the application of data assimilation as outlined in detail above.  Moreover, there are scenarios that reach and/or halt (at) a much lower level of diffusion. See sensitivity analysis of parameters in the supplementary material section 2.  Furthermore, the model starts from entirely data-grounded initial conditions – 13 countries had adopted a lockdown already on 1st march 2020 – the day our model scenarios begin.  The reviewer asserts the model scores “extremely poorly” and has ”serious validity concerns” especially on the micro level. While we acknowledge the limitations of the model and agree that improvements are possible and necessary future research directions, we see the advantages of the model as follows:   1. ABM diffusion models are mostly just use to produce stylize-fact and illustrate general principles of diffusion, hardly ever to produce agent-specific diffusion sequences (see point 3 below). Our model does produce plausible shape diffusion curves based on a peer mimicry mechanism between countries. 2. Being the first of its kind, despite limitations, the model adds to the literature by illustrating that such an interaction-based mimicry mechanism is in principle sufficient to explain the policy diffusion during the pandemic – a fact that econometric/statistical methods can hint at, but not demonstrate. We added the following clarification in the introduction section paragraph 6: “*A significant advantage an ABM has over previous econometric approaches is to emphasize the role of country-interaction and the ability to study whether this interaction is a sufficient mechanism for the emergence of diffusion.”* 3. Predicting sequences of diffusion across agents with 100% accuracy , who exactly adopts 1st, 2nd ,3rd, and so forth, would be an extremely ambitious goal and is usually not expected from ABMs. Running a 100% data-grounded agent-based model, with agents representing specific tractable real-world entities, is already a substantial step beyond most agent-based models. For example, as a comparison, the study Rapaport, O., Levi‐Faur, D., & Miodownik, D. (2009) intends to explain the diffusion of central banks policies over countries but achieves “only” a generic diffusion curve pattern among hypothetical countries without any explicitly judgement based on the data and sequence of who adopted when. At any time- point, our model predicts more than 50% of countries in their correct lockdown state (yes or no), while not perfect, usually ABMs do not have this kind of rigorous judgement against agent-specific behaviour and hence we believe our model is more than sufficient for this study. |
| Regarding particle filter:   As our age is dominated by data, methodological tools/approaches that facilitate the use of (high-volume) empirical data in simulation models is a very promising line of research with great potential. In that respect, the particle filter approach sounds interesting. However, it was difficult for mie to comprehend the details of the filtering and resampling process. The true contribution of the submission in this regard would be easier to grasp if the method could be elaborated a bit more. | Thank you very much for the positive comment. |
| For example, as far as I understand each model replication of a particular model instance is a particle. In that respect, what does “re-sampling” somewhere in the middle of the simulation time horizon mean? Same question holds for “projecting forward”. | Thanks for your comment. We have now clarified parts of section 3.3 which describes the particle filter including the re-sampling procedure. We also updated Figure 3 to clarify. The corresponding text reads now as follows:  *“*  *After every reweighting procedure, a resampling of particles is undertaken to optimise the estimation of the system state. Here, Sequential Importance Resampling is used~\parencite{doucet\_sequential\_2000}. The resampling procedure updates the model ensemble at a specified interval, let us say every fifth time step for example, by discarding the worst performing model runs and replicating the best. During this procedure, the weights are cumulatively counted, so they constitute a cumulative distribution function (CDF). This distribution of weights is compared against a uniformly random partition of the interval [0,1], constituting a uniformly random CDF. Then \(N\_p\) points along the uniform distribution are selected, exactly at the mean step size of \(1/N\_p\) and compared against the CDF of the weights at that particular point. For example, if there are 10 particles then the uniform CDF is evaluated at \(x = 0.1\), \(x = 0.2\) and so forth. Let us say at \(x = 0.1\) the uniform yields exactly \(y\_u = 0.1\) and the weight distribution yields \(y\_w = 0.2\). The uniform partition therefore makes a smaller step than the weight distribution. Then the weight of the particle is large, its error small and correspondingly the particle should be resampled. If on the other hand \(y\_w = 0.05\) then the particle weight is less than the expected uniform average (which is 0.1) and thus the particle is discarded from the future particle population. Overall by conducting this procedure until 100\% of the uniform distribution are reached, it is very likely that particles with small weights are discarded because the `room' they make up in the cumulative weight distribution is very small. In analogy, this procedure might be considered a Roulette in which the ball has exactly uniform probability to end up anywhere on the Roulette, but the Roulette regions are not of uniform size. Clearly then, the probability is higher that the ball lands up on one of the larger regions. The principle applied here is the same.*  *Figure~\ref{Figure 3: Particle Filter steps} depicts the particle procedure and its iterative nature. A certain number of particles are projected forward in time and then considering new observations, a new particle population is resampled and again this particle population is projected forward.”* |
| As the proposed model has an adaptive nature, its trajectory is revised. In that respect, the basis for error calculation is not clear to me. At t=0, the model is ran n time, which give n estimates for output level at t=100, for example. Then, the approach includes a re-sampling and filtering at t=5. After that we will have a set of altered trajectories, and revised estimates for t=100. This continues all along. If this is a correct depiction of the process, how do we calculate the estimation error for t=100? If it is based on the most recent revision, it means that the error for t=100 is based on the filtered and resampled output at t=95, with an estimation horizon of 5 time units. In the conventional approach, we have a 100 period estimation horizon. If this is the case, comparing these two error terms would be misleading. If the process is misunderstood by me, that should serve as a good indication of lack of clarity in the text. | Thanks very much for your comment.  In general, to clarify, our model has a maximum number of 31 time steps (only the period of March 2020).  We believe we have clarified how the resampling and filtering procedure works in the comment above.  On the appropriateness of the error calculation: Yes, it is correct that every k-th steps, in our medium-case that is every five steps, the particle population is revised/updated by discarding the worst and replicating the best particles (see section 3.3 for details). It is standard procedure to apply the same model-to-data error metric at any time-step in the data assimilation literature. Compare for instance Malleson, N., Minors, K., Kieu, L. M., Ward, J. A., West, A. A., & Heppenstall, A. (2019) who also consistently apply the same error metric. In fact, it is necessary so that the particle weights are computed in a consistent manner throughout. We hope that clarifies your query.  References  Malleson, N., Minors, K., Kieu, L. M., Ward, J. A., West, A. A., & Heppenstall, A. (2019). Simulating crowds in real time with agent-based modelling and a particle filter. *arXiv preprint arXiv:1909.09397*. |