RUHR
ECONOMIC PAPERS

Tom Bauermann
Michael Roos
Frederik Schaff

POSA: Policy Implementation Sensitivity Analysis

RUB RGS econ

#854

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #854

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-990-2

Ruhr Economic Papers #854

Tom Bauermann, Michael Roos, and Frederik Schaff

POSA: Policy Implementation Sensitivity Analysis

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788990
ISSN 1864-4872 (online)
ISBN 978-3-86788-990-2

Tom Bauermann, Michael Roos, and Frederik Schaff[1]

POSA: Policy Implementation Sensitivity Analysis

Abstract

Agent-based computational economics (ACE) is gaining interest in macroeconomic research. Agent-based models (ABM) are increasingly able to replicate micro- and macroeconomic stylised facts and to extend the knowledge about real-world economic systems. These advances allow ABM to become a valuable and more frequently used tool for policy analysis in academia and economic practice. However, ACE is a rather complex approach to already complex investigations like policy analyses, i.e. the analyses on how a variety of policy measures affects the (model) economy, which makes policy analyses in ABM prone to critique. The following research paper addresses these problems. We have developed a procedure for policy experiments in ACE which helps to conceptualise and conduct policy experiments in macroeconomic ABM efficiently. The procedure makes policy implementation decisions and their consequences transparent by conducting what we term the policy implementation sensitivity analysis (POSA). The application of the procedure produces graphical and/or numerical reports that should be included in the appendix of the original research paper in order to increase the credibility of the research, similar to proofs and protocols in analytical and empirical research.

JEL-Code: C63, E6, B4

Keywords: Agent-based macroeconomics; policy experiments; sensitivity analyses

June 2020

1 Tom Bauermann, RUB and RGS Econ; Michael Roos, RUB; Frederik Schaff, RUB. – All correspondence to: Tom Bauermann, RGS Econ, Hohenzollernstr. 1/3, 45128 Essen, Germany, e-mail: tom.bauermann@ruhr-uni-bochum.de

1 Introduction

Agent-based computational economics (ACE) is used more and more often in macroeconomic research (Richiardi, 2017). Agent-based models (ABM) are increasingly able to replicate stylized facts on the micro- and macroeconomic level(e.g., Dosi et al. (2006) and Caiani et al. (2016)) and to extend knowledge about real-world systems due to representing the heterogeneity and interdependence of real-world agents within the model (LeBaron and Tesfatsion, 2008). These advances allow ABM to become a valuable and more frequently used tool for policy analysis in academia (i.e., (Tesfatsion, 2001),(Dawid and Fagiolo, 2008) and (Dawid et al., 2018)). Such models are used to show how the economy is affected by the introduction of a certain policy, for example a new labor market regime (Dosi et al., 2018) or a new tax (Cincotti et al., 2012). Furthermore, ACE can be used for policy analyses in economic practice (Baptista et al., 2016).

However, ACE is a rather complex approach to already complex investigations like policy analyses. As in political reality, policy experiments often consider many simultaneous changes in the model, i.e. a bundle of changes in the parameters or behavioral rules. Authors have to reduce the complexity of policy analyses when presenting the results in a research paper and sometimes also when conducting the analyses. Accordingly, policy scenarios are defined and analyzed in policy experiments as a set of unambiguously defined policy parameter levels and their changes. The downside of this general practice is that it becomes difficult for the readers, including academic reviewers and political practitioners, to judge how the concrete implementation of the policy was chosen and whether slight variations in the implementation or the underlying model would change the results of the policy experiment. Thus, the complexity of ABM and the necessary strategies to cope with it make it prone to criticisms like the *lucky parameter configuration* (Dosi et al., 2018), which states that simulation output is the result of a "lucky configuration" of the underlying parameters, or at least very sensitive to the implementation choices which are often opaque to the reader/reviewer (Edmonds and Hales, 2003).

In the following, we develop a procedure for policy experiments in macroeconomic ACE. The procedure makes decisions on the implementation of the respective policy in a research paper, as well as the robustness of the implemented policy and the consequences of alternative policy scenarios (i.e. alternative parameter changes), transparent by conducting what we term the *policy implementation sensitivity analysis* (POSA). The procedure can be rigorously carried out and with reasonable effort. The application of the procedure produces graphical and/or numerical reports that should be included in the appendix of the original research paper in order to increase the credibility of the research, similar to proofs and protocols in more traditional analytical and empirical research. It should be noted that the aim of this paper differs from the literature on design protocols (e.g., ODD (Grimm et al. (2010) and Laatabi et al. (2018)) or the Dahlem protocol for economic research (Wolf et al. (2013)), sensitivity analyses (e.g., Campolongo et al. (2011)) and design of experiments (e.g., Montgomery (2017) and Lorscheid et al. (2012)). We neither want to provide guidelines on how to present and describe ABM nor provide tools to get a better understanding of the underlying model. Our aim with POSA is to balance the two objectives. On the one hand, it should make the underlying implementation decisions for the policy (experiment) and their impact on the results of the experiment transparent. On the other, it shows how to reduce the complexity of policy experiments in large-scale macroeconomic agent-based models to keep efforts manageable in providing transparency. In other words, POSA does not reduce the complexity of macroeconomic ABM, but addresses the complexity of the policy analyses. POSA seems to be important, since ACE is attracting a lot of interest not only from academic macroeconomists, but also from the public authorities (Turrell (2016) and Laliotis et al. (2019)). Further, approaches like POSA contribute to debates about the uncertainty in policy analyses (Steinmann et al., 2020). To the best of our knowledge, similar procedures have not yet been proposed in the literature. The remainder of this paper is organized as follows. Section 2 of this paper describes goals and the underlying concepts of policy analyses, as well as POSA. Section 3 presents the individual steps for the procedure. In section 4, we provide a short summary. Finally, in the appendix of this paper (section A), the POSA procedure is illustrated using an example. We provide a practical illustration by using it on the model of Bauermann (2020) (section A.1).

2 Goals of the policy analysis

In what follows, we define a *policy* as a specific real-world policy (policy measure), e.g. taken by the government or a changing behavior by another class of agents, that can be implemented with different levels of strength. Each policy is related to a set of *policy parameters* which are supposed to affect the *policy target*. The literal meaning of a strong (weak) implementation is that the set of policy parameters as a whole is strongly (weakly) tuned in the direction that is expected to favour the policy target. Here, *expected* refers to the expectations of the initiators of the policy, e.g. by political decision takers, not necessarily the actual direction of the policy target. The number of the parameters and their individual implementation strength make the policy analysis complex. As will be shown later in details, a major element in reducing the complexity is to fix the mapping of the individual policy parameters on the scale from a weak to a strong implementation. In short, a *policy* is a specific set of *policy* *parameters* with a *fixed mapping* from weak to strong implementation. A policy scenario, or *scenario*, is a specific policy with a specific implementation strength. The goal of the policy analysis is to analyze (potential) real-world policy scenarios[1] by simulating them in the macroeconomic ABM. We can identify two main tasks that are related to the implementation of a real-world policy in an ABM. First, instead of finding the optimal policy, researchers can present and analyze the spectrum of potential implementation strength levels. Structured simulation experiments need to be conducted that allow the identification of the *correct* implementation of the real-world policy *without* adjusting the implementation to some kind of desired outcome. Second, researchers select from this spectrum of potential implementation strength levels for later in-depth analysis in the paper and provide criteria for the selection. The decisions taken to identify the *correct* implementation of a policy need to follow scientific criteria, and this needs to be conveyed to the audience, i.e. other scientists and policy makers.

The aim of POSA is to provide a convenient, yet scientifically rigorous procedure to deal with these two tasks when the policy is implemented in an ABM. The following sections will explain the POSA procedure in detail. An additional third task related to the policy implementation is to test and guarantee the robustness of the POSA procedure when we account for the uncertainty that is present in the model already. We will briefly discuss this task in the final subsection 3.5.

3 POSA: policy implementation sensitivity analysis

In the following sections, we will lay out the policy implementation sensitivity analysis (POSA). The individual subsections represent the steps of the POSA procedure. Figure 3 on page 14 provides an overview over the procedure.

3.1 Prerequisites and preliminary notes (preparatory step)

We assume the agent-based macro model to be fully specified, analyzed (e.g. sensitivity analyses) and calibrated to the baseline parameter setting, i.e. before the above-mentioned policy is implemented/introduced. Further, the real-world policy that needs to be implemented is sufficiently complex, i.e. it is a portfolio of various (singular) policy parameter changes. A single policy parameter can be a single continuous parameter, for example, the (level of) unemployment benefits or the tax rate. It can also be a parameterized set of differ-

[1] We refrain from calling them predictions, as we consider ABM to be tools to explore potential scenarios without putting forth probabilities in the traditional sense of predictions.

ent categorial mechanisms, for example, switching from a tax regime to quantity regulation in foreign trade, or switching between various investment strategies on the asset market (Chakraborti et al., 2011). The term mechanism also includes behavioral rules (Dawid and Delli Gatti, 2018).

If there are only very few parameters, traditional factor variation can be used conceptually similar to comparative statics in general equilibrium models despite the fact that a descriptive approach is taken (see e.g. (Schaff, 2016, part 3)). As a rule of thumb, there should be more than three policy parameters present and the scale of measurement of these parameters should be at least ordinal, but *not* categorial. With up to three parameters, the complete space of potential policy implementations can easily be visualized in a single graph (or set of graphs, differing for the categorial variable) and thus analyzed and evaluated without the need to further reduce the complexity. If a policy can be implemented with different mechanisms, which implies a categorial "switch" variable like, for example, switching from a tax regime to a quantity regulation in foreign trade, each mechanism (or combination of mechanisms) is a variation of the policy and should be treated as a potential policy on its own. The POSA procedure should be applied to each such variation.

The main idea of the POSA procedure is to define a set of policies (it can also be only one policy) whereby each policy can be varied on a scale from a *weak* to a *strong* implementation according to external sources. The authors of the policy experiment need to be able to define a weak and strong implementation value for each policy parameter, as well as the way it changes continuously from a weak to a strong implementation. The policy needs to reflect the description of the policy in the real world, i.e. a party program (e.g., a proposed change in the social security system) or regulations that are in place but not yet triggered (e.g., (automatic) stabilization mechanisms, which are activated when a recession occurs).

Before the POSA procedure can be implemented, we have to define the criteria on which the *correct* implementation strength (for later in-depth analyses) is selected. The criteria must be specified externally and laid out prior to the analysis. By defining the criteria prior to the analysis, based on some external criteria, potential criticism aimed at the selection of the scenarios, i.e. aligning the simulation, can be addressed. ABM research is in this respect not different from empirical research. If the goal is to explore the consequences of a hypothetical policy in the theoretical model, the data gathering process has to be defined prior to the act of data gathering, which, in case of an ABM, is running the actual simulations with the specified, i.e. calibrated/estimated, model. The problem is that the policy implementation itself needs to undergo some kind of calibration process itself. Thus, there is the danger (or appeal) of adjusting the policy implementation in such a way that a desired outcome will show up. For example, *counterintuitive* results are often desired because they imply a high relevance for the research conducted. Some potential criteria for selecting the strength of the policy implementation are:

Descriptive implementation strength

Select the point on the implementation strength scale, i.e. from a weak to a strong implementation that is close to the real, i.e. planned, policy.

Optimistic implementation strength

Select the point on the implementation strength scale that is expected to yield the most favourable results with respect to the policy target.

Pessimistic implementation strength

Select the point on the implementation strength scale that is expected to yield the least favourable results with respect to the policy target.

The comparative effectiveness of the set of policy scenarios, i.e. the effect of a single policy at different levels of the implementation strength on the (model) economy, needs to be evaluated with output measure y, which can also be an index measure composed of several measures. More precisely, the specific output measure y can be identical to the (above-mentioned) policy target and additional measures if the potential side effects of a policy matter.

Further, the researchers have to decide at which point in (model) time the policy implementation starts and when the measurement of y ends. The time when the measurement y is taken reflects the type of simulation model with respect to its interpretation of time. In general, two kinds of simulation can be distinguished: terminating and steady-state (non-terminating) simulations (Alexopoulos, 2006; Law, 1980, 2015). First, in the case of a terminating simulation, the quantities of interest are defined relative to the interval of simulated time $[0, T_E]$, with T_E being the time that a specified event E occurs, a possibly degenerate random variable. Second, in the case of a steady-state simulation, the quantities of interest are defined as limits, as the length of the simulation goes to infinity. Since there is no natural event to terminate the simulation, the length of the simulation is made large enough to get "good" estimates of the variables of interest (Law, 1980), i.e. the output measures.

The objective of a steady-state simulation is the long-run/steady-state behavior of the system. The measurement should reflect this (stochastic) state and needs to be conducted after the burn-in phase, considering a sufficiently long observation period that the stationary dynamics are covered in an unbiased way (Law, 2015). An example would be a change in the long-run unemployment rate in the model economy. If the model can be considered a non-terminating simulation, i.e. the interest lies within the limit (stationary) distribution, it has to be guaranteed that the limit state is actually captured by the measurement period. The objective of a terminating simulation is not to study the long-run behavior of a system, but the system at a specific event. This implies that simulation time is comparable to real time, although not necessarily in an absolute manner. An example would be the time that is necessary to recover from a recession. In this case, when the system is in a steady state, the interval in which the measurement is taken, e.g. the first year after the policy is implemented, and the aggregation method, e.g. the maximum or the mean, need to be specified.

3.2 Definition of scenarios

Once the preliminary steps (section 3.1) are completed, the real world policy description needs to be translated into the model language as it will be shown in the following steps, before the policy implementation sensitivity analysis (section 3.3) is done:

Policy factors/parameters

The policy factors, i.e. the model parameters that are manipulated by the policy, are defined as a subset of the total number of parameters of the model. Let us call the subset of policy factors $\mathbf{x} = \{x_i, \ldots, x_n\}$ with a total of n factors.

Individual implementation strength

For each policy factor, the weakest *realistic* implementation strength $\underline{x}_i$ and the strongest *realistic* implementation strength $\overline{x}_i$ are defined. A strong (weak) implementation strength is associated with the expectation of a strong (only weak) effect in the desired direction of the policy target. The modeler also needs to decide on extreme values for each policy factor, i.e. extremely weak $\underline{\underline{x_i}}$ and extremely strong $\overline{\overline{x_i}}$. $\underline{x}_i$ and $\overline{x}_i$ are associated with the underlying case (for example a specific country), i.e. there might be constraints to go above/ fall below these values. $\underline{\underline{x_i}}$ and $\overline{\overline{x_i}}$ are extreme values which are possible in general but unrealistic for the underlying case. It should be noted (again) at this point that a weak or strong policy is associated with the *expectation* underlying the real world policy, for example a party program, not necessarily the *actual* (degree of) attainment of the policy target. The *actual* (degree of) attainment of the policy target is, in the model and in the real economy, *ex ante* unknown, which is also the reason for the policy analysis. It should be further noted that the gradient of x_i can be positive or negative depending on the expected relationship between the policy parameter and the policy target.[2]

[2] For example, the reader should assume a positive relationship between unemployment and unemployment benefits, and a negative one between unemployment and search efforts as in Pissarides (2000). A

Global implementation strength

We introduce the global parameter $s \in [0,1]$ to characterize the implementation strength from the theoretically weakest ($\underline{\underline{s}} = 0$) to the theoretically strongest ($\overline{\overline{s}} = 1$) policy implementation strength. It serves as a mapping for the individual parameter strength level to a (global) policy implementation strength. We divide this interval into three line-segments, $[\underline{\underline{s}}, \underline{s}), [\underline{s}, \overline{s}], (\overline{s}, \overline{\overline{s}}]$ with an arbitrary number $K = 21$ of evenly spaced design points such that $\underline{\underline{s}} = 0,\ \underline{s} = 0.25,\ \overline{s} = 0.75,\ \overline{\overline{s}} = 1.0$. The left interval, i.e. $[\underline{\underline{s}}, \underline{s})$, and right interval, i.e. $(\overline{s}, \overline{\overline{s}}]$, are attributed to only theoretically relevant extreme levels for the individual policy factors, whereas the main focus is on the behavior in the interim interval, $[\underline{s}, \overline{s}]$. $\underline{\underline{s}}, \underline{s}, \overline{s}, \overline{\overline{s}}$ are four of the mentioned 21 design points.

Individual strength mapping

For each policy factor x_i, we define a mapping $x_i(s) = x_i \to s$ such that $x_i(\underline{s}) = \underline{x}_i$ and $x_i(\overline{s}) = \overline{x}_i$, and $x_i(\underline{\underline{s}}) = \underline{\underline{x_i}}$ and $x_i(\overline{\overline{s}}) = \overline{\overline{x_i}}$, and a monotonous (discretized) mapping from the weakest to the strongest implementation. The discretization allows us to consider ordinal variables together with discrete and continuous variables. The exact shape of this mapping has to be decided by the modeler with the specific policy in mind, but also considering the resulting global mapping. The reader should note that we recognize that complex interactions between the policy parameters, and between the policy parameters and the policy target (other than monotonous) are possible. A typical policy does not consider this relationship, however. The complex interactions are rather part of the sensitivity analyses to explore the model before policy analyses are carried out.

Border case mapping

The selected criteria, including the *optimistic*, s_O, and *pessimistic*, s_P, implementations (section 3.1) should at best also lie within the interim continuum from $\underline{x}_i$ to $\overline{x}_i$ and not at the border, that is $\underline{s} \ll s \ll \overline{s}\ \forall s \in \{s_\mathrm{O}, s_\mathrm{P}\}$ is desired. The same is true for potential other criteria not discussed. However, this cannot be known in advance and the selection of the border cases for the individual policy factors, i.e. $\underline{\underline{x_i}}$ and $\overline{\overline{x_i}}$, should be based on the data alone and fixed prior to the analysis.[3] Therefore it is important to also cover unrealistic policy options by extrapolating the single policy factors towards their perspective theoretical extreme positions at the weak and the strong spectrum.

policy targeting a reduction in unemployment would assign the parameter *unemployment benefit* a negative gradient and the parameter *search activity* a positive gradient.

[3] If all the theoretical extremes are the same as the realistic borders, treating border cases can be omitted. In this case, it might be more descriptive to refine $\underline{s} = 0$ and $\overline{s} = 1$.

Global strength mapping

The combination of the individual policy strength mappings for all x_i results in a global matrix $\mathbf{X}(i, s)$, where each row represents a single policy factor x_i and each column represents the global implementation strength s. The individual strength mappings may need some refinement as to allow that all strength levels of interest, e.g. the *descriptive* implementation strength level ($s = s_D$) from actual policy, is represented.

Figure 1 provides an example with five parameters. Four parameters are continuous and one is discrete. The parameters are already mapped to the global implementation strength ($x_i(s) = x_i \to s$). The rectangular represents the practically relevant implementation levels, i.e. what was called realistic levels ($\underline{x_i}$ to $\overline{x_i}$). The areas to the left and right represent the theoretical limits for the sensitivity analysis ($\underline{\underline{x_i}}$ at strength level 0.0 and $\overline{\overline{x_i}}$ at strength level 1.0). The five policy parameters at one strength level (e.g. $x_A = ... = x_E = 0.2$) mark one policy strength level (*global implementation strength mapping*), i.e. a specific policy scenario. As noted in section 3.1, similar steps can be carried out for behavioral rules.

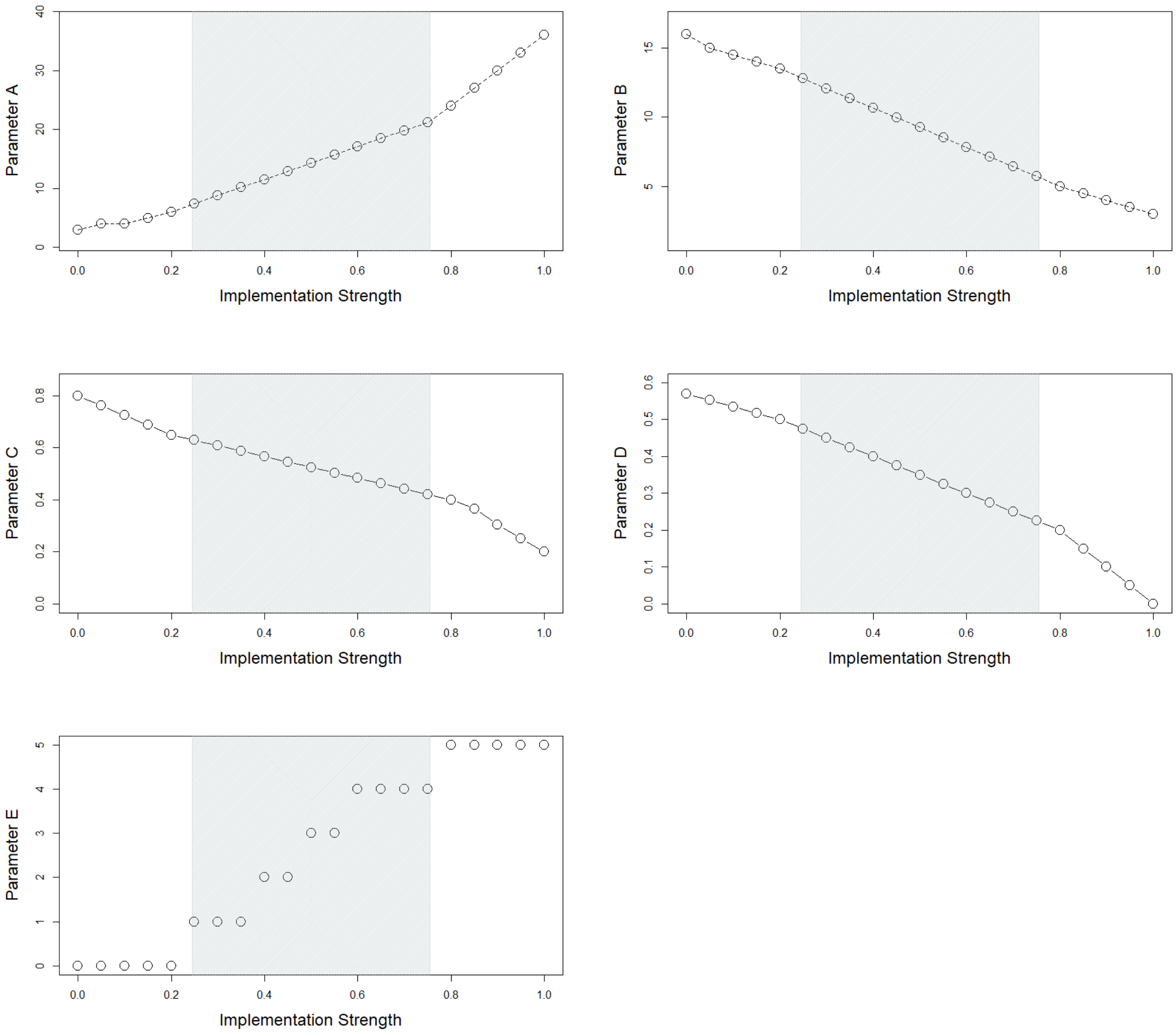
3.3 Conducting the implementation sensitivity analyses

We can now analyze the effect of the implementation strength of the policy on the output measure y for each policy mapping $y \overset{s}{\leftarrow} \mathbf{X}$. We call this effect the **implementation sensitivity**, as it quantifies the effect that the specific choice of strength has on output measure y given the mapping implied by the matrix $\boldsymbol{X}$. The implementation sensitivity analysis relies on the following setting:

- Sections 3.1 and 3.2 defined the setup to take out the simulations for the implementation sensitivity analysis, i.e. the simulations of the policy at various strength levels. We selected a fixed baseline parameter setting (i.e. the calibrated model), the output variable and type of simulation. The number of factor combinations is equivalent to the number of different strength levels of policy s, which we call parameter K (section 3.2). The policy parameters at one strength level define one policy scenario.

- Now, we select a suitable Monte Carlo sample size m for each strength level, such that the variance within each level of s is sufficiently stable. A suitable size has to be determined as usual. A general rule of thumb is to use $m = 10 \times n$ as an initial sample size (Loeppky et al., 2009), where n is the number of policy factors, and then adjust the size upwards if a screening implies that there is too much variance within the sample. Accordingly, the whole experiment will be conducted on the $K \times m$ design point matrix.

Figure 1: Example of a POSA design point matrix with $K = 21$ and 5 policy parameters (A-E)



Source: Own presentation.
Description: The parameters are already matched to the implementation strength on a 0-1 scale. The grey rectangle contains the realistic values with $\underline{x_i}$ and $\overline{x_i}$ being on the lower and upper border. The border cases, i.e. $\underline{\underline{x_i}}$ and $\overline{\overline{x_i}}$, are assigned to 0.0 and 1.0 respectively. We chose 21 strength levels (K=21). The assembly of all parameters (A to E) at their strength level (s) (e.g., $x_A = x_B = x_C = x_D = x_E = 0.2$) marks one policy scenario, a policy at a certain strength level.

- We reduce the variance between the levels by using common random numbers for all levels of s (Schruben et al., 2010). Although the effectiveness of this step is not guaranteed, it does no harm in this setting.[4]

The next step is to conduct the simulation experiment for each policy, resulting in a two-dimensional $m \times K$-matrix $\mathbf{Y}$ where the column vectors $\mathbf{Y}_k$ correspond to the factor levels, i.e. m-times the output variable y at policy strength level K. We suggest analysing the results with a graphical representation such as in figure 2, using the variable s as a factor and depicting the output values, using both, a scatter- and a violin plot with a box plot.

3.4 Selecting the policy scenarios (for further analysis)

Beyond the background of POSA, the researcher can now select the specific policy scenarios, i.e. policies with given implementation strength s, which is the policy scenario that should be analyzed and presented in the main part of this research paper. No matter on what criteria the selection is based on (section 3.1), it is now transparent. Reviewers, policy makers and other readers who get access to the results of this analysis (e.g. in form of figures like figure 2) can judge the effect of this decision and are comforted that the decision is based on a scientific procedure, i.e. it is rigorous, transparent and reproducible, and not an unjustified selection, a *lucky parameter configuration*, or the result of only a cursory analysis.

3.5 Testing the robustness of the selected scenarios

The proposed policy implementation sensitivity analysis rests upon the assumption that the baseline model is calibrated and tested, especially in terms of y. This implies that a global sensitivity analysis has been performed for the calibrated baseline model and it has been proven that it is robust against reasonable changes in the input data, i.e. the input uncertainty. Keeping this in mind, the analyses in sections 3.3 and 3.4 face two limitations:

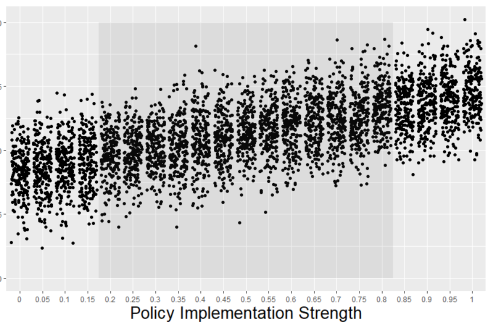
1) The policy implementation sensitivity analysis performed is, strictly speaking, only valid for the exact baseline parameter setting.

2) The policy implementation sensitivity analysis is, strictly speaking, a one-factor-at-a-time (OAT) sensitivity analysis. Even though we defined the factor s as a multitude of factors, there is only a single fixed mapping. The limits of such an approach are well-known (e.g. Saltelli and Annoni (2010)).

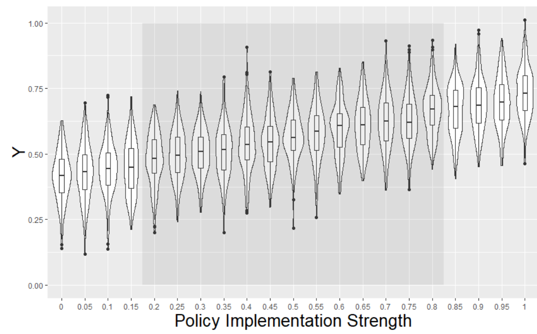
[4] The effectiveness of common random numbers to reduce variance depends on the utilization of random numbers. If the sequence of draws (i.e. the number of draws per simulation step) is also a function of the input parameters, using common random numbers does not reduce the variance. This happens, for example, if the number of agents depends on the policy input factors.

Figure 2: Example POSA graphical analysis for linear, non-linear and no relationship between strength level s and output measure y

(a) Example scatterplot for linear relationship

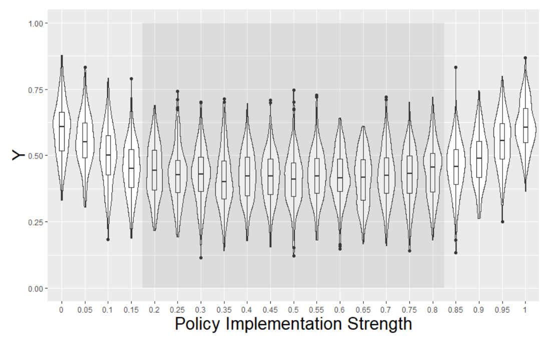


(b) Example violin plot for linear relationship

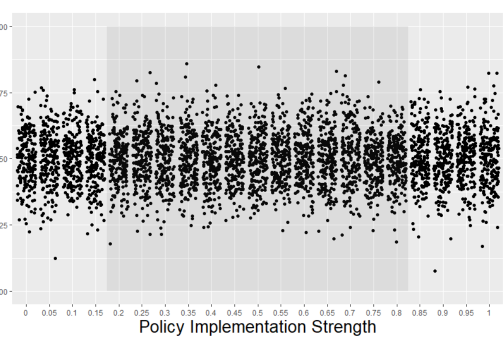


(c) Example scatterplot for non-linear relationship

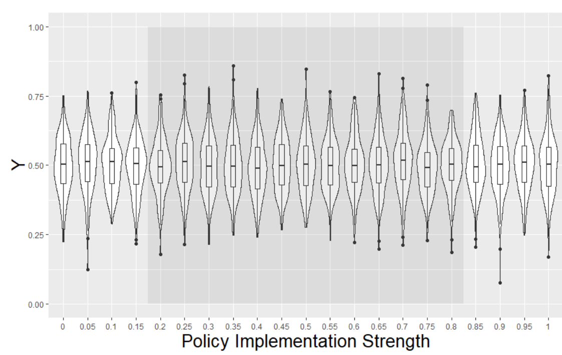
(d) Example violin plot for non-linear relationship



(e) Example scatterplot for an undefined relationship



(f) Example violin plot for an undefined relationship



Source: Own calculation and presentation of random draws to exemplify K strength levels and the measure of output y.
Description: Scatter- and violin plots demonstrate a linear (figures 2a and 2b), a non-linear (figures 2c and 2d) and undefined relationship (figures 2e and 2f) between $K = 21$ strength levels of the policy and output measure of y, with $m = 200$ common random numbers.

We acknowledge that there is a lack of time and space (in a report or paper) that can be attributed to the analyses like POSA. Therefore, the procedure proposed makes only use of the baseline parameter setting (without a sensitivity range), and we limit the analysis to a *combined* OAT analysis, as laid out above. This implies that the results of the policy implementation sensitivity analysis could be different if a different baseline parameter setup or a range of baseline settings would have been used. This is not a problem in and of itself as long as this is clearly kept in mind and the analysis is repeated each time the baseline model is changed.

1. Based on the results of the sensitivity analysis for the underlying model, we propose two strategies to deal with potential criticism for issue 1):

    (a) If the model is very sensitive to reasonably small variations in the input factors, define a set of alternative baseline parameter values (alternative calibrations) and perform the policy sensitivity analysis for each of these baseline parameter settings, presenting the above-mentioned statistics and graphs (figure 2, section 3.3) on the overall set. This allows us to check if the results of the policy are valid for (related) alternative settings.

    (b) If the model remains robust to changes in the calibration, it may instead be reasonable to perform the same kind of global sensitivity analysis that has been performed with the baseline model, but now on the model with the implemented policy scenario (i.e. for each policy scenario that was selected in 3.4). Thus, it is possible to show whether the model maintains its robustness after implementing the policy, which allows to generalize the findings to different settings (or not!).

2. Regarding issue 2), to test the local stability of the results derived from the POSA with respect to the mapping $\mathbf{x} \rightarrow s$, for each policy parameter x_i, one can define an interval of reasonable uncertainty around the selected strength level, i.e. the selected implementation strength (section 3.4), and conduct a local sensitivity analysis:

    (a) If the policy parameters do not interact strongly with the other parameters, guidance can be found in the implementation sensitivity analysis itself: Take the values that result when varying s (see section 3.4) by one or two steps in each direction (the number of steps depends on the sensitivity of the y to s). If the total number of policy parameters is small, a factorial analysis is possible with 3^n combinations, i.e. matching all combinations of low, middle and high values around s.

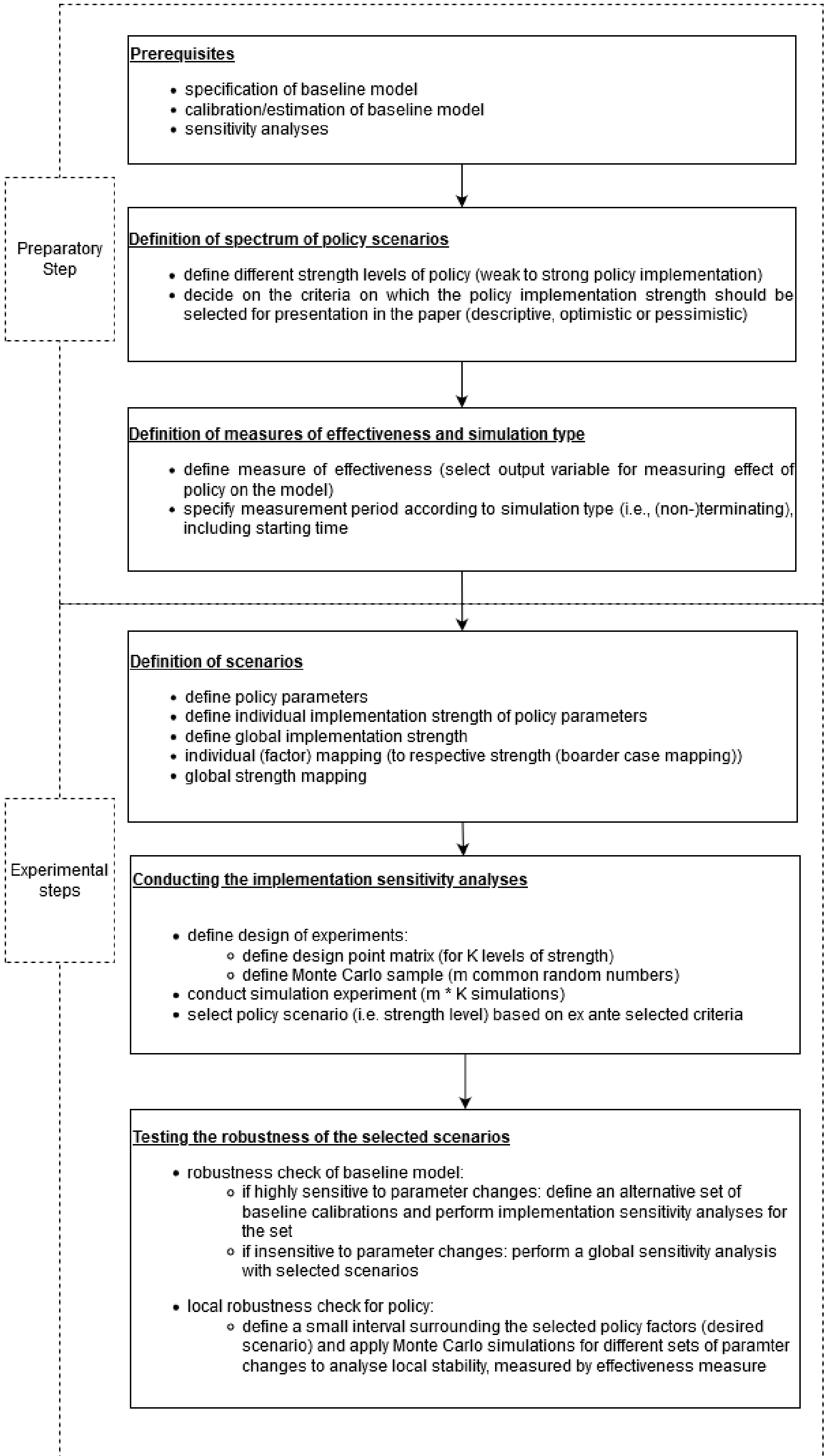
    (b) If there is reason to assume that the policy parameters interact strongly with the uncertainty in the model parameters left over (i.e. except for the policy parameters), which can be tested with repeated sensitivity analyses as suggested in the former paragraph, it may be useful to add the non-policy model parameters to the local sensitivity analysis of 2a, despite that it increases the overall variance.

There are two potential outcomes from 2a and 2b. Either, the model's reaction is *not* too sensitive to the changes, i.e. a minor increase in the variance, but no significant change in y. In this case, nothing else has to be done and the results of POSA are independent of the mapping $(x \rightarrow s)$. If the model exhibits high sensitivity, further analyses of the individual causes are necessary. Perhaps the model itself needs refinement, perhaps the translation of the policy is problematic, or it can be concluded that in fact the effectiveness of the policy is extremely sensitive to some issues related to the mapping of the individual factors, which is also an important result.
At this point, it becomes obvious how POSA addresses the critique of the *lucky parameter configuration*. First, the decisions on the policy implementation strength are based on rigorous scientific criteria set before the policy experiment was implemented rather than an unjustified selection in its aftermath, i.e. when the effects of the implementation strength on the output measure y were revealed. Second, by varying the strength of the policy parameters in the local stability analysis (analysis 2a of this section) the impact of small deviations is visualized to ensure the local stability of the results for variations.

4 Summary

Agent-based simulations allow testing for how policies affect a policy target under the consideration of heterogeneous agents, the macroeconomic environment and the interplay between the micro- and the macroeconomic level. However, the complexity of agent-based models paired with the complexity of possible policies makes the results of policy analyses subject to criticism, e.g. that the results are due to a lucky parameter configuration. This paper provides a procedure that accounts for this critique and improves transparency and comprehensibility for academic readers and (political) decision-makers unfamiliar with ACE. We developed a procedure for policy experiments that helps to conceptualize and conduct policy experiments in macroeconomic ABM. The procedure makes the decisions on how to implement a policy experiment in ABM as well as the robustness of the selected policy scenario and the consequences of alternative specifications transparent. The application of the *policy implementation sensitivity analysis*, POSA, produces graphical and/or numerical reports that should be used in the appendix of the original research paper in order to increase its credibility. The full procedure is summed up and visualized as a checklist in figure 3.

Figure 3: POSA checklist



Source: Own presentation.
Description: Summary of the steps of the POSA procedure.

A Appendix

A.1 Illustration: applying POSA to an example

After we laid out the procedure in the previous sections, we apply POSA to Bauermann (2020)'s model to improve the comprehensibility and its steps with a practical example.

A.1.1 Prerequisites and preliminary notes (preparatory step)

The model of Bauermann (2020) is a modification of the baseline model of Delli Gatti et al. (2011). It was set up to analyze the effects of different labor market policies on unemployment when carried out during a recession. The effects on short-run unemployment, the average rate over periods 1 to 10 following the introduction of the policy, and on the long-run development, the average over periods 40 to 100 following the policy, were analyzed. The model will not be presented in details, since the focus of this paper is on the POSA procedure. Readers are asked to consult Bauermann (2020) for more detailed information on the model. The baseline parameter values of the model were found via indirect calibration (Windrum, 2007) to match a selection of stylized facts on the German economy between the second half of the 1960s and the early 1990s (Fiorito and Kollintzas, 1994; Brandner and Neusser, 1992). The robustness of the model parameter setup was checked via a nearly-orthogonal latin hypercube sampling (NOL).[5] In the following example, we want to focus on labor market policies that affect the search efforts of unemployed, the generosity of long- and short-term unemployment benefits and the eligibility of the latter type of benefits. In other words, these are the *policy parameters* carried out in the policy analysis. As some of the most important parameters in the labor market reforms, these parameters have been frequently targeted in the recent decades, especially in Southern and Continental Europe (Turrini et al., 2014). We define the *strong* and the *weak implementation*. The weak policy implementation is supposed represent a labor market that is associated with rather generous short- and long-term unemployment benefits, a long period of drawing short-term unemployment benefits and low job search efforts. The strong policy implementation is associated with reduced unemployment benefits, especially low long-term unemployment benefits, a short duration of drawing the more generous short-term unemployment benefits and high job search efforts, which point in the direction of the labor market reforms of the recent decades (Turrini et al., 2014). The *policy target* in our analysis is the reduction in the unemployment rate. We analyze the *optimistic* (policy) implementation strength, which is associated to the idea, or

[5] Bauermann (2020) used an NOL to control how small deviations from the baseline parameter setup affect central statistics. Details and the results can be found in the appendix of Bauermann (2020). For the Python code to produce the settings, please consult: `https://github.com/FrederikSchaff/DOEsimple`.

the *expectations*, that labor market rigidities have to be reduced to meet the policy target (e.g., Siebert (1997)), i.e. a strong implementation. The *measures of effectiveness* of the policies are the short- and long-run unemployment rates. The aim is to see short-run effects, i.e. how long it takes until unemployment is back to its long-run value (i.e. the statistical equilibrium (Grazzini, 2012)), and whether the long-run value changes. This qualifies the investigations as *steady-state* analyses (see: section 3.1). Regarding the timing of the policy experiment, the shift in the policy is triggered by the event of a recession. In other words, the policy is introduced at different strength levels when the model economy faces a recession.[6]

A.1.2 Definition of scenarios

The *policy parameters* in our model, i.e. the parameters that change during the experiments, are a) short-term unemployment benefits (UB), via changes in κ_{short}; b) long-term unemployment benefits (UB), via changes in κ_{long}; c) the duration of the more generous short-term UB, denoted by DUR; and d) the search efforts of unemployed, denoted by SE. Short-term UB is defined in relation to the average wage in the economy, e.g., 55% ($\kappa_{short} = 0.55$) of the average wage in the baseline model. Long-term UB is defined in a similar way (i.e. $\kappa_{long} = 0.4$). The duration of short-term UB is defined as periods of eligibility for short-term UB, and search efforts of the unemployed is defined as the number of applications that unemployed agents send to firms. The baseline parameter values, before the policies take place, are: $\kappa_{short} = 0.55$, $\kappa_{long} = 0.4$, $DUR = 6$ and $SE = 7$. For the above-mentioned parameters, Tab. 1 depicts the values for the baseline parameter setup (calibrated to the German economy between the 1960s and the early 2000s), the *weakest realistic implementation strength* ($\underline{x_i}$), the *theoretically weakest implementation strength* ($\underline{\underline{x_i}}$), the *strongest realistic implementation strength* ($\overline{x_i}$) and the *theoretically strongest implementation strength* ($\overline{\overline{x_i}}$). Except for search efforts, the values are mainly taken from OECD (2012), OECD (2014) and van Vliet and Caminada (2012). Due to the absence of precise criteria or values, the values for search efforts are chosen to represent significant but still plausible increases. $\underline{x_i}$ and $\overline{x_i}$ are case-specific. In other words, these are plausible values for a specific country, e.g. Germany. $\underline{\underline{x_i}}$ and $\overline{\overline{x_i}}$ are chosen on theoretical base. $\underline{\underline{x_i}}$ and $\overline{\overline{x_i}}$ are extreme values, which could appear in theory or in some countries, e.g. $\underline{\underline{\kappa_{long}}} = 0$ in Italy. Similar to section 3.2 in the main part of the paper, we choose $K = 21$ (strength levels). As shown in figure 4, we map the parameters into the global implementation strength levels (s). Section A.2 in the appendix provides an overview of the exact individual mapping (section A.2, matrix (1)).

[6] We can rely on an empirical definition of NBER (2010). Following NBER (2010), "*[...]a recession is a significant decline in economic activity spread across the economy, lasting more than a few months,*

Table 1: Magnitudes of policy reforms

Parameter	Baseline	$\underline{\underline{x}}_i$	$\underline{x}_i$	$\overline{x}_i$	$\overline{\overline{x}}_i$
Duration Short-term UB, DUR	6	12	8	4	1
Short-term UB(% of average wage), κ_{short}	0.55	0.8	0.65	0.4	0.2
Long-term UB (% of average wage), κ_{long}	0.4	0.57	0.5	0.2	0
Search Efforts, SE	7	3	3	13	13

Source: Own figure based on OECD (2012), OECD (2014), van Vliet and Caminada (2012) and own description.

Combining the individual strength mappings yields the *global strength mapping* (figure 4 below and matrix (1) (see section A.2; columns denote the strength level with $\overline{\overline{s}}$ at the top)). As described in subsection A.1.2 above, a policy scenario consists of four parameter values. In figure 4, the strength of a policy is depicted by four parameter values at one strength level s. As in section 3.2 of the main part of the paper, the shaded areas of figure 4 depict the area of the realistic policy implementation strength levels while the white area depicts the rather extreme implementation strength levels. For example, the weakest realistic policy is depicted by the parameter values at strength level 0.2, which is the lower edge of the grey area.

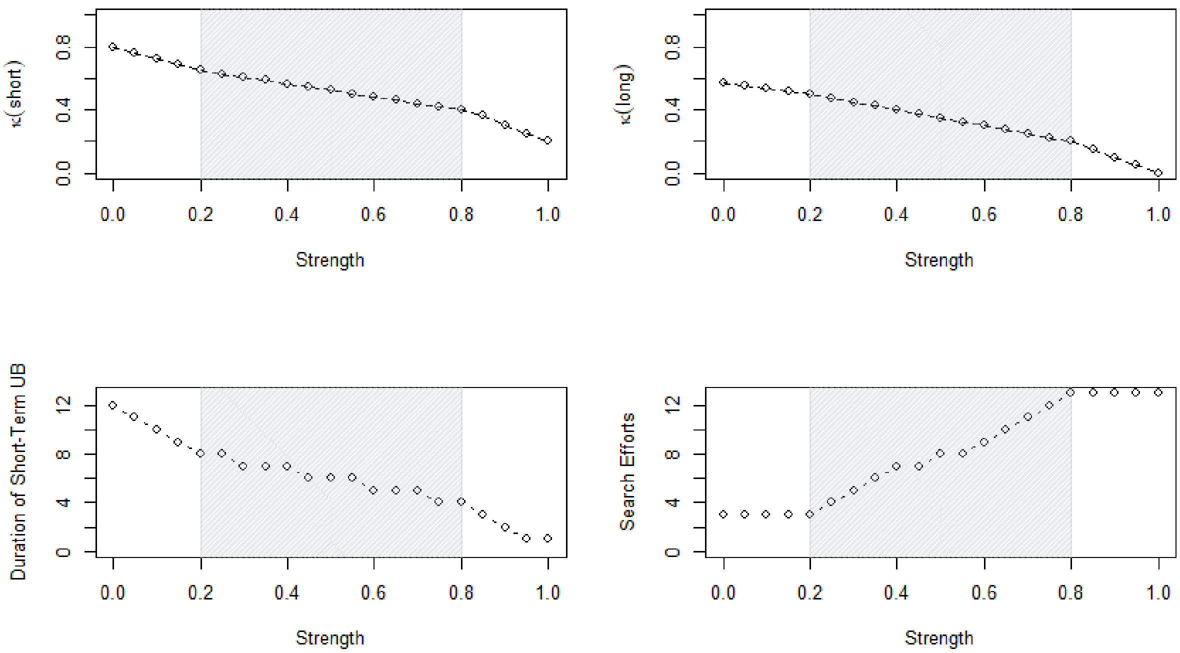
A.1.3 Conducting the implementation sensitivity analysis

After defining the scenarios in section A.1.2, the *policy implementation sensitivity analysis* is carried out in which the effect of the policy strength on the measures of effectiveness, i.e. the long- and short-run unemployment, are analyzed. The baseline calibration is fixed and, as it was described above, the policy experiment starts when the economy enters a recession. In other words, we look at how the output variables change when one of the aforementioned K strength levels are introduced. We select a Monte Carlo sample size of 128 runs.[7]
Figures 5 and 6 show the short- and long-run unemployment for 128 runs per 21 strength levels, depicted as violin plots. From figure 5, it becomes visible that the varying magnitudes of our policy measures do not yield significant effects on the average short run unemployment. Neither the average unemployment nor the variance across the magnitudes seem to be affected. However, the picture changes for the long-run unemployment, especially for the more extreme strengths, i.e. the violin plots outside the grey rectangle in figure 6. Taken

normally visible in real GDP [...], employment, [...].” Accordingly, a recession in the model is identified, when unemployment exceeds its long-run average by two times the standard deviation, more precisely 12%.

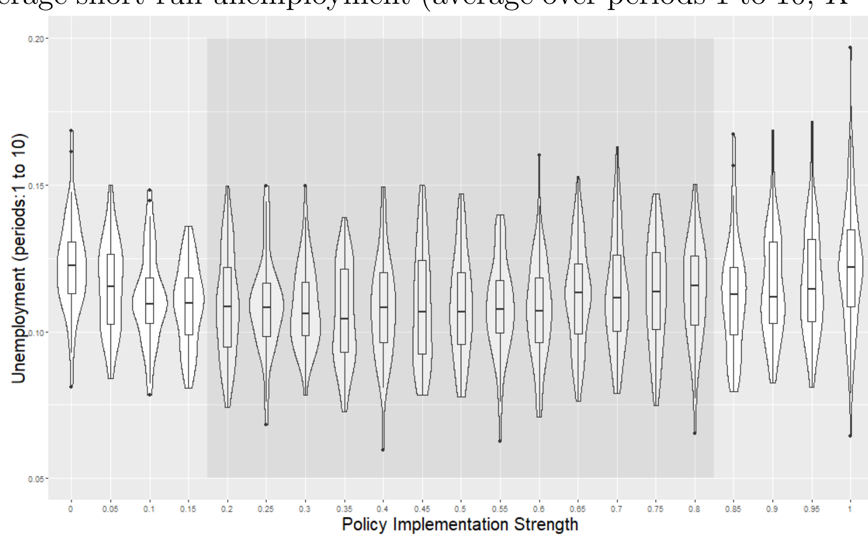
[7] Via the procedure of variance stability (Lee et al., 2015), we found that the variance of the long-run unemployment in our model (baseline parameter setting) stabilizes at 128 runs. Further simulations reduce the variance, but not significantly. The policy experiment starts when the burn-in phase has passed and unemployment has reached the critical value of 0.12.

Figure 4: Strength levels for all parameters (K points)



Source: Own figure based on OECD (2012), OECD (2014), van Vliet and Caminada (2012) and own description.
Description: The horizontal axis depicts the 21 strength levels. While the left axis depicts the level of the UB in relation to the average wage, the right axis depicts the number of applications sent per unemployed and number of period for the short-term UB.

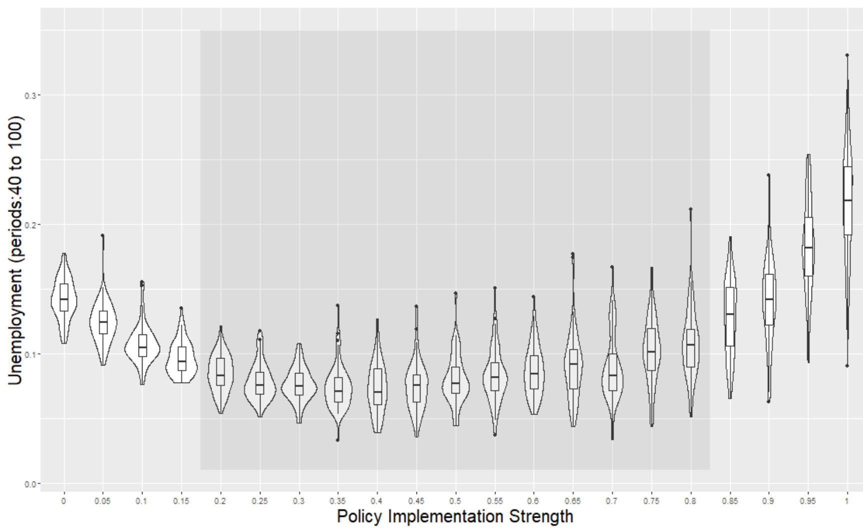
Figure 5: Average short-run unemployment (average over periods 1 to 10, $K = 21$, 128 runs)



Source: Own figure.
Description: Every violin plot depicts the average short-run unemployment for the first 10 periods following the policy experiment for 128 runs. There are 21 violin plots since we analyze $K = 21$ strength levels.

from figure 6, a U-shaped relationship between the strength level and the long-run unemployment seems to appear. Recalling that the changes in a policy are lasting, unemployment increases, especially, for an extremely weak ($\underline{\underline{s}}$) and an extremely strong policy implementation strength level ($\overline{\overline{s}}$).[8] A much smaller U-shaped relationship also appears if we focus on the more realistic strength levels of the policy experiments, i.e. the more realistic changes (grey boxes of figure 6).

Figure 6: Average long-run unemployment (average over periods 40 to 100, $K = 21$, 128 runs)



Source: Own figure.
Description: Every violin plot depicts the unemployment for periods 40 to 100 following the policy experiment for 128 runs. There are 21 violin plots, since we analyze $K = 21$ strength levels.

At this point, we note that the U-shaped relationship could not have been found if only a selected scenario instead of a range of changes would have been depicted, which proves the usefulness of this procedure.

A.1.4 Selecting the policy scenario

In this example of a policy experiment, we would like to implement policies that favour the idea of labor market liberalization. As such, low search efforts in combination with generous unemployment benefits are seen as the main cause of high unemployment rates in Europe (e.g., Turrini et al. (2014), Bassanini and Duval (2006) and Siebert (1997)). More precisely, we want to implement an *optimistic* implementation strength, which assumes that the reduction of labor market rigidities lead to a reduction in unemployment. For this

[8] The strong policy is denoted by a low unemployment benefits, high search efforts of unemployed and short duration of short-term unemployment benefit eligibility. The weak policy is denoted by a high unemployment benefits, low search efforts of unemployed and a long duration of short-term unemployment benefit eligibility.

purpose, the long-term unemployment benefits (κ_{long}) are reduced to 20% of the average wage, short-term unemployment benefits (κ_{short}) drop to 40%, the duration of the short-term unemployment benefits drop to four periods (DUR), and the search efforts (SE) increase from 7 to 12 applications per unemployed and per period, which is similar to implementation strength $\bar{s}$.

A.1.5 Testing the robustness of the selected scenario

The model is sensitive to small variations in the input factors (Bauermann (2020), Appendix). Therefore, we apply the robustness check 1a (section 3.5). We define a set of alternative baseline parameter values (alternative calibrations) and perform the above-mentioned policy sensitivity analysis for these alternative baseline parameter sets (see subsection A.1.3). For this purpose, we select 128 different baseline parameter settings via a nearly-orthogonal latin hypercube sampling (NOL).[9] Table 2 shows the baseline calibration and between which values the 128 different baseline calibrations are selected, i.e. between which upper and lower values.

Table 2: Setting for finding alternative baseline parameterizations (via nearly-orthogonal latin hypercube sampling)

Description	**Parameter**	**Baseline value**	**Lower values**	**Upper values**	**Increment**
Number of firms	F	100	80	120	5
Number of workers	H	1668	1550	1750	10
Number of banks/aggregated banking sector	Z	1	1	1	-
Number of governments	G	1	1	1	-
Number of central banks	CB	1	1	1	-
Price mark(-down)-up	η	$U(0;(-)0.07)$	$U(0;(-)0.05)$	$U(0;(-)0.09)$	0.005
Quantity mark(-down)-up	χ	$U(0;(-)0.1)$	$U(0;(-)0.08)$	$U(0;(-)0.12)$	0.005
Wage mark-up (if insufficient applications)	δ	$U(0;0.09)$	$U(0;(-)0.06)$	$U(0;(-)0.12)$	0.005
Price mark-down (entrants)	γ	$U(0;-0.05)$	$U(0;(-)0.03)$	$U(0;(-)0.07)$	0.005
Reservation wage mark-down	ζ	$U(0;-0.08)$	$U(0;(-)0.07)$	$U(0;(-)0.09)$	0.005
Ratio lower bound reservation wage to UB	ν	$U(0.1;0.9)$	$U(0.0;0.7)$	$U(0.3;0.9)$	0.1
Ratio: Buffer-stock savings/wage	ξ	2.0	1.0	3.0	1
Wage tax	τ	0.05	0.04	0.06	0.005
Factor marginal propensity to consume	β	0.5	0.4	0.6	0.05
Marginal propensity to consume from excess wealth	λ	0.5	0.3	0.6	0.1
Capital requirement coefficient	v	0.15	0.1	0.5	0.005
Fragility weight of firm debt	θ	0.4	0.2	0.5	0.1
Maximal leverage of firm accepted	ι	2.0	1.0	3.0	0.5
Base (interest) rate	$\hat{r}$	0.01	0.008	0.012	0.001
Percentage of firms with lowest price checked	α	5%	1%	7%	1%-point
Ratio: short-term unemployment benefits to average wage	κ_{short}	0.55	0.5	0.6	0.05
Ratio: long-term unemployment benefits to average wage	κ_{long}	0.4	0.37	0.42	0.01
Maximal of (job) applications to firms	M	7	6	8	1
Number of suppliers checked	O	7	5	9	1
Duration of short-term unemployment benefits (periods)	DUR	4	3	5	1

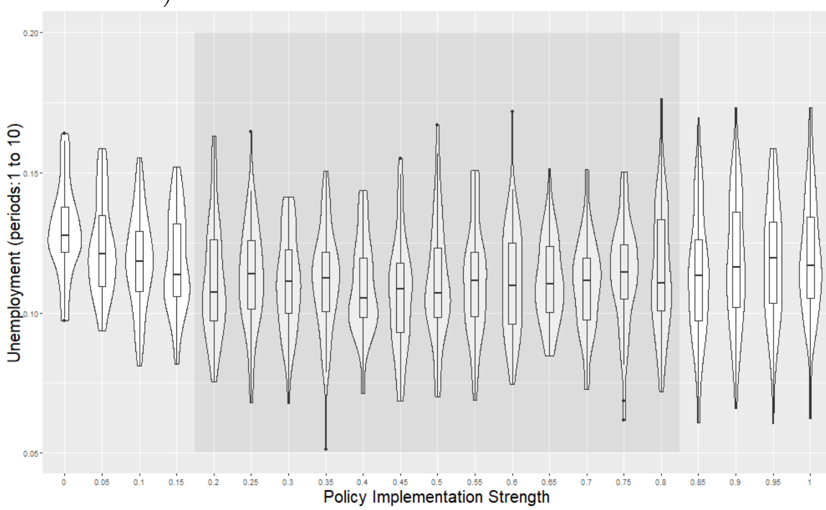
Description: Table 2 presents the parameters of the model and their upper and lower boundaries that the latin hypercube design was based on.

Figures 7 and 8 show the respective graphics of the policy implementation sensitivity analyses

[9] We used the same procedure as noted in footnote 5.

for the variations in the baseline parameterizations. These figures do not differ significantly from figures 5 and 6 above, which implies that the results are stable for reasonable (small) variations in the underlying baseline parameterization, and that the interpretations from above hold for such variations. Only, figures 7 and 8 show a slightely larger variance in the output variables compared to figures 5 and 6. However, it is comprehensible given that the parameters and not only the random seeds vary as in figures 7 and 8.

Figure 7: Average short-run unemployment (average over periods 1 to 10, $K = 21$, 128 baseline parameterizations)



Source: Own figure.

Description: Every violin plot depicts the average short-run unemployment rate for the first 10 periods following the policy experiment for 128 runs. There are 21 violin plots, since we analyze $K = 21$ strength levels.

Further, the local stability of the selected policy scenario derived from the implementation sensitivity analysis is tested (2b in section 3.5). For this purpose, we apply a 3^k full factorial design (Montgomery, 2017), since the number of factors is relatively small. Based on our selected scenario (from section A.1.4), we analyze whether small, *non-linear* deviations in the policy parameters from the selected policy scenario ($\bar{s} = 0.75$) have an effect on the results shown in section A.1.3. Starting from the model with the baseline parameter values, the policy parameters vary discretely, i.e. $\kappa_{short} \in \{0.15, 0.2, 0.23\}$, $\kappa_{long} \in \{0.36, 0.4, 0.421\}$, $DUR \in \{3, 4, 5\}$ and $SE \in \{12, 13, 14\}$, when the economy enters a recession. The local stability analysis checks whether the results of a certain strength level (section A.1.3) hold when the policy parameters vary in different directions instead of a fixed mapping. If there are strong interaction effects between the policy parameters, changing the variations is supposed to have a strong effect on the output measure(s) y, the short- and long-run unemployment. Therefore, in the case of strong interaction effects, the distribution of the short and long-run unemployment should differ between the values from the selected scenario and those of the

Figure 8: Average long-run unemployment (average over periods 40 to 100, $K = 21$, 128 baseline parameterizations)

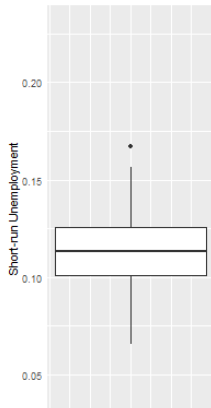
Source: Own figure.
Description: Every violin plot depicts the unemployment for periods 40 to 100 following the policy experiment for 128 runs. There are 21 violin plots, since we analyze $K = 21$ strength levels.

local stability analyses.

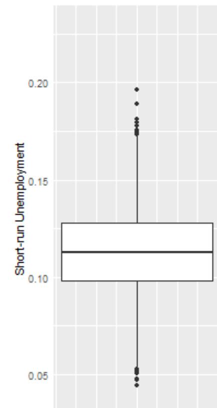
Figures 9 and 10 show the results of the local sensitivity analyses, presented as boxplots. Figure 9a shows the boxplot for the short-run unemployment when the policy strength s equals 0.75, 0.8 or 0.85, all included in one boxplot. Figure 9b shows the respective boxplot, given the above-mentioned 3^k full factorial design. Comparing figures 9a and 9b, the boxplots are failry similar to each other, which implies that variations in the mapping of the policy parameters do not affect the results of the above-mentioned analyses (results from POSA, section A.1.3). Only the variance of the right graph (figure 9b) is a little bit larger, but not noticeably so. Figure 10 shows the results for the long-run unemployment. Figure 10a shows the boxplot for the short-run unemployment when the policy strength s equals 0.75, 0.8 or 0.85, again, all included in one boxplot. Figure 10b shows the respective boxplot for the above-mentioned 3^k full factorial design. Again, the median, the first and the third quartile do not differ significantly, which implies that variations in the mapping of the policy parameters does not affect the results of the above-mentioned analyses (results from POSA, section A.1.3). In addition to the graphical local stability analyses, we estimate a fixed effects regression model as statistical local stability analyses with the random seeds being the fixed effects (see table 3). Table 3 shows the effects of the policy parameters themselves and in combination (interaction effects). The table shows that small (local) deviations from the policy scenario, selected in section A.1.4, do not have a significant effect on the output variables, which supports the graphical analyses above.

Figure 9: Local stability analysis (short-run unemployment)

(a) Boxplot for policy implementation strength levels $s \in \{0.75, 0.8, 0.85\}$



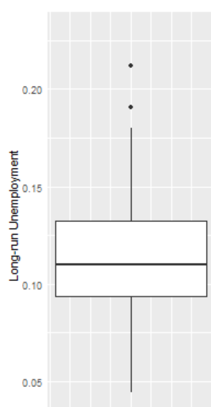
(b) Boxplot for full factorial design



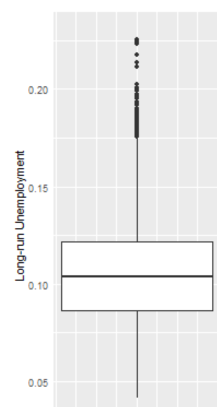
Source: Own computations and simulations.
Description: Boxplots 9a and 9b show the short-run unemployment, mean unemployment between periods 1 to 10 after the policy is introduced (starting from the baseline parameterization). Figure 9a shows the boxplot for the short-run unemployment when the policy strength s equals 0.75, 0.8 or 0.85. Figure 9b shows the boxplot for the short-run unemployment given the 3^k full factorial design (i.e. $\kappa_{Short} \in \{0.15, 0.2, 0.23\}$, $\kappa_{Short} \in \{0.36, 0.4, 0.421\}$, $DUR \in \{3, 4, 5\}$ and $SE \in \{12, 13, 14\}$). It depicts 81 combinations and 96 runs (common random numbers) for each.

Figure 10: Local stability analysis (long-run unemployment)

(a) Boxplot for policy implementation strength levels $s \in \{0.75, 0.8, 0.85\}$



(b) Boxplot for full factorial design



Source: Own computations and simulations.
Description: Boxplots 10a and 10b show the long-run unemployment, mean unemployment between periods 40 to 100 after the policy is introduced (starting from the baseline parameterization). Figure 10a shows the boxplot for the long-run unemployment when the policy strength s equals 0.75, 0.8 or 0.85. Figure 10b shows the boxplot for the long-run unemployment given the 3^k full factorial design (i.e. $\kappa_{long} \in \{0.15, 0.2, 0.23\}$, $\kappa_{long} \in \{0.36, 0.4, 0.421\}$, $DUR \in \{3, 4, 5\}$ and $SE \in \{12, 13, 14\}$). It depicts 81 combinations and 96 runs (common random numbers) for each.

Table 3: Local stability analyses ((fixed-effects) regression analyses for small deviations from selected policy scenario)

	Dependent variable:	
	Short-run unemployment	Long-run unemployment
	(1)	(2)
κ_{long}	−0.108	−2.543
	(6.249)	(9.539)
κ_{short}	−0.204	−1.234
	(3.093)	(4.721)
DUR	0.003	−0.016
	(0.300)	(0.458)
SE	0.005	−0.038
	(0.094)	(0.144)
$\kappa_{long} * \kappa_{short}$	0.932	6.401
	(15.769)	(24.070)
$\kappa_{long} * DUR$	−0.051	0.160
	(1.531)	(2.337)
$\kappa_{short} * DUR$	0.010	0.045
	(0.758)	(1.156)
$\kappa_{long} * SE$	−0.022	0.194
	(0.480)	(0.732)
$\kappa_{short} * SE$	0.001	0.100
	(0.237)	(0.362)
$DUR * SE$	−0.001	0.002
	(0.023)	(0.035)
$\kappa_{long} * \kappa_{short} * DUR$	0.026	−0.531
	(3.863)	(5.896)
$\kappa_{long} * \kappa_{short} * SE$	0.006	−0.529
	(1.211)	(1.848)
$\kappa_{long} * DUR * SE$	0.011	−0.014
	(0.118)	(0.179)
$\kappa_{short} * DUR * SE$	0.002	−0.007
	(0.058)	(0.089)
$\kappa_{long} * \kappa_{short} * DUR * SE$	−0.021	0.052
	(0.297)	(0.453)
Observations	7,776	7,776
R^2	0.014	0.073
Adjusted R^2	0.0002	0.059
F Statistic (df = 15; 7665)	7.417***	39.986***
Model	fixed effects model (random seeds)	fixed effects model (random seeds)
Note:		*p<0.1; **p<0.05; ***p<0.01

Description: Table 3 shows the effects of the policy parameters on the output variables, short- and long-run unemployment. The model is a fixed-effects regression estimation, the random seeds are the fixed effects. The variables follow the descriptions and notations of section A.1.1 to A.1.5.

A.2 Global strength mapping

The global strength mapping the policy factors described as matrix

$$X(s) = [\vec{DUR} \quad \vec{\kappa}_{short} \quad \vec{\kappa}_{long} \quad \vec{SE}] = \begin{bmatrix} 12 & 0.8 & 0.57 & 3 \\ 11 & 0.763 & 0.55 & 3 \\ 10 & 0.725 & 0.54 & 3 \\ 9 & 0.688 & 0.52 & 3 \\ 8 & 0.65 & 0.5 & 3 \\ 8 & 0.629 & 0.48 & 4 \\ 7 & 0.608 & 0.46 & 5 \\ 7 & 0.588 & 0.44 & 6 \\ 7 & 0.567 & 0.42 & 7 \\ 6 & 0.550 & 0.40 & 7 \\ 6 & 0.525 & 0.37 & 8 \\ 6 & 0.504 & 0.34 & 8 \\ 5 & 0.483 & 0.31 & 9 \\ 5 & 0.463 & 0.28 & 10 \\ 5 & 0.442 & 0.25 & 11 \\ 4 & 0.421 & 0.23 & 12 \\ 4 & 0.4 & 0.2 & 13 \\ 3 & 0.366 & 0.15 & 13 \\ 2 & 0.35 & 0.1 & 13 \\ 1 & 0.25 & 0.05 & 13 \\ 1 & 0.2 & 0.0 & 13 \end{bmatrix} \tag{1}$$

References

Alexopoulos, C. (2006). A Comprehensive Review of Methods for Simulation Output Analysis: Proceedings of the 2006 Winter Simulation Conference. In Perrone, L. F., Wieland, F. P., Liu, J., Lawson, B. G., Nicol, D. M., and Fujimoto, R. M., editors, *Proceedings of the 38th conference on Winter simulation*, pages 168–178. Winter Simulation Conference, s.l.

Baptista, R., Farmer, J. D., Hinterschweiger, M., Low, K., Tang, D., and Uluc, A. (2016). Macroprudential policy in an agent-based model of the UK housing market. Bank of England working papers 619, Bank of England.

Bassanini, A. and Duval, R. (2006). Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions. OECD Economics Department Working Papers 486, OECD Publishing.

Bauermann, T. (2020). Governmental policies to reduce unemployment during recessions: Insights from an ABM. Ruhr Economic Papers 847, Essen.

Brandner, P. and Neusser, K. (1992). Business cycles in open economies: Stylized facts for austria and germany. *Review of World Economics (Weltwirtschaftliches Archiv)*, 128(1):67–87.

Caiani, A., Godin, A., Caverzasi, E., Gallegati, M., Kinsella, S., and Stiglitz, J. E. (2016). Agent based-stock flow consistent macroeconomics: Towards a benchmark model. *Journal of Economic Dynamics and Control*, 69:375–408.

Campolongo, F., Saltelli, A., and Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182(4):978–988.

Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011). Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7):1013–1041.

Cincotti, S., Raberto, M., and Teglio, A. (2012). Macroprudential Policies in an Agent-Based Artificial Economy. *Revue de l'OFCE*, 124(5):205.

Dawid, H. and Delli Gatti, D. (2018). Agent-Based Macroeconomics. In Hommes, C. and LeBaron, B., editors, *Handbook of Computational Economics*, volume 4 of *Handbook of Computational Economics*, chapter 2, pages 63 – 156. Elsevier.

Dawid, H. and Fagiolo, G. (2008). Agent-based models for economic policy design: Introduction to the special issue: Agent-based models for economic policy design - Agent Based Models. *Journal of Economic Behavior & Organization*, 67(2):351–354.

Dawid, H., Gemkow, S., Harting, P., van der Hoog, S., and Neugart, M. (2018). Agent-Based Macroeconomic Modeling and Policy Analysis: The Eurace@Unibi Model. In Chen, S.-H., Kaboudan, M., and Du, Y.-R., editors, *The Oxford Handbook of Computational Economics and Finance*, pages 490–519. Oxford University Press.

Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P., and Gallegati, M. (2011). *Macroeconomics from the Bottom-up*, volume 1 of *New Economic Windows*. Springer Milan, Milano.

Dosi, G., Fagiolo, G., and Roventini, A. (2006). An Evolutionary Model of Endogenous Business Cycles. *Computational Economics*, 27(1):3–34.

Dosi, G., Pereira, M., Roventini, A., and Virgillito, M. E. (2018). What if supply-side policies are not enough? The perverse interaction of flexibility and austerity.

Edmonds, B. and Hales, D. (2003). Replication, Replication and Replication: Some Hard Lessons from Model Alignmen. *Journal of Artificial Societies and Social Simulation*, 6.

Fiorito, R. and Kollintzas, T. (1994). Stylized facts of business cycles in the g7 from a real business cycles perspective. *European Economic Review*, 38(2):235–269.

Grazzini, J. (2012). Analysis of the emergent properties: Stationarity and ergodicity. *Journal of Articial Societies and Social Simulation*, 15(2):1–15.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., and Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, 221(23):2760–2768.

Laatabi, A., Marilleau, N., Nguyen-Huu, T., Hbid, H., and Ait Babram, M. (2018). ODD+2D: An ODD Based Protocol for Mapping Data to Empirical ABMs. *Journal of Artificial Societies and Social Simulation*, 21(2).

Laliotis, D., Buesa, A., Leber, M., and Poblacin Garca, F. J. (2019). An agent-based model for the assessment of LTV caps. Working Paper Series 2294, European Central Bank.

Law, A. M. (1980). Statistical analysis of the output data from terminating simulations. *Naval Research Logistics Quarterly*, 27(1):131–143.

Law, A. M. (2015). Statistical analysis of simulation output data: The practical state of the art. In Yilmaz, L., Chan, W. K. V., Moon, Roeder, T. M. K., Macal, C., Rossetti, M. D., and Yilmaz, L., editors, *Proceedings of the 2015 Winter Simulation Conference*, 2015 Winter Simulation Conference (WSC), pages 1810–1819. IEEE, Piscataway, NJ.

LeBaron, B. and Tesfatsion, L. (2008). Modeling Macroeconomies as Open-Ended Dynamic Systems of Interacting Agents. *The American Economic Review*, 98(2):246–250.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, J. G., Sun, Z., and Parker, D. C. (2015). The Complexities of Agent-Based Modeling Output Analysis. *Journal of Artificial Societies and Social Simulation*, 18(4):4.

Loeppky, J. L., Sacks, J., and Welch, W. J. (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, 51(4):366–376.

Lorscheid, I., Heine, B.-O., and Meyer, M. (2012). Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments. *Computational & Mathematical Organization Theory*, 18(1):22–62.

Montgomery, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & Sons Inc., Hoboken, 8th edition.

NBER (2010). US Business Cycle Expansions and Contractions. .

OECD (2012). Maximum duration of unemployment benets in oecd countries. `https://www.oecd-ilibrary.org/maximum-duration-of-unemployment-benefits-in-oecd-countries-2012_5jxrfg7hd0lq.xls?itemId=%2Fcontent%2Fcomponent%2F9789264227200-graph35-en&mimeType=vnd.ms-excel`. Cited at 02nd September 2018.

OECD (2014). Benets and wages: Statistics. http://www.oecd.org/els/benefits-and-wages-statistics.htm. Cited ar 02nd September 2018.

Pissarides, C. (2000). *Equilibrium Unemployment Theory*. MIT Press.

Richiardi, M. G. (2017). The Future of Agent-Based Modeling. *Eastern Economic Journal*, 43(2):271–287.

Saltelli, A. and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517.

Schaff, F. (2016). *Pure Agent-based Computational Economics of Time, Knowledge and Procedural Rationality with an Application to Environmental Economics*. PhD thesis, FernUniversität in Hagen, Hagen.

Schruben, L. W., Cochran, J. J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P., and Smith, J. C. (2010). Common Random Numbers. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc.

Siebert, H. (1997). Labor market rigidities: At the root of unemployment in europe. *The Journal of Economic Perspectives*, 11(3):37–54.

Steinmann, P., Wang, J. R., van Voorn, G. A. K., and Kwakkel, J. H. (2020). Don't try to predict COVID-19. If you must, use Deep Uncertainty methods. *Review of Artificial Societies and Social Simulation*.

Tesfatsion, L. (2001). Introduction to the special issue on agent-based computational economics. *Journal of Economic Dynamics and Control*, 25(3-4):281–293.

Turrell, A. (2016). Agent-based models: understanding the economy from the bottom up. *Bank of England Quarterly Bulletin*, 56(4):173–188.

Turrini, A., Koltay, G., Pierini, F., Goffard, C., and Kiss, A. (2014). A decade of labour market reforms in the eu: Insights from the labref database. *EUROPEAN ECONOMY Economic Papers*, 522:52.

van Vliet, O. and Caminada, K. (2012). Unemployment replacement rates dataset. https://www.universiteitleiden.nl/en/law/institute-for-tax-law-and-economics/economics/data-sets/unemployment-replacement-rates-dataset. Cited at 2nd September 2018.

Windrum, P. (2007). Neo-Schumpeterian simulation models: Chapter 26. In Hanusch, H. and Pyka, A., editors, *Elgar companion to neo-Schumpeterian economics*, pages 405–439. Edward Elgar, Cheltenham.

Wolf, S., Bouchaud, J.-P., Cecconi, F., Cincotti, S., Dawid, H., Gintis, H., van der Hoog, S., Jaeger, C., Kovalevsky, D., Mandel, A., and Paroussos, L. (2013). Describing economic agent-based models – Dahlem ABM documentation guidelines. *Complexity Economics*, 2:63–74.