Journal of Economic Behavior and Organization 228 (2024) 106786

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research paper

Calibration and validation of macroeconomic simulation models by statistical causal search ☆

Mario Martinoli, Alessio Moneta *, Gianluca Pallante

Institute of Economics & L'EMbeDS, Sant'Anna School of Advanced Studies, Italy

ARTICLE INFO

JEL classification:
C32
C52
E37

Keywords:
Model evaluation
Identification
Independent component analysis
Causal inference
Model confidence set
Minimum distance index

ABSTRACT

We introduce a general procedure for macroeconomic models' calibration and validation. Configurations of parameters are selected on the basis of a loss function involving a distance between model-derived structural coefficients and their empirical counterparts. These, in both cases, are locally identified by exploiting non-Gaussianity in a structural vector autoregressive framework under a data-driven approach. We use model confidence set to account for the uncertainty in the selection procedure. We provide a measure of validation by comparing (model's and empirical) shocks-variables structure. We apply our procedure to a complex macroeconomic simulation model that studies the link between climate change and economic growth.

1. Introduction

Policy evaluation in macroeconomics is traditionally carried out within the framework of formal models. Such models serve as surrogates of laboratories in which, through simulation, counterfactual questions can be addressed. Questions may concern the effects of systematic changes in fiscal or monetary policy, but also the economic consequences of climate change. It is evident that the results of simulations are reliable and useful insofar as the models are empirically plausible; namely to the extent that they are taken to the data through estimation, calibration or validation (see, e.g., Ireland, 2004; Christiano et al., 2018). In this paper, we propose a general procedure to both calibrate and (at a subsequent stage) validate macroeconomic models that are sufficiently complex that they must be analysed through simulations.

Calibration has a long tradition in empirical macroeconomics (Kydland and Prescott, 1996; Hansen and Heckman, 1996; Cooley, 1997; Gomme and Rupert, 2007). Its scope is to restrict the parameters of a model so that the latter is made consistent with empirical properties of the data (e.g., stylized facts about long-run growth, or moments of selected time series) or microeconomic observations. We follow in part this tradition but introduce the novel idea that, when the scope of the model is policy analysis, parameters' values should be selected so that the model reflects key properties of the causal structure underlying the data, where such properties are identified via a statistical identification approach, that is, under a minimal set of assumptions, not related to economic theory.

☆ We acknowledge financial support from the PRIN grant no. 20177FX2A7 of the Italian Ministry of University and Research. We are grateful to Fulvio Corsi, Cristiano Ricci, Raffaello Seri, as well as participants of the International Conference Computing in Economics and Finance (CEF 2023), the International Association for Applied Econometrics (IAAE 2022), the Workshop on Model Evaluation and Causal Search (Pisa 2022), the Workshop on Economic Science with Heterogeneous Interacting Agents (WEHIA 2022), the European Society for Ecological Economics (ESEE 2022) conference, the Statistics & Econometrics Seminar, Humboldt University, and the Insubria Economics seminars, for insightful comments. A special thank goes to Francesco Lamperti for his valuable help and advice in running the DSK model. Codes and data for replication can be retrieved following the instructions in the supplementary material.

* Correspondence to: Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa, Italy.
E-mail address: a.moneta@santannapisa.it (A. Moneta).

https://doi.org/10.1016/j.jebo.2024.106786
Received 21 September 2023; Received in revised form 10 October 2024; Accepted 15 October 2024
Available online 2 November 2024

Our idea of calibration has some overlapping with the strand of literature on estimation of dynamic stochastic general equilibrium (DSGE) models that involves the minimization of the distance between the impulse response functions of the models and the empirical impulse response functions (see, in particular, Christiano et al., 2005; Del Negro et al., 2007; Dridi et al., 2007; Hall et al., 2012; Guerron-Quintana et al., 2017). At odds with these studies, however, we do not rely on indirect inference or simulated minimum-distance. In fact, the method we propose should be classified as calibration, rather than an estimation. The key difference between the two notions lies in the fact that a model can be consistently estimated only when the model is identified, whereas calibration, which can be seen as a complementary tool, can be applied to non-identified (and misspecified) models.

Having discussed the term calibration, we now need to introduce validation. Validation is a notion that is used to address the following question: *How good is your model?* The assessment is *relative* when model's goodness is relative to other models and *absolute* when the model's performance is measured by fixing a unit of measure. The literature on DSGE modelling has devised important tools for comparing different models and evaluating the model's capacity of fitting data, mainly adopting a Bayesian approach. Comparisons of posterior marginal likelihoods and comparisons of the model's implied characteristics with a benchmark DSGE-Vector Autoregressive (VAR) model are prominent examples of relative and absolute validation tools, respectively (Del Negro et al., 2006; Cantore et al., 2013). The literature on agent-based models (ABMs) has also duly discussed the question of validation (see Windrum et al., 2007; Fagiolo et al., 2019), as the inherent complexity of these models poses a challenge in empirically validating them against observed data (Delli Gatti and Grazzini, 2020). Here, the emphasis has been posed on the idea that validation is about measuring the extent to which the data generating process (DGP) associated to the calibrated theoretical model is a good representative of the actual (''real-world'') DGP.

In the last decades, a large literature has emerged on calibration and estimation of complex simulation models, where key notions useful for validation have been discussed. We have mentioned above indirect inference (Gouriéroux et al., 1993; Smith, 1993) and simulated minimum-distance (Altissimo and Mele, 2009). Related approaches are the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989; Zila and Kukacka, 2023), simulated maximum likelihood method (Lee, 1992; Kristensen and Shin, 2012; Kukacka and Sacht, 2023), and approximate Bayesian computation (Grazzini et al., 2017; Frazier et al., 2018). Frameworks based on surrogate meta-models have also been developed (Lamperti et al., 2018b), which can address computational issues emerging from simulation and improve the performance of the above-mentioned methods.

In the present work, in the spirit of Guerini and Moneta (2017), we claim that not only calibration, but also validation should be designed by taking into account the adequacy for purpose of model building (Parker, 2020). If the objective is policy analysis, and, specifically, the prediction of the effect of a policy intervention on some variables of interest, a model should be considered ''valid'' by the extent of which the causal structure associated to the model's DGP matches the causal structure underlying the ''real-world'' DGP.

Therefore, our general approach necessarily hinges on tools for causal inference. Causal inference in macro-econometrics is intertwined with the discussion of identification of structural equation models (Hoover, 2012), which most economists see as plagued by the two famous critiques of Lucas (1976) and Sims (1980) (see Favero, 2001). We tackle here causal inference from a very ''agnostic'' perspective, in tune with the discussion of identification in structural vector autoregressive (SVAR) analysis (Kilian and Lütkepohl, 2017). For the sake of calibration and validation, we do not need, indeed, to identify a fully-fledged structural equation model. Nor is our scope to uncover the entire network of causal relationships among time series variables. We aim at identifying a set of structural shocks and how they impact a set of variables of interests.

We do this both for the model's and the ''real-world'' DGP: we estimate VAR models both from synthetic (i.e., generated by the model) and actual data and we identify the corresponding SVAR model by adopting a statistical identification approach. Specifically, local identification of the impact matrix is achieved by exploiting non-Gaussianity in the data, i.e., by applying independent component analysis (ICA) to SVAR modelling, as proposed by Moneta et al. (2013), Lanne et al. (2017), Gouriéroux et al. (2017) and Herwartz (2018). Our identification strategy is agnostic because, not only we do not rely on economic-theoretic restrictions, but also, differently from Guerini and Moneta (2017), we do not impose a recursive causal structure on the variables, which can be difficult to justify from an economic point of view. This comes, however, with a price, since we cannot perform shock labelling: an identified shock cannot be directly attributed to a specific variable (e.g., output). Nevertheless, by calculating a minimum distance index (MDI) between impact matrices, we show that it is possible to match shocks between the SVAR models derived from synthetic data and the ones derived from actual data.

This result is suited to our objective because, on the basis of the MDI, we can build a model confidence set (MCS) procedure (Hansen et al., 2011; Seri et al., 2021; Barde, 2020) that selects a set of model's configurations of parameters containing the most appropriate (*best*) one with a given level of confidence. In other words, MDI enters as loss function in MCS. This step allows us to achieve calibration of model's parameters that is consistent with causal analysis. Furthermore, by comparing the causal links between shocks and variables — the *shocks-variables structure* — associated with the calibrated configurations of parameters with the one derived by the actual data, we can propose an absolute measure of validation.

The proposed approach can be applied to any macroeconomic numerical simulation model, including, e.g., DSGE models, heterogeneous agent (HA) models and ABMs. The only requirements are that the theoretical model under scrutiny can be represented or adequately approximated by a state–space model and that both the data generated from the model and the actual data have non-Gaussian features (see full details below). We use our protocol to validate the ''Dystopian Schumpeter meeting Keynes'' (DSK) model by Lamperti et al. (2019a), which can be considered the first agent-based large-scale integrated-assessment model (IAM). IAMs try to formalize key processes at the intersection of the socio-economic and environmental system, with the aim of providing policy-relevant insights for decision-making. Indeed, the DSK model features a consumption-good, a capital-good and an energy sector

Table 2.1
Main notation and acronyms.

VAR	Vector autoregression	SHD	Structural hamming distance
SVAR	Structural vector autoregression	VM	Validation measure
DGP	Data generating process	ICR	Initial Independent component representation
ABM	Agent-based model	$\mathcal{M}^0$	Initial set of CoPs
DSGE	Dynamic stochastic general equilibrium	i	Index of CoPs
IAM	Integrated-assessment model	j	Index of Monte Carlo runs
DSK	Dystopian Schumpeter meeting Keynes	t	Index of time steps
ICA	Independent Component Analysis	θ_i	Vector of model's parameters
MDI	Minimum distance index	$\mathbf{y}_t$	Vector of observed time series
MCS	Model confidence set	$\mathbf{z}_{jt}(\theta_i)$	Vector of simulated time series
CoPs	Configuration of parameters	$\mathcal{M}^\star$	Set of selected CoPs
$\widehat{\boldsymbol{\Psi}}_0$	Matrix of real-world contemporaneous shocks (actual *mixing matrix*)	$\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$	Matrix of simulated contemporaneous shocks (simulated *mixing matrix*)

which jointly contribute to the Carbon Dioxide emissions process accounting for climate-induced damages. ABMs are interesting candidates for our approach as it hinges on non-Gaussianity, which is a common feature in data generated by these models.[1]

The contributions of this paper can be summarized as follows: first, we introduce a general protocol that, in subsequent steps, can perform both calibration and validation. We spell out its theoretical underpinnings based on SVAR-ICA, MDI, and MCS. The proposed method turns out to be faster than other procedures based on optimization or the exploration of the parameter space and reduces the risk to deviate from the (pseudo-)true values (e.g., the probability of incurring in multiple local minima, tipping points or flat regions of the objective function). Second, we propose a novel employment of a statistical (i.e., data-driven) identification procedure in the context of calibration and validation. We show that in such context, differently from other settings, lack of global identification does not create any hurdle. Furthermore, the fact that the proposed procedure allows recovering the shocks-variable structure — attempting to open the causality black box — constitutes an advantage with respect to approaches based on predictive ability. Third, we present an implementation of MCS which allows the possibility of ranking model's causal structures from the most to the least plausible. Notice that MCS is also in tune with the data-driven approach, as it focuses on the informativeness of the real-world data (Hansen et al., 2011; Seri et al., 2021). Moreover, our empirical application contributes to the literature that attempts at calibrating and validating IAMs. Agent-based IAMs, in particular, seem to offer a new paradigm for the assessment of climate-induced outcomes and climate policy (Lamperti et al., 2019b; Castro et al., 2020; Lamperti and Roventini, 2022). However, a unified protocol for calibration and validation is still missing. It is worth noting that our application identifies, in the model, a set of shocks that hit energy and investment and match quite accurately the empirical counterpart found in U.S. data.

The rest of the paper is structured as follows. In Section 2, we summarize the different steps involved in our calibration and validation technique. In Section 3, we introduce the statistical framework and we provide the SVAR representation for both the model and the actual data. In Section 4, we present our general protocol of calibration and validation. In particular: in Section 4.1, we describe the SVAR-ICA approach to identification; in Section 4.2, we discuss the MDI used as loss function in the MCS; in Section 4.3, we describe the MCS-based calibration procedure; in Section 4.4, we discuss the validation step. In Section 5, we briefly illustrate the DSK model. This model is calibrated and validated by applying our general protocol in Section 6. Section 7 concludes.

2. Sketch of the protocol

In Table 2.1, we report the different acronyms with their meaning, alongside the main notation, used throughout the paper. Below, we summarize our general protocol for calibration and validation. Two steps (1–2) can be seen as preliminary:

1. Select a discrete set $\mathcal{M}^0 := \{1, \dots, m_0\}$ of configurations of parameters (henceforth, CoPs) from the parameter space of the theoretical model object of the study. A vector of parameters θ_i $(i = 1, \dots, m_0)$ is associated to each CoP. From the same model, for each CoP i, simulate n Monte Carlo runs, denoted by $\mathbf{z}_{jt}(\theta_i)$, with $j = 1, \dots, n$ and $t = 1, \dots, T$. The vector $\mathbf{z}_{jt}(\theta_i)$ is $K-$dimensional.
2. Select a $K \times 1$ vector $\mathbf{y}_t$ of observed time-series macroeconomic data, with $t = 1, \dots, \tau$. We refer to $\mathbf{y}_t$ as the real-world or actual data. Estimate a reduced-form VAR model both from $\mathbf{z}_{jt}(\theta_i)$ (for each i and j) and $\mathbf{y}_t$.

The next 3 steps (3–5) refer to calibration:

3. For each estimated VAR model, estimate the impact matrix, i.e., the matrix that describes the contemporaneous impact of the shock on the variable of interest. This matrix, which we refer to as the *mixing matrix*, is locally identified by ICA applied to the VAR residuals. We call $\widehat{\boldsymbol{\Psi}}_0$ the mixing matrix estimated from $\mathbf{y}_t$, and $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ the one estimated from $\mathbf{z}_{jt}(\theta_i)$.
4. Calculate the MDI between $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ and $\widehat{\boldsymbol{\Psi}}_0$ and record the unique signed-permutation matrix $\mathbf{C}_{ji}$ associated to it, for each j and i.

[1] Examples of DSGE models characterized by non-Gaussian shocks can be found in An and Schorfheide (2007) and Cúrdia et al. (2014).

5. Apply the MCS using the MDI as loss function and select the set $\mathcal{M}^\star$ of CoPs that minimizes the expected loss. The selected CoPs are statistically indistinguishable given a level of confidence.

The last 2 steps (6–7) concern only validation:

6. For each $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)\mathbf{C}'_{ji}$, with $i \in \mathcal{M}^\star \subseteq \mathcal{M}^0$, test the significance of its entries, exploiting distributions obtained by Monte Carlo simulations across j. In a similar manner, test the significance of the entries of $\widehat{\boldsymbol{\Psi}}_0$ via bootstrap. For each CoP and the actual data, infer a causal structure (which we call independent component representation) representing the significant influences from shocks to variables.
7. Compare the shocks-variables structure associated to CoPs $i \in \mathcal{M}^\star \subseteq \mathcal{M}^0$ to the "real-world" shocks-variables structure, using a validation measure (VM) based on the Structural Hamming Distance (SHD). SHD measures how many entries of the matrices representing the two structures do not coincide.

3. SVAR representation

Our method moves from the assumption that both the stochastic process underlying a set of observed macroeconomic data (what we call the "real-world" DGP) and the process underlying a macroeconomic simulation model (model DGP) can be approximated by a SVAR model.

Example 1 (*DSGE Representation*)**.** A DSGE model can be represented by a reduced-form VAR following the conditions devised in Fernández-Villaverde et al. (2007) and Ravenna (2007).

Example 2 (*ABM Representation*)**.** Analogously, the relationship between an ABM and a SVAR (see, e.g., Guerini and Moneta, 2017 and Delli Gatti and Grazzini, 2020) can be justified by the fact that an ABM can be approximated by a state–space model (Hinkelmann et al., 2011), and the latter can in turn be approximated by a finite-order VAR model (Giacomini, 2013).[2] Examples of state–space representation are common in the data assimilation literature (see Ward et al., 2016). Moreover, Gusella and Ricchiuti (2024) introduce a formal framework for the state–space representation of heterogeneous interacting agents models.

In the following, we consider a set of K time series variables $\mathbf{y}_t = (y_{1t}, \dots, y_{Kt})'$, corresponding to a set of K observed macroeconomic variables and a set of K time series variables $\mathbf{z}_{jt}(\theta_i)$ corresponding to a set of data generated by a (simulated) theoretical model for vector of parameters θ_i and Monte Carlo run j ($i = 1, \dots, m_0$; $j = 1, \dots, n$).

The process generating $\mathbf{y}_t$ is represented by the following SVAR model:

$$\boldsymbol{\Gamma}_0\mathbf{y}_t = \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_P\mathbf{y}_{t-P} + \boldsymbol{\varepsilon}_t \tag{3.1}$$

where $\boldsymbol{\Gamma}_p$ (for lag $p = 0, \dots, P$) are $K \times K$ matrices denoting the contemporaneous and lagged structural coefficients, and $\boldsymbol{\varepsilon}_t$ is a K-dimensional vector of i.i.d. structural error terms (or shocks) with covariance matrix $\boldsymbol{\Sigma}_\varepsilon$, which we assume to be diagonal. Eq. (3.1) may also contain a constant (or even a deterministic trend), which we omit here for convenience, not being relevant for the present discussion. This model can be rewritten in a form that omits contemporaneous causality. This is the reduced-form VAR model, which turns out to be more convenient for estimation:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_P\mathbf{y}_{t-P} + \mathbf{u}_t \tag{3.2}$$

where $\mathbf{A}_p = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_p$ ($p = 1, \dots, P$), and $\mathbf{u}_t = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t$, i.e. $\mathbf{u}_t$ is a vector of i.i.d. processes with covariance matrix $\boldsymbol{\Sigma}_\mathbf{u} = \mathbb{E}\{\mathbf{u}_t\mathbf{u}_t'\} = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Gamma}_0^{-1,\prime}$. We call the impact matrix $\boldsymbol{\Psi}_0 = \boldsymbol{\Gamma}_0^{-1}$ the *real-world mixing matrix*.

Analogous representation holds for data generated by the simulation model:

$$\boldsymbol{\Gamma}_{j,0}(\theta_i)\,\mathbf{z}_{j,t}(\theta_i) = \boldsymbol{\Gamma}_{j,1}(\theta_i)\,\mathbf{z}_{j,t-1}(\theta_i) + \cdots + \boldsymbol{\Gamma}_{j,P}(\theta_i)\,\mathbf{z}_{j,t-P}(\theta_i) + \boldsymbol{\varepsilon}_{jt}(\theta_i) \tag{3.3}$$

$$\mathbf{z}_{jt}(\theta_i) = \mathbf{A}_{j,1}(\theta_i)\,\mathbf{z}_{j,t-1}(\theta_i) + \cdots + \mathbf{A}_{j,P}(\theta_i)\,\mathbf{z}_{j,t-P}(\theta_i) + \mathbf{u}_{jt}(\theta_i), \tag{3.4}$$

where $\boldsymbol{\varepsilon}_{jt}(\theta_i)$ and $\mathbf{u}_{jt}(\theta_i)$ are the model's shocks and the reduced-form residuals, respectively. We call the impact matrix $\boldsymbol{\Psi}_{j,0}(\theta_i) = \boldsymbol{\Gamma}_{j,0}^{-1}(\theta_i)$ the *model mixing matrix* associated to the jth Monte Carlo run of the ith CoP of the simulated model. As well known in the SVAR analysis, the mixing matrix is key for identification.

It is important to notice that the shocks $\boldsymbol{\varepsilon}_{jt}(\theta_i)$ may have different interpretations and origins, depending on whether the SVAR model is used for representing an ABM or, rather, a DSGE model. In a DSGE model there is a close correspondence between the structural shocks referring to the equations of the linearized theoretical model and the structural shocks obtained from the VAR model fitted on the simulated data. Instead, if we fit a VAR model on data generated by an ABM, the SVAR model's identified shocks represent macro-level forces resulting from the aggregation of micro-level, idiosyncratic shocks. Examples of micro-level shocks include idiosyncratic and heterogeneous technological improvements that firms may enjoy, unexpected cuts

[2] Giacomini (2013) specifies that the approximation of a DSGE in terms of a VAR model proceeds in three stages: (i) from DSGE to state–space model; (ii) from state–space model to VAR(∞); (iii) from VAR(∞) to finite-order VAR. For our purposes, we focus on the second and third steps: as long as the ABM can be approximated be a state–space representation, the latter can be carved into a finite-order VAR.

in credit supply, unforeseen revisions in investment plans. Moreover, as in Delli Gatti and Grazzini (2020) and, more generally, in macro ABM standard practice, the modeller can externally introduce unanticipated policy shocks during the simulation to see how the contemporaneous causal structure embedded in the model dynamically unfolds in the system. This is rather distinct with respect to DSGE models, where equilibrium conditions — which theoretically determine the contemporaneous causal structure of the model — are externally hit by shocks whose correspondence with VAR model's innovations is close by construction (making their interpretation easier and straightforward). On this matter, for the purpose of complex-simulation models' calibration and validation, it becomes of interest to assess, on the basis of the statistical identification approach that we pursue, whether there is correspondence between the shocks derived by the SVAR model that represents the ABM and the shocks identified from the real-world data.

4. Calibration and validation protocol

We now enter the core of our calibration-validation procedure. Moving from the (S)VAR representability of our DGPs, in this section we provide the theoretical background for the steps 3–7 of Section 2.

4.1. The SVAR-ICA approach to identification

Our general protocol is based on a comparison between the SVAR models estimated from the synthetic data (one for each Monte Carlo run) and the one derived from the actual data. Thus, there is a problem of identification to be faced. We adopt here a data-driven approach to identification, which allows us to avoid strong a priori restrictions (e.g., theoretical short-run or sign restrictions). Specifically, we use independent component analysis, which exploits non-Gaussianity. With this approach, we obtain local identification but, as we will explain in the next subsection, our index of comparison between SVAR models remains invariant to lack of global identification.

ICA is a statistical method that models a set of observed random variables as a linear combination of independent latent random variables, called the independent components (Comon, 1994; Hyvärinen et al., 2001). In line with the applications of ICA to SVAR analysis (see, for instance, Moneta et al., 2013; Gouriéroux et al., 2017; Lanne et al., 2017; Herwartz, 2018), the input data are the estimated reduced-form residuals $\mathbf{u}_t$ (or $\mathbf{u}_{jt}(\theta_i)$) and the latent independent components are the structural shocks $\boldsymbol{\varepsilon}_t$ (or $\boldsymbol{\varepsilon}_{jt}(\theta_i)$).

Given that $\mathbf{u}_t = \boldsymbol{\Psi}_0 \boldsymbol{\varepsilon}_t$, the ICA model recovers (up to some indeterminacy, see below) $\boldsymbol{\Psi}_0$ and $\boldsymbol{\varepsilon}_t$ from realizations of $\mathbf{u}_t$ under the assumptions that the components $\boldsymbol{\varepsilon}_t$ are non-Gaussian (with at most one exception) and that $\boldsymbol{\Psi}_0$ is invertible (see, e.g., Hyvärinen et al., 2001 and Hyvärinen, 2013). Notice that the same assumptions hold for $\boldsymbol{\varepsilon}_{jt}\left(\theta_i\right)$ and $\boldsymbol{\Psi}_{j,0}\left(\theta_i\right)$. The indeterminacy is related to the fact that ICA identifies $\boldsymbol{\Psi}_0$ up to the post-multiplication of a generalized permutation matrix $\mathbf{DP}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix.[3] This means that the order and scale of the shocks are not identified. Our choice of the minimum-distance index allows us to tackle this issue.[4]

In the ICA literature, many methods have been developed to estimate $\boldsymbol{\Psi}_0$ from $\mathbf{u}_t$. Some of them are based on the minimization of a contrast function whose argument is a vector of parameters ω determining the rotation angles of the orthogonalized input data. The method based on the minimization of the Cramér–von-Mises statistics proposed by Herwartz and Plödt (2016) and the method based on distance covariance developed by Matteson and Tsay (2017) use this approach. Another established technique considers semi-parametric estimators of the pseudo-maximum likelihood function (Gouriéroux et al., 2017). A different approach exploiting information theory techniques has been developed by Hyvärinen (1999) and Hyvärinen and Oja (2000). The authors formulate a fixed-point algorithm called *fastICA*. This technique relies on the maximization of the non-Gaussianity of $\boldsymbol{\gamma}_k' \mathbf{u}_t$, where $\boldsymbol{\gamma}_k'$ is the kth row of the matrix $\boldsymbol{\Gamma}_0$, for each $k = 1, \ldots, K$. Moneta and Pallante (2022) provide a performance evaluation study comparing *fastICA* with other ICA estimators, showing its relative robustness and reliability in a SVAR setting. We therefore choose to adopt *fastICA* algorithm to estimate and identify our SVAR-ICA model. In Appendix A, we state the assumptions underlying the ICA model and the *fastICA* estimator.

4.2. Minimum distance index

We present here the minimum distance index, which allows us to calculate the distance between impact matrices identified by ICA, tackling the issue of the scale/order indeterminacy. The MDI is inspired by Matteson and Tsay (2017), who suggest to measure the error between the estimate $\widehat{\boldsymbol{\Psi}}_0$ and the true value $\boldsymbol{\Psi}_0$ exploiting the metric proposed by Ilmonen et al. (2010). Here, instead, we want to measure the discrepancy between the model mixing matrix and the real-world mixing matrix. The index finds the shortest discrepancy by searching across all the possible column-signed permutations of the model mixing matrix, by keeping the real-world mixing matrix as reference matrix.[5] In other words, the MDI is invariant to all possible column's permutations and changes of sign of the estimated model mixing matrix. To simplify the notation, in the following we write $D_{ji} := D\left(\widehat{\boldsymbol{\Psi}}_{j,0}\left(\theta_i\right), \widehat{\boldsymbol{\Psi}}_0\right)$.

[3] We recall that a generalized permutation matrix is a matrix that has exactly one non-zero element in each row and each column.

[4] It is customary to normalize SVAR models so that the structural shocks have unit standard deviations, meaning that impulse response functions refer to one standard-deviation shock. In this manner, the scale problem is resolved (this normalization involves a re-scaling of the columns of the mixing matrix), but not completely because the sign of shocks (or of their impacts) remains undetermined. Therefore, one can conclude that $\boldsymbol{\Psi}_0$ is identified up to the post-multiplication of a signed permutation matrix $\mathbf{JP}$ (where $\mathbf{J}$ is a sign-change matrix, i.e. a diagonal matrix with only +1 or −1 entries on the main diagonal, and $\mathbf{P}$ is a permutation matrix).

[5] In our application of the procedure, for convenience, the columns of the real-world mixing matrix are signed-permuted by applying the *Maxfinder* criterion in a hierarchical manner, as proposed by Bruns et al. (2021). However, results are not sensitive to any signed-permutation of the columns of the matrix $\widehat{\boldsymbol{\Psi}}_0$.

Definition 1. The minimum-distance index for $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ is:

$$D_{ji} := \frac{1}{\sqrt{K-1}} \inf_{\mathbf{C}_{ji} \in \mathcal{C}} \left\| \mathbf{C}_{ji} \widehat{\boldsymbol{\Psi}}_{j,0}^{-1}(\theta_i) \widehat{\boldsymbol{\Psi}}_0 - \mathbf{I}_K \right\|_F \tag{4.1}$$

where

$$\mathcal{C} = \left\{ \mathbf{C}_{ji} \in \mathcal{G} : \mathbf{C}_{ji} = \mathbf{P}_{ji}\mathbf{J}_{ji} \text{ for some } \mathbf{P}_{ji} \text{ and } \mathbf{J}_{ji} \right\},$$

$\mathcal{G}$ is the set of full-rank $K \times K$ matrices, $\mathbf{P}_{ji}$ is a permutation matrix, $\mathbf{J}_{ji}$ is a sign-change matrix, $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ is the estimator of the model mixing matrix $\boldsymbol{\Psi}_{j,0}(\theta_i)$, $\widehat{\boldsymbol{\Psi}}_0$ is the estimator of the real-world mixing matrix $\boldsymbol{\Psi}_0$ (from real data), $\mathbf{I}_K$ is the identity matrix and $\|\cdot\|_F$ is the Frobenius norm. When the value of D_{ji} approaches 0, we have that $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ is close to $\widehat{\boldsymbol{\Psi}}_0$.

Since it implies the minimization over all choices $\mathbf{C}_{ji} \in C$, D_{ji} seems to require high computational costs, especially when the number of variables K increases. However, this is not a real drawback in our case. First, VAR models that are usually treated in the macroeconomic literature considers a limited number of variables (typically $K < 10$). Second, we compute the MDI following the two-steps procedure described by Ilmonen et al. (2010, pp. 234–235), which reduces the optimization problem over all permutation matrices $\mathbf{P}_{ji}$ of Eq. (4.1) to a linear programming problem that can be solved using specific algorithms (e.g., the Hungarian method).

4.3. Model confidence set

We now present our calibration procedure, which is based on the Model Confidence Set. MCS is a statistical procedure that allows the researcher to find the best CoPs, with a given level of confidence, among a discrete set of candidates (Hansen et al., 2011). To perform this selection, the researcher needs to specify a loss function, a selection criterion, and an elimination rule. Since our purpose is to select the set of CoP(s) delivering causal structures that match as close as possible the structure underlying the actual data, we use the MDI as loss function.

From the set of CoPs $\mathcal{M}^0$, MCS selects a set $\mathcal{M}^*$ with cardinality greater or equal than one. We recall that to each CoP is associated a vector of parameters (to be calibrated) θ_i, for $i = 1, \dots, m_0$. For each CoP i: (i) we run n Monte Carlo simulations $\mathbf{z}_{jt}(\theta_i)$ ($j = 1, \dots, n$); (ii) we derive the model mixing matrix $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ (for each Monte Carlo j); and (iii) we compute the MDI between $\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i)$ and the real-world mixing matrix $\widehat{\boldsymbol{\Psi}}_0$ (for each j).

Let $\overline{D}_i := \mathbb{E}_{\widehat{\boldsymbol{\Psi}}_0(\theta_i)} D\left(\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i), \widehat{\boldsymbol{\Psi}}_0\right)$ be the expected MDI relative to CoP i, where the expectation term is taken over the values that the estimated model mixing matrix takes across Monte Carlo runs. Let $\overline{\mathbf{D}} := (\overline{D}_1, \dots, \overline{D}_{m_0})'$ be the m_0-dimensional vector of these expected values for the m_0 CoPs. Let $\overline{D}_i^{(n)} := \frac{1}{n}\sum_{j=1}^n D_{ji}$ and $\overline{\mathbf{D}}^{(n)} := \left(\overline{D}_1^{(n)}, \dots, \overline{D}_{m_0}^{(n)}\right)'$ be the sample counterparts of $\overline{D}_i$ and $\overline{\mathbf{D}}$ respectively. Defining $\mathbf{D}_j := \left(D_{j1}, \dots, D_{jm_0}\right)'$, the sample average distance can be rewritten as $\overline{\mathbf{D}}^{(n)} := \frac{1}{n}\sum_{j=1}^n \mathbf{D}_j$.

We aim at finding the CoPs achieving the minimal MDI. Let

$$\mathcal{M}^\star := \left\{ h \in \mathcal{M}^0 : \overline{D}_h = \min_{i \in \mathcal{M}^0} \overline{D}_i \right\} \tag{4.2}$$

be the set of parameters minimizing the distance $\overline{D}_i$. For $i \in \mathcal{M}^0$, the estimator $\hat{i}^{(n)}$ is the value that minimizes the sample average distance $\overline{D}_i^{(n)}$. Note that $\hat{i}^{(n)}$ is a singleton while $\mathcal{M}^*$ is not necessarily so.

To achieve a given level of confidence in the selection procedure, we need to formulate a statistical test. To this aim, we estimate, via Gaussian quasi-likelihood, $\overline{D}_i$ and $\sigma_i^2 := \mathbb{V}_{\widehat{\boldsymbol{\Psi}}_0(\theta_i)} D\left(\widehat{\boldsymbol{\Psi}}_{j,0}(\theta_i), \widehat{\boldsymbol{\Psi}}_0\right)$, following Seri et al. (2021). To do that, $\mathbf{D}_j$ must be independent and identically distributed and, for each $i \in \mathcal{M}^0$, the distances D_{ji} must be independent. Moreover, the mean $\mathbb{E}D_{ji}$ must exists and be finite for each $i \in \mathcal{M}^0$. These requirements guarantee the consistency and measurability of $\hat{i}^{(n)}$. These properties derive directly by the fact that each simulation $\mathbf{z}_{jt}(\theta_i)$ is independent across Monte Carlo runs. Moreover, fixed $i \in \mathcal{M}^0$, $\mathbf{z}_{jt}(\theta_i)$ are identically distributed (see also Choirat and Seri, 2012, Proposition 1, p. 280). At last, we need finite σ_i^2. All these requisites are either fulfilled by the construction of the simulation model or verified in the data. Therefore, we can use standard statistical hypothesis testing to test m_0 restrictions of the model at the same time.

We define the equivalence test $\delta_{\mathcal{M}}$ and the selection rule $e_{\mathcal{M}}$ associated to the set $\mathcal{M} \subseteq \mathcal{M}^0$. The test has a null $\mathsf{H}_{0,\mathcal{M}}$ and an alternative hypothesis $\mathsf{H}_{1,\mathcal{M}}$:

$$\mathsf{H}_{0,\mathcal{M}} : \overline{D}_i = \overline{D}_h, \forall i, h \in \mathcal{M}; \tag{4.3}$$

$$\mathsf{H}_{1,\mathcal{M}} : \exists i, h \in \mathcal{M} \text{ such that } \overline{D}_i \neq \overline{D}_h. \tag{4.4}$$

If the test rejects the null hypothesis, then $\delta_{\mathcal{M}} = 1$, else $\delta_{\mathcal{M}} = 0$. When $\delta_{\mathcal{M}} = 1$ we use $e_{\mathcal{M}} := \arg\max_{h \in \mathcal{M}} \overline{D}_h^{(n)}$ to remove a CoP from $\mathcal{M}$ (i.e., we select the index $h \in \mathcal{M}$ which provides the largest value $\overline{D}_h^{(n)}$). Now, we introduce the sequence of subsets of $\mathcal{M}^0$, $\mathcal{M}_{i+1} = \mathcal{M}_i \setminus e_{\mathcal{M}_i}$ for $i = 1, \dots, m_0 - 1$, and the p-values of the test procedure $p_{\mathsf{H}_{0,\mathcal{M}_i}}$, where we impose that $p_{\mathsf{H}_{0,\mathcal{M}_{m_0}}} \equiv 1$. Therefore, the MCS p-value can be defined as follows:

$$\widehat{p}_{e_{\mathcal{M}_h}} := \max_{i \leq h} p_{\mathsf{H}_{0,\mathcal{M}_i}}, \tag{4.5}$$

for $h = 1, \dots, m_0$. The algorithm for the implementation of the MCS is reported in Appendix B.

Note that our MCS-based calibration procedure can be easily adapted to a MDI which refers not just to the mixing matrices, i.e. $\boldsymbol{\Psi}_0$, $\boldsymbol{\Psi}_{j,0}\left(\theta_i\right)$, but rather to structural moving-average matrices at different lags, $\boldsymbol{\Psi}_\ell$, $\boldsymbol{\Psi}_{j,\ell}\left(\theta_i\right)$ (with $\ell = 1, \ldots, H$). However, we focus on the former matrices, since the identification of the latter depends on the mixing matrices, which ICA is able to locally identify in a data-driven fashion. Therefore, the comparison of structural matrices at time horizons greater than zero does not provide additional information about structural identifiability.

4.4. Validation step

Once the MCS-based calibration is performed, it is possible to investigate the behaviour of the causal structures associated to the CoPs which pass the test. By comparing such behaviour with the causal structure associated with the real-world DGP, we propose a measure of model validation. Such measure fulfils two desirable criteria. First, it is a measure that is bounded by construction between zero and one. Thus, it delivers an absolute assessment and can be used to compare the performance of models of different nature. Second, it focuses on properties of the causal structures that both are significant from a statistical point of view and can be inferred from the data without further theoretical restrictions.

Notice that we have only partial information about the causal structure underlying the real-world DGP, which is our reference point. From ICA, as already pointed out, we get an estimate of $\boldsymbol{\Psi}_0$ which is underdetermined by permutations and changes of sign of its columns. Thus, since we do not want to impose further restrictions, we do not obtain labels of shocks, i.e., we cannot relate shocks with variables. But, by bootstrap, we can recover the causal structures between the (mutual independent) real-world shocks and variables, by inferring which shocks $\varepsilon_{k_s,t}$ (with $k_s = 1, \ldots, K$) have significant impacts on variables $\mathbf{y}_t = (y_{1,t}, \ldots, y_{K,t})'$. From the simulated data, we get an estimate of $\boldsymbol{\Psi}_{j,0}\left(\theta_i\right)$, which contains in principle the same column/sign indeterminacy. However, the calculation of the MDI D_{ji} has provided a unique matrix $\mathbf{C}_{ji}$ for each j and i. From $\widehat{\boldsymbol{\Psi}}_{j,0}\left(\theta_i\right)\mathbf{C}'_{ji}$, we get a one-to-one mapping between the impacts of the simulated shocks and the impacts of the real-world shocks. This warrants the possibility of comparing the real-world shocks-variables structures with the model's shocks-variables structures, which can be inferred by exploiting the Monte Carlo simulations.

We represent the shocks-variables structure via a matrix called "independent component representation" (ICR) (see Casini et al., 2021, for a graph-theoretic definition). An ICR is a $K \times K$ matrix whose entries are zeros or ones. The entry $\langle k_v, k_s \rangle$ is 1 if and only if there is a significant impact of the shock $\varepsilon_{k_s,t}$ (or $\varepsilon_{k_s,t}\left(\theta_i\right)$) on the variable $y_{k_v,t}$ (or $z_{k_v,t}\left(\theta_i\right)$), for $k_v, k_s = 1, \ldots, K$.

Whether an impact is significant or not is based on significance tests on the coefficients which enter in the matrices $\widehat{\boldsymbol{\Psi}}_0$ and $\widehat{\boldsymbol{\Psi}}_{j,0}\left(\theta_i\right)\mathbf{C}'_{ji}$, with $i \in \mathcal{M}^* \subseteq \mathcal{M}^0$. As regards $\widehat{\boldsymbol{\Psi}}_0$, the significance tests is based on the wild bootstrap procedure (see Kilian and Lütkepohl, 2017, Sec. 12.2.3): at each bootstrap iteration n^* ($n^* = 1, \ldots, N^*$), the bootstrap-estimated mixing matrix $\widehat{\boldsymbol{\Psi}}_0^{n^*}$ is right-multiplied by a signed permutation matrix $\mathbf{C}'_{n^*}$, where $\mathbf{C}_{n^*}$ corresponds to the *arg inf* of the MDI between $\widehat{\boldsymbol{\Psi}}_0^*$ and $\widehat{\boldsymbol{\Psi}}_0$ (consistently to the scheme we apply to the model mixing matrices).

Once we have obtained ICRs for both synthetic and real data, we calculate the Structural Hamming Distance to be used in the proposed validation measure. SHD originates from information theory and is generally used to compare the similarity of blocks of words of equal length. In the field of causal networks, SHD has been introduced by Acid and de Campos (2003) and Tsamardinos et al. (2006) to confront directed acyclic graphs.

Let ICR_{rw} be the ICR representing the real-world shocks-variables structure and ICR_{sim} the analogous structure for the model's shocks and variables. We adapt SHD such that it counts how many entries of the two matrices do not coincide. We define our validation measure as follows:

Definition 2. The validation measure of ICR_{sim} with respect to ICR_{rw} is:

$$\text{VM} := 1 - {}^{SHD}/_{K^2}, \tag{4.6}$$

where K^2 is the number of entries in each ICR.

If SHD $\to 0$, then VM $\to 1$. For a given K, the smaller is SHD, the closer is, under this interpretation, the model's causal structure to the real-world causal structure. In particular, SHD counts, in a theory-free fashion, how many edges differ between ICR_{rw} and ICR_{sim}. Even though the resulting validation measure does not distinguish between matching due to common presence or common absence of edges, it is helpful in providing a synthetic score of how good is a model in capturing a reference causal structure. Finally, this measure is alternative to the measures proposed by Guerini and Moneta (2017) (namely, sign-based, size-based and conjunction measures) as it is both general and more in tune with the literature on causal search.[6]

[6] It may be argued that using the same empirical data both to calibrate and to validate the model can yield a sort of "double-counting" and can raise issues of circularity. Steele and Werndl (2013) show that in fact double-counting is not problematic, building on a Bayesian theory of confirmation, but also showing its legitimacy in a frequentist approach. Note that Steele and Werndl (2013) develop their argument for the calibration-confirmation problem, but that this holds *a fortiori* for calibration-validation, since our validation step is merely based on a distance between causal structures.

Table 6.1
Parameter notations and interval values.

Description	Parameter	Values
Beta distribution support (innovation)	$\underline{x}$	$[-0.15, -0.05]$
	$\overline{x}$	$[0.05, 0.15]$
Beta distribution support (energy)	$\underline{x}_{\text{en}}$	$[-0.1, -0.01]$
	$\overline{x}_{\text{en}}$	$[0.01, 0.1]$
Firm search capabilities parameters	$\zeta_{1,2}$	$[0.3, 1]$
Payback period (industrial)	b	$[2, 3.75]$
Consumption-good firm initial mark-up	$\overline{\mu}_0$	$[0.2, 0.3]$
Mark-down for bank deposits	μ^{dep}	$[0.75, 1]$
Mark-down on the bank reserves at Central Bank	μ^{res}	$[0.5, 0.9]$

5. The DSK model

In this section, we briefly illustrate the DSK model by Lamperti et al. (2019a), which is the object of our application. The DSK family of models represents the first attempt to provide an agent-based integrated assessment model, in the spirit of contributions in environmental economics (Weyant, 2017), as it combines energy, climate and economic modelling to offer an integrated perspective on emission trajectories, decarbonization pathways and the corresponding policies to implement. It has been recently used to study scenarios under which green transitions are more likely to occur (Lamperti et al., 2020) and to analyse the public costs of climate-induced financial instability (Lamperti et al., 2019a), as well as to evaluate financial policies aimed at dealing with increasing climate risks. In particular, DSK models allow tackling several problems that plague traditional general-equilibrium integrated assessment models, by enhancing the degree of heterogeneity, improving the representation of radical uncertainties, refining the technological change process, and obtaining an accurate assessment of climate scenarios (Stern and Stiglitz, 2021).

The DSK model by Lamperti et al. (2019a) features a manufacturing sector, populated by heterogeneous and interacting firms, devoted to the production of either capital or consumption goods and receiving inputs from an energy sector. The financial system is represented by a banking sector in which banks — heterogeneous in number of clients, balance-sheet structure, and lending conditions — decide the amount of credit to provide to their clients subject to a capital requirement and leverage conditions. The energy sector is populated by heterogeneous plants embracing different energy generation technologies ("clean" and "dirty") which possess diverse cost structures and emission intensities. Moreover, it is characterized by an exogenous fossil fuel sector which provides dynamics and boundary conditions (reflecting scarcity) on the price of an undifferentiated fossil fuel. The production activities of energy and manufacturing firms lead to CO2 (equivalent) emissions, which increase temperature in a nonlinear way. Technical change occurs both in the manufacturing and energy sectors. Capital-good firms develop new vintages of machines that are both more productive and more "green". The energy sector can improve both the "brown" and "green" energy generation technologies. Innovation determines the cost of energy produced by dirty and green technologies, which, in turn, affect the energy-technology production mix and the total amount of CO2 emissions. Finally, the government sector collects taxes on profits and pays unemployment benefits. A detailed description of the model is provided in Section 1 of the Supplementary Material.

Our approach to calibration and validation puts emphasis on the ability of the model to deliver empirically reliable causal structures concerning the real side of the economy, the energy sector, and climate-related outcomes. The DSK model (Lamperti et al., 2019a), as well as its two-sector predecessor, the "Schumpeter meeting Keynes" (K+S) model (Dosi et al., 2010, 2013, 2015), have undertaken a wide, yet manual, exploration of the parameter space that ensures, at a macro level, reasonable and stable long-run economic trajectories, along with endogenous business cycle dynamics. In particular, the DSK version that we analyse in this paper (Lamperti et al., 2019a) has been indirectly calibrated by the authors to reproduce stylized facts such as energy demand and CO2 emission patterns that are consistent with the shared socioeconomic pathway 5 (SSP5, see Riahi et al., 2017)

For this reason, the K variables of interest are: aggregate output (GDP), consumption ($Cons$), investments (Inv), unemployment rate (UR), a price index (CPI), demand of energy ($Ener$), and total emissions of Carbon Dioxide ($Emiss$).

6. Application of the general protocol

In this section, we show the results of the application of the general protocol for calibration and validation, presented in Section 4, to the DSK model illustrated in Section 5.

The starting point is the choice of the discrete set of parameters to be calibrated. This choice hinges on the detection of those features that have the highest influence on the behaviour of the macroeconomic output (see Lamperti et al., 2019a). In Table 6.1, we summarize the selected parameters and the intervals in which they vary.

The range of variation of the parameters is defined considering previous experiments (see Lamperti et al., 2019a, and references therein). We refer to Section 1 of the Supplementary Material (SM) for details regarding the description and the main equations of the model.

In the spirit of Dosi et al. (2010) and Lamperti et al. (2018a), the support of the *Beta* distribution ($\underline{x}$ and $\overline{x}$) models the notional possibilities of technological advance and the relative frequency of (un)successful innovations. This translates into the technological opportunities for capital-good firms to improve upon the productivity, energy efficiency and environmental friendliness of the machineries produced. The parameter $\underline{x}$ is the lower bound of the support of the *Beta* distribution and it varies between −0.15 and

-0.05, while $\overline{x}$ is the upper bound of the support and it ranges between 0.05 and 0.15. The parameters $\underline{x}$ and $\overline{x}$ are symmetric, hence when $\underline{x} = -0.15$, $\overline{x} = 0.15$ (SM, Section 1.1). Similarly, in the energy sector the support of the *Beta* distribution ($\underline{x}_{en}$, $\overline{x}_{en}$) dictates the likelihood of building a new green or dirty plant. Differently from previous indirect calibration procedures, we allow for the possibility for the capital-good and energy sector to have different access to innovation opportunities. Accordingly, $\underline{x}_{en}$ varies between -0.1 and -0.01, while $\overline{x}_{en}$ varies between 0.01 and 0.1 (SM, Section 1.3). To a similar extent, the search capabilities parameters ($\zeta_{1,2}$) influence the growth engine of the model by enlarging (or restricting) the possibilities of accessing "innovations", regardless they constitute an improvement or not. In particular, higher values of these parameters imply a lower degree of technological asymmetry, as all firm access more easily to innovations, thus easing the effects of the Schumpeterian "creative destruction" (Dosi et al., 2010). The firm search capabilities parameters $\zeta_{1,2}$ vary between 0.3 and 1 (SM, Section 1.1). Finally, the process of technical change transmits to downstream consumption-good firms when they invest in more productive machineries. Thus, the frequency at which consumption-good firms replace machinery equipment, matters for both growth and business cycle dynamics. This choice is determined following a payback period routine. The parameter b influence this decision as firms compare newly available machineries to the set of the existing ones in terms of prices and costs. The range of variation of the payback period parameter b for the industrial sector, goes from 2 to 3.75 (SM, Section 1.2).

The price of consumption-good firms is determined as a mark-up on the unit production cost that changes over time. The initial level of mark-up μ_0 is crucial in governing the dynamics of growth and investment. In the model, mark-up level influences firm profitability and, correspondingly, the propensity of the firm to rely on external source of financing for production and investment. Moreover, higher mark-ups over unit cost of productions also influence income distribution by favouring profits over wages. As aggregate consumption in the model is directly linked to total wages, higher mark-ups can dampen demand for final goods (Dosi et al., 2013). We let the initial mark-up $\overline{\mu}_0$ vary between 0.2 and 0.3 (SM, Section 1.2). As higher (lower) mark-ups determine less (more) dependence of external finance, the well-functioning and the profitability of the banking sector can play a prominent role in fuelling investments and by smoothing the volatility of internal flows of funds. This motivates calibration upon bank margins, i.e. the mark-down for bank deposits μ^{dep} and the mark-down on bank reserves deposited at the Central Bank μ^{res}. We vary μ^{dep} between 0.75 and 1 and μ^{res} between 0.5 and 0.9 (SM, Section 1.4).

Although we keep fixed some parameters, the strategic relevance of the chosen parameters in affecting the properties of the model is proved by the substantial number of experiments that have been conducted for exploring the relationship between economic growth and business cycle fluctuations.

Once the parameters are defined, we draw $m_0 = 200$ CoPs using Quasi Monte Carlo with sampling based on Sobol' sequence (QMCS). Although other sampling methods are possible (e.g., Monte Carlo with pseudo-random numbers and Latin Hypercube Sampling), we decide to exploit QMCS as it gives a better way of arranging points in high-dimensional spaces than standard Monte Carlo methods and standard Latin Hypercube Sampling, having the advantage of a safer rate of convergence (Kucherenko et al., 2015, p. 10). According to Sobol' (1967), QMCS is convenient for many reasons: (i) it allows to reach the best uniformity of distribution as the number of points in the parameter space $N \to \infty$; (ii) it has a good distribution also for small initial sets; (iii) it is very fast in terms of computation time, as its rate of convergence is close to $O\left(N^{-1}\right)$, as opposed to Monte Carlo techniques, where the convergence rate is of $O\left(N^{-1/2}\right)$.[7]

The appropriate number of Monte Carlo runs for each CoP can be determined following the power analysis for ANOVA described in Secchi and Seri (2017) and Seri and Secchi (2017). To proceed with the power analysis, we must consider two features: the number of CoPs and the effect size f. Given $m_0 = 200$, we choose the values of the significance level α (the probability of rejecting the null hypothesis when it is true) and the power of test $1-\beta$ (the probability of rejecting the null when it is false). The value of $1-\beta$ depends on f, which measures the ability to discern between the null and the alternative hypothesis (see, e.g., Cohen, 1988). Generally, the effect size can have different impacts: *small*= 0.1, *medium*= 0.25 and *large*= 0.4. In order to be conservative, and in line with the literature (see Secchi and Seri, 2017), we consider $\alpha = 0.01$, $1 - \beta = 0.95$ and $f = 0.1$. These values lead to an optimal number of Monte Carlo runs $n = 46$ per configuration, for a total of $n \times m_0 = 9200$ runs. However, to reduce as much as possible the effect of the model's stochasticity, we simulate the ABM $n = 200$ times for each CoP. Therefore, the total number of Monte Carlo runs considered in the exercise is $n \times m_0 = 40000$.

We generate $T = 500$ synthetic observations for each Monte Carlo run and we delete the first 105 observations to remove the transients. Therefore, the final sample size is $T = 395$. We then inspect the parameter space to check whether some simulated time series provide unexpected values (e.g., "N/A", "NaN", "-Inf", "Inf", etc.). However, we do not encounter such cases.

The actual data we use for calibration and validation are U.S. data. We rely on two different sources. We draw the macroeconomic variables (i.e., GDP, consumption, investments, unemployment rate, and CPI) from the FRED-QD Database of the Federal Reserve Bank of St. Louis (McCracken and Ng, 2020), and the energy variables (i.e., total energy consumption and total Carbon Dioxide emissions) from the U.S. Energy Information Administration.[8] We take logs of all variables except for the unemployment rate. A description of the empirical dataset is provided in Table 6.2.

All the variables are on quarterly basis and the time series go from 1973:Q1 to 2021:Q1, for a total of $\tau = 193$ observations.[9]

[7] In Section 3 of the Supplementary Material we investigate the space-filling and orthogonality properties of the initial set of CoPs $\mathcal{M}^0$.

[8] Macroeconomic variables are downloaded from: https://research.stlouisfed.org/econ/mccracken/fred-databases/, total energy consumption are downloaded from: https://www.eia.gov/totalenergy/data/monthly/index.php, and total Carbon Dioxide emissions from energy consumption are downloaded from: https://www.eia.gov/environment/.

[9] Energy variables are available either on yearly or monthly basis, therefore we compute the quarterly data summing up the monthly values in each quarter.

Table 6.2
Empirical dataset.

Variable	Unity of measure	Description
GDP	Billions of chained 2012 Dollars	Real Gross Domestic Product
Consumption	Billions of chained 2012 Dollars	Real personal consumption expenditures
Investment	Billions of chained 2012 Dollars	Real Gross Private Domestic Investment
Unemp rate	Percent	Civilian Unemployment Rate
CPI	Index 1982 – 84 = 100	Consumer Price Index for all consumers
Energy	Trillion Btu	Total energy consumption
Emissions	Million Metric Tons of CO2	Total Carbon Dioxide emissions

We start by fitting a reduced-form VAR model on the actual data, selecting the number of lags with the Akaike Information Criteria (AIC). Then, we perform the Ljung–Box test to check whether the VAR residuals are uncorrelated. For all the variables considered in the empirical application, we cannot reject the null hypothesis of uncorrelatedness. In light of this, we fit a VAR(2) model in levels on both the actual data and the simulated time series. Imposing the same number of lags on both VAR models guarantees coherence in the calibration step.

As explained in Section 4.1, the estimation of the matrices $\widehat{\boldsymbol{\Psi}}_0$ and $\widehat{\boldsymbol{\Psi}}_{j,0}\left(\theta_i\right)$ is achieved via *fastICA* from the estimated reduced-form residuals $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{u}}_{jt}\left(\theta_i\right)$. On these residuals, we perform the Jarque–Bera test. The hypothesis of normality is rejected both for actual data and (in the vast majority of cases) for simulated data. Hence, we conclude in favour of non-Gaussianity. We then compute MDI for each Monte Carlo run of each CoP and we take the mean across Monte Carlo runs:

$$\overline{D}_i^{(200)} = \frac{1}{200}\sum_{j=1}^{200}\sqrt{\mathrm{tr}\left[\left(\mathbf{C}_{ji}\widehat{\boldsymbol{\Psi}}_{j,0}^{-1}\left(\theta_i\right)\widehat{\boldsymbol{\Psi}}_0 - \mathbf{I}_K\right)\left(\mathbf{C}_{ji}\widehat{\boldsymbol{\Psi}}_{j,0}^{-1}\left(\theta_i\right)\widehat{\boldsymbol{\Psi}}_0 - \mathbf{I}_K\right)'\right]}, \tag{6.1}$$

for $i = 1, \ldots, 200$.

We then use the MDI as input for the MCS. We select only CoPs that pass the testing procedure, so we discard CoPs with p-value < 0.05. In Table 6.3, we report the order of elimination of the CoPs, the p-values of the test procedure, the MCS p-values and the sample average distances $\overline{D}_i^{(200)}$ for the contemporaneous causal structures. CoPs are ranked according to their p-values; these p-values measure the likelihood of the simulated causal structures with respect to the ones embodied in the real-world data. For readability, we report only the first ten (the last ten) eliminated CoPs. The only configuration selected for the validation procedure is CoP 35. The fact that the MCS procedure selects a single CoP (the "best" model) suggests that, as pointed out by Hansen et al. (2011, p. 454), the actual data used in our application are very informative. This warrants the reliability of our data-driven approach. The values of the parameters associated to CoP 35 are reported in Table 6.4. If compared with the baseline parametrization in Lamperti et al. (2019a), CoP 35 suggests that, in order to better match the contemporaneous causal structure embedded in the model: (i) bank profitability margins should be tighter (μ^{dep}, μ^{res})[10]; (ii) firms' capabilities in innovative activities are stronger because $\zeta_{1,2}$ are higher and the support upon which innovation-driven productivity shocks are drawn ($\underline{x}$ and $\overline{x}$) is larger; (iii) on the other hand, in the energy sector the same support is very close to the baseline specification ($\underline{x}_{\mathrm{en}}, \overline{x}_{\mathrm{en}}$); (iv) finally, the payback parameter b seems to be slightly lower than for the baseline calibration, suggesting that firms are willing to replace their old machineries more often.

In the validation step of our protocol (steps 6 and 7 in Section 2), we measure the goodness-of-match of the shocks-variables structure embodied in CoP 35, with respect to the actual shocks-variables structure. In so doing, we test the significance of the coefficients of the matrix $\widehat{\boldsymbol{\Psi}}_0$, using bootstrap, and the significance of the entries of the matrix $\widehat{\boldsymbol{\Psi}}_{j,0}\left(\theta_{35}\right)$, relying on the distributions of the Monte Carlo runs. Then, we infer ICR for both the simulated and the actual causal structures. The $\mathrm{ICR}_{\mathrm{rw}}$ and $\mathrm{ICR}_{\mathrm{sim}}$ are reported in Table 6.5. To simplify the notation, we write $\varepsilon_{k_s,t} = \varepsilon_{k_s}$, with $k_s = 1, \ldots, 7$.

The analysis of the ICRs provides the following outcomes: (i) as regards the $\mathrm{ICR}_{\mathrm{rw}}$, ε_1 has a significant (contemporaneous) impact on investment, ε_2 has an impact on GDP, consumption, investment, unemployment rate, demand of energy and emissions, ε_3 on GDP, investment, unemployment rate, demand of energy and emissions, ε_4 and ε_5 impact only on investment, and ε_7 has an impact on the demand of energy and emissions; (ii) as regards the $\mathrm{ICR}_{\mathrm{sim}}$, ε_1 influences investment, ε_2 has an impact on investment, ε_3 impacts consumption, investment and unemployment rate, ε_6 hits investment and ε_7 has an impact on investment and demand of energy emissions; (iii) the validation measure $\mathrm{VM} = 1 - {}^{14}/_{49} = 0.714$, therefore, the simulated model is able to recover the 71.4% of the real-world shocks-variables structures.[11]

It is worth noting that CoP 35 is the "best" configuration across different SVAR specifications, confirming the robustness of the calibration exercise (see Appendix D, Table D.3). When focusing on the identified and statistically significant causal relationships, the model delivers a lower performance in matching the emerging shocks-variable structure regarding GDP, CPI and Emissions at the macro-level. However, a set of shocks hitting energy and investment is consistently identified in both the simulated model and

[10] Lamperti et al. (2019a) do not report a baseline value for the parameters affecting bank margins (see Supplementary Information, Section D, Table 2 therein). We retrieved the parameters from the original code. These two parameters play an important role on the banks profitability margins and thus on the capacity of the banking sector to steadily fuel investments. Accordingly, we decided a range for μ^{res} and μ^{dep} that we think is more representative of the interest rate structure.

[11] Inference conducted on $\mathrm{ICR}_{\mathrm{rw}}$ and $\mathrm{ICR}_{\mathrm{sim}}$ reveals that many coefficients are not statistically different from zero, leading to a contemporaneous impact matrix that is sparse. Despite the relatively high score of the validation measure is driven by the "common absence" of statistically relevant causal ties, the validation measure VM provides, in a theory-agnostic manner, a relevant informational content.

Table 6.3
Order of elimination of the different CoPs with p-values and sample average distances.

k	$e_{\mathcal{M}_k}$	p-value of $\delta_{\mathcal{M}_k}$ ($p_{\mathrm{H}_{0,\mathcal{M}_k}}$)	MCS p-value ($\hat{p}_{e_{\mathcal{M}_k}}$)	$\overline{D}_i^{(200)}$
1	108	0.0000	0.00000	0.5037
2	44	0.0000	0.00000	0.4993
3	112	0.0000	0.00000	0.4960
4	166	0.0000	0.00000	0.4945
5	188	0.0000	0.00000	0.4922
6	48	0.0000	0.00000	0.4857
7	199	0.0000	0.00000	0.4851
8	129	0.0000	0.00000	0.4816
9	13	0.0000	0.00000	0.4732
10	20	0.0000	0.00000	0.4652
⋮	⋮	⋮	⋮	⋮
191	159	0.00000	0.00000	0.0313
192	63	0.00000	0.00000	0.0312
193	65	0.00000	0.00000	0.0311
194	195	0.00000	0.00000	0.0310
195	133	0.00000	0.00000	0.0302
196	27	0.00011	4.69e-05	0.0301
197	179	0.00012	0.00011	0.0286
198	99	4.69e-05	0.00012	0.0284
199	139	0.00015	0.00015	0.0275
200	35	1.00000	1.00000	0.0268

The shaded area identifies the selected CoP.

Table 6.4
CoP 35 values compared to baseline in Lamperti et al. (2019a).

Empirical calibration		
Parameter	CoP 35	Baseline
$\overline{\mu}_0$	0.2875	0.28
μ^{dep}	0.8669	1
μ^{res}	0.5525	0.33
$\zeta_{1,2}$	0.7597	0.3
$\underline{x}$	−0.1417	−0.08
$\overline{x}$	0.1417	0.08
$\underline{x}_{\mathrm{en}}$	−0.0566	−0.058
$\overline{x}_{\mathrm{en}}$	0.0566	0.058
b	2.59	3

Table 6.5
Independent Component Representation from actual data (left) and simulated data associated to CoP 35 (right).

	$\mathrm{ICR}_{\mathrm{rw}}$								$\mathrm{ICR}_{\mathrm{sim}}$						
	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7		ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7
GDP	0	1	1	0	0	0	0	GDP	0	0	0	0	0	0	0
$Cons$	0	1	0	0	0	0	0	$Cons$	0	0	1	0	0	0	0
Inv	1	1	1	1	1	0	0	Inv	1	1	1	0	0	1	1
UR	0	1	1	0	0	0	0	UR	0	0	1	0	0	0	0
CPI	0	0	0	0	0	0	0	CPI	0	0	0	0	0	0	0
$Ener$	0	1	1	0	0	0	1	$Ener$	0	0	0	0	0	0	1
$Emiss$	0	1	1	0	0	0	1	$Emiss$	0	0	0	0	0	0	0

its empirical counterpart. As summarized in Appendix D, Table D.3, the same common shocks-variable structure is also identified in lower-dimensional SVARs. This finding partially aligns with the modelling purpose of Lamperti et al. (2019a), more precisely in replicating stylized facts on energy demand and CO2 emissions patterns that are consistent with the Shared Socioeconomic Pathway 5 (SSP5, Riahi et al., 2017).

7. Conclusions

In this paper, we propose a new general protocol for calibration and validation of complex simulation models by searching causal structures both from synthetic and actual data. The emphasis on causal search is linked to the importance that policy analysis, specifically, the prediction of the effects of policy interventions, plays in macroeconomic simulation models. Our procedure combines MCS and causal inference: first, we estimate reduced-form VAR models from both the data generated by a macroeconomic simulation model and a set of observed data, and we identify, through ICA, a vector of structural shocks and a mixing matrix; then, we compute

the MDI between, on the one hand, the mixing matrix associated to each CoP and Monte Carlo run and, on the other hand, the mixing matrix estimated from real data, and we apply the MCS to the distribution of the MDIs to select the set of CoPs that best approximate actual data; finally, for the selected CoP(s), we infer an ICR describing which shocks have a significant impact on the variables, and we compare such ICR with the analogous ICR derived from the actual data.

We apply our method to the DSK model of Lamperti et al. (2019a). The results show that the MCS procedure based on the MDI discriminates well among different CoPs (only CoP 35 passes the test). According to our validation measure, the best CoP turns out to mimic the 71.4% of the shocks-variables structure underlying the actual data.

Our protocol can be seen as a complement and a generalization of other existing calibration and validation methods, for at least three reasons: (i) it allows the researcher to rank causal structures associated to different CoPs of a simulation model from the most to the least plausible according to a statistical measure; (ii) it is faster than other procedures based on the optimization of an objective function or the exploration of the parameter space; (iii) it applies both to calibration and validation.

Further developments are possible. First, one can replace the minimum-distance index with another metric which accounts for the long-run dynamics of the macroeconomic variables. Second, SVARs can be estimated through Vector Error Correction Models to account for potential cointegration among the variables. More in general, causal structures between shocks and variables can be estimated by econometric time series models that relax many features of the standard linear VAR model.

Declaration of competing interest

The authors declare that we have no conflicts of interest to disclose.

Appendix A. The ICA model and fastICA

In this section, we detail the assumptions and theoretical background underlying the ICA model and the *fastICA* estimator.

Given that $\mathbf{u}_t = \boldsymbol{\Psi}_0 \boldsymbol{\varepsilon}_t \iff \boldsymbol{\varepsilon}_t = \boldsymbol{\Gamma}_0 \mathbf{u}_t$ (with $\boldsymbol{\Gamma}_0^{-1} = \boldsymbol{\Psi}_0$), the ICA model recovers, up to scale and order indeterminacy, $\boldsymbol{\Gamma}_0$ ($\boldsymbol{\Psi}_0$) and $\boldsymbol{\varepsilon}_t$ from realizations of $\mathbf{u}_t$ under the following assumptions (see, e.g., Hyvärinen et al., 2001 and Hyvärinen, 2013):

Assumption 1.

(i) The components $\boldsymbol{\varepsilon}_t$ are statistically independent;

(ii) The components $\boldsymbol{\varepsilon}_t$ are non-Gaussian with at most one exception;

(iii) The matrix $\boldsymbol{\Psi}_0$ is invertible.

The *fastICA* estimator is a fixed-point algorithm that relies on the maximization of the non-Gaussianity of $\boldsymbol{\gamma}_k' \mathbf{u}_t$, for each $k = 1, \dots, K$, where $\boldsymbol{\gamma}_k'$ is the kth row of the matrix $\boldsymbol{\Gamma}_0$ (see Hyvärinen, 1999). Without loss of generality, in the following we will write $\boldsymbol{\gamma}'$ instead of $\boldsymbol{\gamma}_k'$.

The non-Gaussianity is measured using negentropy, which relies on the notion of entropy. In statistics and social sciences, entropy represents the amount of uncertainty associated with a probability distribution or, simpler, the loss of information deriving by using a model to approximate reality. More in general, entropy is a measure of disorder or uncertainty (see, e.g., Seri and Martinoli, 2021). Let $\mathbf{y}$ be a continuous random vector, the differential entropy (Shannon, 1948) is defined as follows:

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) \, \mathrm{d}\mathbf{y}, \tag{A.1}$$

where $f(\cdot)$ is the probability density function. The differential entropy can be used to quantify the degree of non-Gaussianity. This comes from the fact that a Gaussian variable has the largest entropy among all random variables of equal variance (see, e.g., Cover and Thomas, 1991).

Given a Gaussian vector $\mathbf{x}$ with the same covariance matrix of $\mathbf{y}$, we can define negentropy as follows:

$$J(\mathbf{y}) = H(\mathbf{x}) - H(\mathbf{y}). \tag{A.2}$$

If both $H(\mathbf{x})$ and $H(\mathbf{y})$ are Gaussian, it follows that $J(\mathbf{y}) = 0$. Moreover, according to the fact that $H(\mathbf{x})$ has the maximum entropy, $J(\mathbf{y})$ is always positive. For these reasons, negentropy is one of the most natural index to measure non-Gaussianity.

Since estimation of (neg)entropy from the data can be difficult, as it involves the nonparametric estimation of $f(\cdot)$, the *fastICA* algorithm exploits the following approximation:

$$J(\mathbf{y}) \approx \left[\mathbb{E}(G(\mathbf{y})) - \mathbb{E}(G(\mathbf{z}))\right]^2, \tag{A.3}$$

where $G(\cdot)$ is a specific nonquadratic function of a random variable, i.e. $G(z) = -\exp\left(z^2/2\right)$ with $z \sim \mathcal{N}(0,1)$. The approximation devised in Eq. (A.3) drastically reduces the computational time to find ICA projections (see Issoglio et al., 2021, for a discussion).

Finally, we have the following estimator:

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \mathbb{E}\left[J\left(\boldsymbol{\gamma}' \mathbf{u}_t\right)\right]. \tag{A.4}$$

According to Hyvärinen et al. (2001), Eq. (A.4) is equivalent to the following maximization problem:

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \mathbb{E}\left[G\left(\boldsymbol{\gamma}' \mathbf{u}_t\right)\right], \tag{A.5}$$

as both Eq. (A.4) and Eq. (A.5) have the same solution $\boldsymbol{\gamma}$. Let $g(\cdot)$ be the first derivative of $G(\cdot)$, the statistical properties of the *fastICA* estimator hold under the following assumptions:

Assumption 2.

(i) $\mathbb{E}\left[\mathbf{u}_t\right] = 0$ and $\mathbf{u}_t$ has all moments up to the fourth;

(ii) $\dot{g}(\cdot)$ and $\ddot{g}(\cdot)$, i.e. the first and second derivatives of $g(\cdot)$, satisfy Lipschitz continuity, which means that $\exists \delta_1, \delta_2 < \infty$ such that $\|\dot{g}(\mathbf{y}_1) - \dot{g}(\mathbf{y}_2)\| \leq \delta_1 \|\mathbf{y}_1 - \mathbf{y}_2\|$ and $\|\ddot{g}(\mathbf{y}_1) - \ddot{g}(\mathbf{y}_2)\| \leq \delta_2 \|\mathbf{y}_1 - \mathbf{y}_2\|$;

(iii) $\ddot{g}(\cdot)$ is bounded.

Assumption 2(i) is related to the fact that negentropy can be approximated by using higher-order moments, i.e. kurtosis (see Hyvärinen, 1999). Assumption 2(ii) and (iii) are regularity conditions on the function $g(\cdot)$ and its derivatives to perform the maximization problem.

Therefore, it can be shown that, under Assumptions 1–2 and given the first-order conditions $\mathbb{E}\left(\mathbf{u}_t g\left(\boldsymbol{\gamma}'\mathbf{u}_t\right)\right) = \mathbf{0}$ of the maximization problem (A.5), the estimator $\hat{\boldsymbol{\gamma}} = \{\boldsymbol{\gamma} : \mathbb{E}\left(\mathbf{u}_t g\left(\boldsymbol{\gamma}'\mathbf{u}_t\right) = \mathbf{0}\right)\}$ is consistent and asymptotically normal (see Reyhani et al., 2012), that is:

$$\hat{\boldsymbol{\gamma}} \rightarrow_{\Pr} \boldsymbol{\gamma}, \tag{A.6}$$

$$\sqrt{T}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \rightarrow_{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{A.7}$$

where $\boldsymbol{\Omega}$ is the positive-definite covariance matrix.

Notice that maximizing non-Gaussianity is strictly related to minimizing mutual statistical independence. This connection has been shown by Hyvärinen and Oja (2000) who demonstrate that the most non-Gaussian directions $\boldsymbol{\gamma}'\mathbf{u}_t$ can be found by minimizing the Kullback–Leibler divergence between the joint density $f\left(\boldsymbol{\gamma}_1'\mathbf{u}_t, \ldots, \boldsymbol{\gamma}_K'\mathbf{u}_t\right)$ and the product of the marginals $f\left(\boldsymbol{\gamma}_1'\mathbf{u}_t\right) \cdot \ldots \cdot f\left(\boldsymbol{\gamma}_K'\mathbf{u}_t\right)$.

Appendix B. MCS algorithm

The implementation of the MCS procedure follows the algorithm below:

1. start from the set $\mathcal{M}^0 := \{1, \ldots, m_0\}$ and test, with level $1 - \alpha$, that all the average distances are equal: if $\hat{p}_{e_{\mathcal{M}_1}} > \alpha$, do not reject $\mathsf{H}_{0,\mathcal{M}}$ and the procedure is over; if $\hat{p}_{e_{\mathcal{M}_1}} \leq \alpha$, reject $\mathsf{H}_{0,\mathcal{M}}$ and go to step 2;
2. use the elimination rule $e_{\mathcal{M}} = e_{\mathcal{M}^0}$ to remove one CoP from $\mathcal{M}^0$, getting $\mathcal{M}^1 := \{1, \ldots, m_0 - 1\}$;
3. test, with level $1 - \alpha$, that all the average distances associated with $i \in \mathcal{M}^1$ are equal; again, if the $\hat{p}_{e_{\mathcal{M}_2}} > \alpha$, do not reject the null hypothesis and the procedure is over; if the $\hat{p}_{e_{\mathcal{M}_2}} \leq \alpha$ reject the null hypothesis, use again the elimination rule, and perform the test with $i \in \{1, \ldots, m_0 - 2\}$;
4. the procedure continues until the null hypothesis is not rejected. The final set of CoPs is defined as $\widehat{\mathcal{M}}^*$.

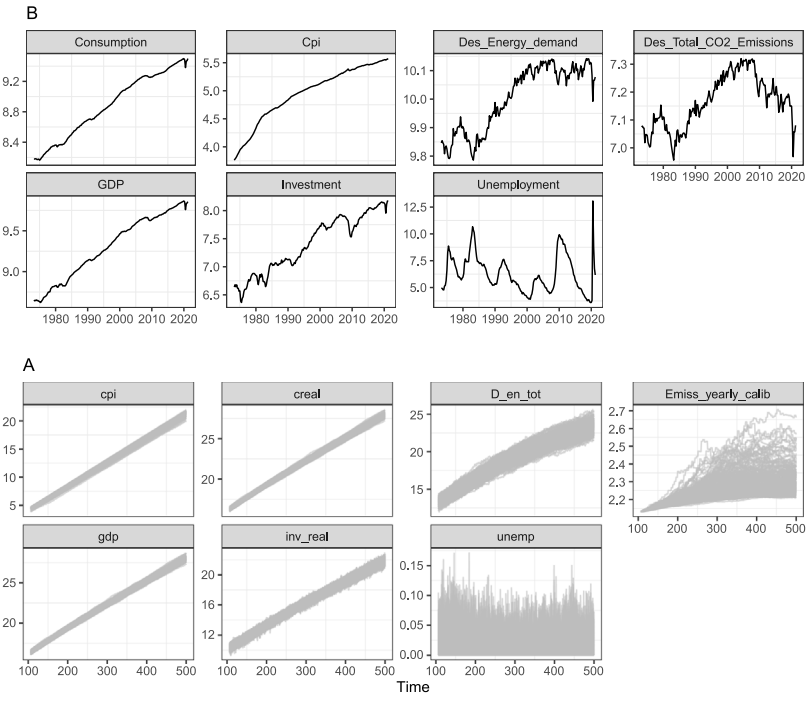

Fig. C.1. Plots of the real-world vs. simulated variables obtained with CoP 35.

Appendix C. Behaviour of the validated simulated variables

Fig. C.1 represents the behaviour of the real-world (upper plots) and the simulated variables (bottom plots) considered in the analysis, obtained with the validated configuration.

Appendix D. Robustness checks

We compute the MCS at 95% for different SVAR models, to investigate whether the calibration procedure is robust across different specifications. To do that, we estimate four different SVAR models, for both simulated and actual data, which we call $SVAR_1$, $SVAR_2$, $SVAR_3$ and $SVAR_4$. The composition of these SVAR models is reported in Table D.1.

Table D.1
SVARs specifications.

Model	Variables
$SVAR_1$	GDP, Inv, $Ener$
$SVAR_2$	GDP, Inv, $Ener$, $Emiss$
$SVAR_3$	GDP, Inv, CPI, $Ener$
$SVAR_4$	GDP, $Cons$, Inv, CPI, $Ener$

The outcomes of the MCS at 95% (the order of elimination of the CoPs, the MCS p-values and the sample average distances) for $SVAR_1$, $SVAR_2$, $SVAR_3$ and $SVAR_4$ are reported in Table D.2 (for readability, we show only the first ten and the last ten eliminated CoPs).

Table D.2
Order of elimination of the CoPs of the different SVAR models with MCS p-values and sample average distances.

	$SVAR_1$			$SVAR_2$			$SVAR_3$			$SVAR_4$		
k	$e_{\mathcal{M}_k}$	$\hat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\hat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\hat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\hat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$
1	108	0.0000	1.2174	108	0.0000	0.8969	108	0.0000	0.8922	108	0.0000	0.7143
2	44	0.0000	1.2119	44	0.0000	0.8922	44	0.0000	0.8874	44	0.0000	0.7095
3	112	0.0000	1.2030	112	0.0000	0.8867	112	0.0000	0.8823	112	0.0000	0.7053
4	166	0.0000	1.2019	188	0.0000	0.8858	166	0.0000	0.8806	166	0.0000	0.7030
5	188	0.0000	1.2014	166	0.0000	0.8843	188	0.0000	0.8787	188	0.0000	0.6992
6	199	0.0000	1.1821	199	0.0000	0.8714	199	0.0000	0.8655	48	0.0000	0.6904
7	48	0.0000	1.1754	48	0.0000	0.8671	48	0.0000	0.8633	199	0.0000	0.6891
8	129	0.0000	1.1659	129	0.0000	0.8597	129	0.0000	0.8549	129	0.0000	0.6841
9	13	0.0000	1.1579	13	0.0000	0.8529	13	0.0000	0.8474	13	0.0000	0.6724
10	20	0.0000	1.1328	20	0.0000	0.8338	20	0.0000	0.8288	20	0.0000	0.6611
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
191	159	0.0000	0.0809	159	0.0000	0.0554	159	0.0000	0.0541	65	0.0000	0.0458
192	63	0.0000	0.0806	65	0.0000	0.0548	65	0.0000	0.0536	63	0.0000	0.0457
193	65	0.0000	0.0796	195	0.0000	0.0546	195	0.0000	0.0533	159	0.0000	0.0453
194	195	0.0000	0.0779	133	0.0000	0.0543	133	0.0000	0.0530	195	0.0000	0.0444
195	133	0.0000	0.0776	63	0.0000	0.0539	63	0.0000	0.0529	133	0.0000	0.0441
196	27	0.0004	0.0756	27	1.6e-08	0.0533	27	2.5e-06	0.0521	27	5.9e-06	0.0432
197	99	0.0005	0.0736	99	5.7e-08	0.0499	99	2.8e-06	0.0489	99	6.8e-06	0.0410
198	179	0.0027	0.0716	179	5.9e-08	0.0498	179	3.1e-06	0.0486	179	6.8e-06	0.0408
199	139	0.0098	0.0708	139	1.6e-07	0.0488	139	1.5e-05	0.0475	139	3.1e-05	0.0399
200	35	1.0000	0.0690	35	1.0000	0.0470	35	1.0000	0.0459	35	1.0000	0.0386

The shaded area highlights the selected configurations, for each SVAR model, with their corresponding MCS p-values and sample average distances.

All these SVAR specifications highlight a common finding; the theoretical model identifies the same set of common shocks hitting the investment and energy variable. To this extent, the model seems to match quite accurately the empirical structure found in U.S data that link energy demand to investment dynamics. In Table D.3, we report the validation measures and the shocks-variables structures common to both simulated and actual data, for all the SVAR specifications.

Table D.3
Validation measures and common shocks-variables structures for different SVAR specifications.

Model	VM	Shocks-variables structures
$SVAR_1$	0.56	$\varepsilon_1 \rightarrow Inv$, $\varepsilon_2 \rightarrow Inv$, $\varepsilon_3 \rightarrow Ener$
$SVAR_2$	0.63	$\varepsilon_1 \rightarrow Inv$, $\varepsilon_2 \rightarrow Inv$
$SVAR_3$	0.75	$\varepsilon_1 \rightarrow Inv$, $\varepsilon_2 \rightarrow Inv$
$SVAR_4$	0.76	$\varepsilon_1 \rightarrow Inv$, $\varepsilon_2 \rightarrow Inv$, $\varepsilon_3 \rightarrow Inv$

Appendix E. Moving-average representation and impulse response functions

Suppose to compare the structural impulse response matrices at different time horizons $\ell = 0, \ldots, H$, then $\mathbf{y}_t$ and $\mathbf{z}_{jt}\left(\theta_i\right)$ must be represented as a moving-average process.

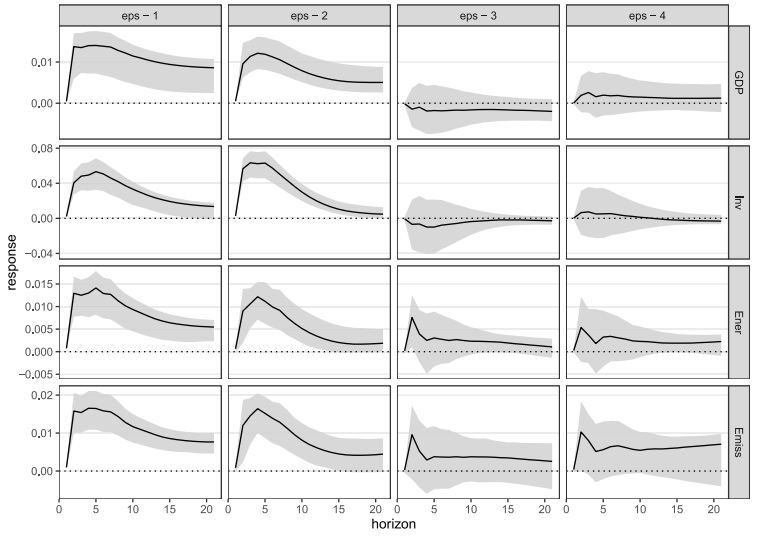

Fig. E.1. Plots of the real-world IRFs of SVAR_2.

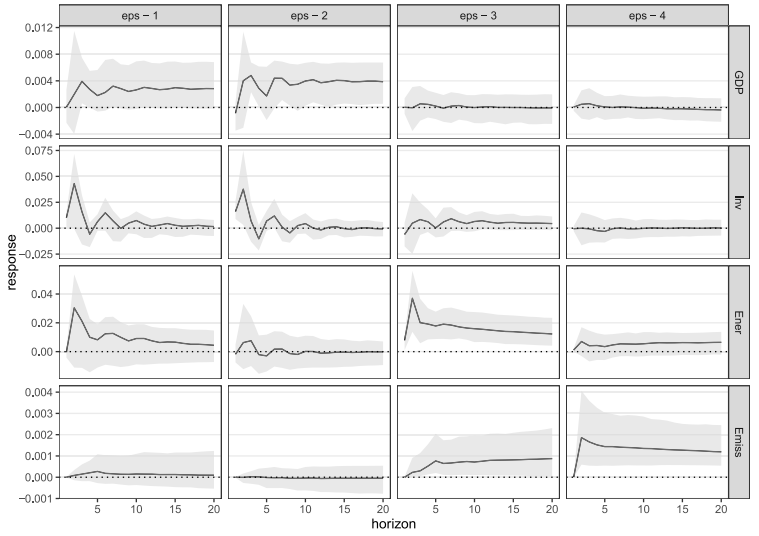

Fig. E.2. Plots of the simulated IRFs of SVAR_2 obtained with CoP 35.

If the process $\mathbf{y}_t$ is stable (i.e., $\det(\mathbf{I}_K - \mathbf{A}_1 z - \cdots - \mathbf{A}_P z^P) \neq 0 \quad \forall z \in \mathbb{C}, \quad |z| \leq 1$), then $\mathbf{y}_t$ admits a Wold moving-average (MA) representation:

$$\mathbf{y}_t = \sum_{\ell=0}^{\infty} \boldsymbol{\Phi}_\ell \mathbf{u}_{t-\ell}, \tag{E.1}$$

where $\boldsymbol{\Phi}_0 = \mathbf{I}_K$ and $\boldsymbol{\Phi}_\ell = \sum_{d=1}^{\ell} \boldsymbol{\Phi}_{\ell-d} \mathbf{A}_d$. We can also write:

$$\mathbf{y}_t = \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_\ell \boldsymbol{\varepsilon}_{t-\ell}, \tag{E.2}$$

where $\boldsymbol{\Psi}_\ell = \boldsymbol{\Phi}_\ell \boldsymbol{\Gamma}_0^{-1}$ and, in particular, $\boldsymbol{\Psi}_0 = \boldsymbol{\Gamma}_0^{-1}$. The entries of the matrices $\boldsymbol{\Psi}_\ell$, for $\ell = 0, \ldots, H$, are referred in the literature as the impulse response functions since $\psi_{dk,\ell} = \frac{\partial y_{d,t+\ell}}{\partial \varepsilon_{kt}}$, where $\psi_{dk,\ell}$ is the (d,k) entry of $\boldsymbol{\Psi}_\ell$.

If $\mathbf{y}_t$ contains processes with unit roots, although the VAR model does not admit a Wold representation, the matrices $\boldsymbol{\Phi}_\ell \boldsymbol{\Gamma}_0^{-1} = (\sum_{d=1}^{\ell} \boldsymbol{\Phi}_{\ell-d} \mathbf{A}_d) \boldsymbol{\Gamma}_0^{-1}$ still represent impulse response functions, which, however, may not approach zero for $\ell \to \infty$ (see Kilian and Lütkepohl, 2017).

Analogous representations hold for data generated by the simulation model, therefore we can write:

$$\mathbf{z}_{jt}\left(\theta_i\right) = \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{j,\ell}\left(\theta_i\right) \boldsymbol{\varepsilon}_{jt-\ell}\left(\theta_i\right). \tag{E.3}$$

The impulse response functions for $SVAR_2$, obtained using real-world data, are displayed in Fig. E.1, while the impulse response functions for $SVAR_2$, estimated using the data simulated from the model, are shown in Fig. E.2.

Appendix F. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2024.106786.

Data availability

We have shared the link to the data/code in the Supplementary Material.

References

Acid, S., de Campos, L.M., 2003. Searching for Bayesian network structures in the space of restricted acyclic partially directed graphs. J. Artificial Intelligence Res. 18, 445–490.

Altissimo, F., Mele, A., 2009. Simulated non-parametric estimation of dynamic models. Rev. Econ. Stud. 76 (2), 413–450.

An, S., Schorfheide, F., 2007. Bayesian analysis of DSGE models. Econometric Rev. 26 (2–4), 113–172.

Barde, S., 2020. Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion. J. Econom. Dynam. Control 111, 103795.

Bruns, S.B., Moneta, A., Stern, D.I., 2021. Estimating the economy-wide rebound effect using empirically identified structural vector autoregressions. Energy Econ. 97, 105158.

Cantore, C., Gabriel, V.J., Levine, P., Pearlman, J., Yang, B., 2013. Handbook of Research Methods and Applications in Empirical Macroeconomics. Edward Elgar Publishing, Cheltenham.

Casini, L., Moneta, A., Capasso, M., 2021. Variable definition and independent components. Philos. Sci. 88 (5), 784–795.

Castro, J., Drews, S., Exadaktylos, F., Foramitti, J., Klein, F., Konc, T., Savin, I., van den Bergh, J., 2020. A review of agent-based modeling of climate-energy policy. WIREs Clim. Change 11 (4), e647.

Choirat, C., Seri, R., 2012. Estimation in discrete parameter models. Statist. Sci. 27 (2), 278–293.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. J. Polit. Econ. 113 (1), 1–45.

Christiano, L.J., Eichenbaum, M.S., Trabandt, M., 2018. On DSGE models. J. Econ. Perspect. 32 (3), 113–140.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Lawrence Erlbaum Associates, Hillsdale, N.J..

Comon, P., 1994. Independent component analysis, A new concept? Signal Process. 36 (3), 287–314.

Cooley, T.F., 1997. Calibrated models. Oxf. Rev. Econ. Policy 13 (3), 55–69.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. John Wiley & Sons, New York, N.Y.

Cúrdia, V., Del Negro, M., Greenwald, D.L., 2014. Rare shocks, great recessions. J. Appl. Econom. 29 (7), 1031–1052.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007. On the fit of new Keynesian models. J. Bus. Econom. Statist. 25 (2), 123–143.

Del Negro, M., Schorfheide, F., et al., 2006. How good is what you've got? DGSE-VAR as a toolkit for evaluating dsge models. Econ. Rev.-Fed. Reserve Bank Atlanta 91 (2), 21.

Delli Gatti, D., Grazzini, J., 2020. Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic agent based models. J. Econ. Behav. Organ. 178, 875–902.

Dosi, G., Fagiolo, G., Napoletano, M., Roventini, A., 2013. Income distribution, credit and fiscal policies in an agent-based Keynesian model. J. Econ. Dyn. Control 37 (8), 1598–1625.

Dosi, G., Fagiolo, G., Napoletano, M., Roventini, A., Treibich, T., 2015. Fiscal and monetary policies in complex evolving economies. J. Econom. Dynam. Control 52, 166–189.

Dosi, G., Fagiolo, G., Roventini, A., 2010. Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. J. Econom. Dynam. Control 34 (9), 1748–1767.

Dridi, R., Guay, A., Renault, E., 2007. Indirect inference and calibration of dynamic stochastic general equilibrium models. J. Econometrics 136 (2), 397–430.

Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., Roventini, A., 2019. Validation of agent-based models in economics and finance. In: Beisbart, C., Saam, N.J. (Eds.), Computer Simulation Validation. Springer, Cham, pp. 763–787.

Favero, C.A., 2001. Applied Macroeconometrics. Oxford University Press, Oxford.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., Watson, M.W., 2007. ABCs (and Ds) of understanding VARs. Amer. Econ. Rev. 97 (3), 1021–1026.

Frazier, D.T., Martin, G.M., Robert, C.P., Rousseau, J., 2018. Asymptotic properties of approximate Bayesian computation. Biometrika 105 (3), 593–607.

Giacomini, R., 2013. The relationship between DSGE and VAR models. In: VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher a. Sims. In: Advances in Econometrics, vol. 32, Emerald Group Publishing Limited, Leeds, pp. 1–25.

Gomme, P., Rupert, P., 2007. Theory, measurement and calibration of macroeconomic models. J. Monet. Econ. 54 (2), 460–497.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. J. Appl. Econometrics 8 (S1), S85–S118.

Gouriéroux, C., Monfort, A., Renne, J.-P., 2017. Statistical inference for independent component analysis: Application to structural VAR models. J. Econometrics 196 (1), 111–126.

Grazzini, J., Richiardi, M.G., Tsionas, M., 2017. Bayesian estimation of agent-based models. J. Econom. Dynam. Control 77, 26–47.

Guerini, M., Moneta, A., 2017. A method for agent-based models validation. J. Econom. Dynam. Control 82, 125–141.

Guerron-Quintana, P., Inoue, A., Kilian, L., 2017. Impulse response matching estimators for DSGE models. J. Econometrics 196 (1), 144–155.

Gusella, F., Ricchiuti, G., 2024. Endogenous cycles in heterogeneous agent models: a state-space approach. J. Evol. Econ. 1–44.

Hall, A.R., Inoue, A., Nason, J.M., Rossi, B., 2012. Information criteria for impulse response function matching estimation of DSGE models. J. Econometrics 170 (2), 499–518.

Hansen, L.P., Heckman, J.J., 1996. The empirical foundations of calibration. J. Econ. Perspect. 10 (1), 87–104.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Herwartz, H., 2018. Hodges–Lehmann detection of structural shocks–an analysis of macroeconomic dynamics in the Euro area. Oxf. Bull. Econ. Stat. 80 (4), 736–754.
Herwartz, H., Plödt, M., 2016. The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. J. Int. Money Finance 61, 30–44.
Hinkelmann, F., Murrugarra, D., Jarrah, A.S., Laubenbacher, R., 2011. A mathematical framework for agent based models of complex biological networks. Bull. Math. Biol. 73 (7), 1583–1602.
Hoover, K.D., 2012. Economic theory and causal inference. In: Mäki, U. (Ed.), Handbook of the Philosophy of Economics. Elsevier/North-Holland, Amsterdam, pp. 89–113.
Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10 (3), 626–634.
Hyvärinen, A., 2013. Independent component analysis: recent advances. Phil. Trans. R. Soc. A 371 (1984), 20110534.
Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. In: Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control, J. Wiley, New York, N.Y..
Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13 (4–5), 411–430.
Ilmonen, P., Nordhausen, K., Oja, H., Ollila, E., 2010. A new performance index for ICA: Properties, computation and asymptotic analysis. In: Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R., Vincent, E. (Eds.), Latent Variable Analysis and Signal Separation. In: Lecture Notes in Computer Science, vol. 6365, Springer, Berlin, pp. 229–236.
Ireland, P.N., 2004. A method for taking models to the data. J. Econom. Dynam. Control 28 (6), 1205–1226.
Issoglio, E., Smith, P., Voss, J., 2021. On the estimation of entropy in the FastICA algorithm. J. Multivariate Anal. 181, 104689.
Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, Cambridge.
Kristensen, D., Shin, Y., 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood. J. Econometrics 167 (1), 76–94.
Kucherenko, S., Albrecht, D., Saltelli, A., 2015. Exploring multi-dimensional spaces: A comparison of latin hypercube and quasi monte carlo sampling techniques. arXiv preprint arXiv:1505.02350.
Kukacka, J., Sacht, S., 2023. Estimation of heuristic switching in behavioral macroeconomic models. J. Econom. Dynam. Control 146, 104585.
Kydland, F.E., Prescott, E.C., 1996. The computational experiment: An econometric tool. J. Econ. Perspect. 10 (1), 69–85.
Lamperti, F., Bosetti, V., Roventini, A., Tavoni, M., 2019a. The public costs of climate-induced financial instability. Nature Clim. Change 9 (11), 829–833.
Lamperti, F., Dosi, G., Napoletano, M., Roventini, A., Sapio, A., 2018a. Faraway, so close: Coupled climate and economic dynamics in an agent-based integrated assessment model. Ecol. Econ. 150, 315–339.
Lamperti, F., Dosi, G., Napoletano, M., Roventini, A., Sapio, A., 2020. Climate change and green transitions in an agent-based integrated assessment model. Technol. Forecast. Soc. Change 153, 119806.
Lamperti, F., Mandel, A., Napoletano, M., Sapio, A., Roventini, A., Balint, T., Khorenzhenko, I., 2019b. Towards agent-based integrated assessment models: examples, challenges, and future developments. Reg. Environ. Change 19 (3), 747–762.
Lamperti, F., Roventini, A., 2022. Beyond climate economics orthodoxy: impacts and policies in the agent-based integrated-assessment DSK model. Eur. J. Econ. Econ. Policies 19 (3), 357–380.
Lamperti, F., Roventini, A., Sani, A., 2018b. Agent-based model calibration using machine learning surrogates. J. Econom. Dynam. Control 90, 366–389.
Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196 (2), 288–304.
Lee, L.-F., 1992. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econometric Theory 8 (4), 518–552.
Lucas, R.E.J., 1976. Econometric Policy Evaluation: A Critique. In: Carnegie-Rochester Conference Series on Public Policy, vol. 1, pp. 19–46.
Matteson, D.S., Tsay, R.S., 2017. Independent component analysis via distance covariance. J. Amer. Statist. Assoc. 112 (518), 623–637.
McCracken, M., Ng, S., 2020. FRED-QD: A Quarterly Database for Macroeconomic Research. Working Paper 26872, National Bureau of Economic Research.
McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57 (5), 995–1026.
Moneta, A., Entner, D., Hoyer, P.O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications. Oxf. Bull. Econ. Stat. 75 (5), 705–730.
Moneta, A., Pallante, G., 2022. Identification of structural var models via independent component analysis: A performance evaluation study. J. Econom. Dynam. Control 144, 104530.
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.
Parker, W.S., 2020. Model evaluation: An adequacy-for-purpose view. Philos. Sci. 87 (3), 457–477.
Ravenna, F., 2007. Vector autoregressions and reduced form representations of DSGE models. J. Monetary Econ. 54 (7), 2048–2064.
Reyhani, N., Ylipaavalniemi, J., Vigário, R., Oja, E., 2012. Consistency and asymptotic normality of FastICA and bootstrap FastICA. Signal Process. 92 (8), 1767–1778.
Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Silva, L.A. Da, Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Glob. Environ. Change 42, 153–168.
Secchi, D., Seri, R., 2017. Controlling for false negatives in agent-based models: a review of power analysis in organizational research. Comput. Math. Organ. Theory 23 (1), 94–121.
Seri, R., Martinoli, M., 2021. Asymptotic properties of the plug-in estimator of the discrete entropy under dependence. IEEE Trans. Inform. Theory 67 (12), 7659–7683.
Seri, R., Martinoli, M., Secchi, D., Centorrino, S., 2021. Model calibration and validation via confidence sets. Econom. Stat. 20, 62–86.
Seri, R., Secchi, D., 2017. How many times should one run a computational simulation? In: Edmonds, B., Meyer, R. (Eds.), Simulating Social Complexity: A Handbook, Understanding Complex Systems. Springer International Publishing, Cham, pp. 229–251.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Smith, A.A.J., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. J. Appl. Econometrics 8 (S1), S63–S84.
Sobol', I., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. 7 (4), 86–112.
Steele, K., Werndl, C., 2013. Climate models, calibration, and confirmation. Br. J. Phil. Sci. 64 (3), 609–635.
Stern, N., Stiglitz, J.E., 2021. The Social Cost of Carbon, Risk, Distribution, Market Failures: An Alternative Approach. Working Paper 28472, National Bureau of Economic Research, Cambridge, MA.
Tsamardinos, I., Brown, L.E., Aliferis, C.F., 2006. The max–min hill-climbing Bayesian network structure learning algorithm. Mach. Learn. 65 (1), 31–78.
Ward, J.A., Evans, A.J., Malleson, N.S., 2016. Dynamic calibration of agent-based models using data assimilation. R. Soc. Open Sci. 3 (4), 150703.
Weyant, J., 2017. Some contributions of integrated assessment models of global climate change. Rev. Environ. Econ. Policy 11 (1), 115–137.
Windrum, P., Fagiolo, G., Moneta, A., 2007. Empirical validation of agent-based models: Alternatives and prospects. J. Artif. Soc. Soc. Simul. 10 (2), 19.
Zila, E., Kukacka, J., 2023. Moment set selection for the SMM using simple machine learning. J. Econ. Behav. Organ. 212, 366–391.