Martinoli, Mario; Moneta, Alessio; Pallante, Gianluca

Working Paper

Calibration and validation of macroeconomic simulation models: A general protocol by causal search

LEM Working Paper Series, No. 2022/33

Provided in Cooperation with:
Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Martinoli, Mario; Moneta, Alessio; Pallante, Gianluca (2022) : Calibration and validation of macroeconomic simulation models: A general protocol by causal search, LEM Working Paper Series, No. 2022/33, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:
https://hdl.handle.net/10419/273635

INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Calibration and Validation of Macroeconomic Simulation Models by Statistical Causal Search

Mario Martinoli [a]
Alessio Moneta [a]
Gianluca Pallante [a]

[a] Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.

2022/33 **October 2022**
ISSN(ONLINE) 2284-0400

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*

Abstract

We introduce a general procedure for 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.

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

JEL classification: C32; C52; E37

[†]Corresponding author. Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa (Italy). Email: *m.martinoli@santannapisa.it*

The authors acknowledge support from the project "How good is your model? Empirical evaluation and validation of quantitative models in economics", PRIN grant no. 20177FX2A7. We are also grateful to Fulvio Corsi, Cristiano Ricci and Raffaello Seri, as well as participants of the conference 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.

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, 1982; Hansen and Heckman, 1996; Cooley, 1997; Gomme and Rupert, 2007). Its scope is to restrict the parameters of a model so that this is 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, if the scope of the model is policy analysis, parameters should be selected so that the model is consistent with 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.

Our idea of calibration has potential overlappings with the strand of literature on calibration 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.

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 to compare different models and evaluating the model's capacity of fitting data, 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). Here, the emphasis has been posed on the idea that validation is about measuring the extent by 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), simulated maximum likelihood method (Lee, 1992; Kristensen and Shin, 2012; Kukacka and Sacht, 2023), and approximate Bayesian computation (Frazier et al., 2018). Frameworks based on surrogate meta-models have also been developed (Lamperti et al., 2018), 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 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); 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 our identification is local and we are not able to label the shocks. 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 through 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). In this work, we use our protocol to validate the "Dystopian Schumpeter meeting Keynes" (DSK) model (Lamperti et al., 2019), a large-scale macroeconomic ABM that includes some characteristics peculiar to "Integrated Assessment Models", such as the energy sector, the Carbon Dioxide emissions produced by manufacturing firms, and the climate damages. We consider a version of the DSK embodying three sectors: the consumption and capital good sector, the banking sector and the energy industry. The focus on ABM is due to the fact that we need non-Gaussianity for identification. While non-Gaussianity is a common feature in data generated by an ABM (see Guerini and Moneta, 2017), is less so in data generated by a general equilibrium model in which a linearized version with normal disturbances is commonly studied (see,

e.g., Smets and Wouters, 2007).[1]

The contributions of this paper can be summarised 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. Third, we present an application 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). Fourth, we apply our general protocol to an agent-based integrated assessment model (Lamperti et al., 2019). This type of models serves as an alternative benchmark to computable general equilibrium models, which have been extensively used in energy and climate policy research, but have faced criticism for their simplistic portrayal of the complex interplay between the economy and the environment (Weyant, 2017; Stern and Stiglitz, 2021). 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

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

actual data. In Section 4, we present our general protocol of calibration and validation. In particular: in Subsection 4.1, we describe the SVAR-ICA approach to identification; in Subsection 4.2, we discuss the MDI used as loss function in the MCS; in Subsection 4.3, we describe the MCS-based calibration procedure; in Subsection 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

We summarize here below our general protocol for calibration and validation. Two steps (1-2) can be seen as preliminary:

1. Select a discrete set $\mathcal{M}^0 := \{1, \ldots, m_0\}$ of configurations of parameters (henceforth, CoPs) from the parameter space of the theoretical model object of the study. A vector of parameters θ_i ($i = 1, \ldots, m_0$) is associated to each CoP. From the same model, for each CoP i, simulate n Monte Carlo runs $\mathbf{z}_{jt}(\theta_i)$, with $j = 1, \ldots, n$ and $t = 1, \ldots, T$. 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 $t = 1, \ldots, \tau$. 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 $\widehat{\mathbf{\Psi}}_0$ the mixing matrix estimated from $\mathbf{y}_t$, and $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ the one estimated from $\mathbf{z}_{jt}(\theta_i)$.

4. Calculate the MDI between $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ and $\widehat{\mathbf{\Psi}}_0$ and record the unique signed-permutation matrix $\mathbf{C}_{ji}$ associated to it, for each j and i.

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 $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)\mathbf{C}'_{ji}$, 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 $\widehat{\mathbf{\Psi}}_0$ 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 represented through a state-space model (Hinkelmann et al., 2011), and the latter can be approximated by a finite-order VAR model (Giacomini, 2013).

In the following, we consider a set of K time series variables $\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})'$, 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 ($i = 1, \ldots, m_0$; $j = 1, \ldots, n$).

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

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

where $\mathbf{\Gamma}_p$ (for lag $p = 0, \ldots, P$) 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 $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$, which we assume to be diagonal. Equation (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 $\mathbf{A}_p = \mathbf{\Gamma}_0^{-1} \mathbf{\Gamma}_p$ ($p = 1, \ldots, P$), and $\mathbf{u}_t = \mathbf{\Gamma}_0^{-1} \boldsymbol{\varepsilon}_t$, i.e. $\mathbf{u}_t$ is a vector of i.i.d. processes with covariance matrix $\mathbf{\Sigma}_{\mathbf{u}} = \mathbb{E}\{\mathbf{u}_t \mathbf{u}_t'\} = \mathbf{\Gamma}_0^{-1} \mathbf{\Sigma}_{\boldsymbol{\varepsilon}} \mathbf{\Gamma}_0^{-1,\prime}$. We call the impact matrix $\mathbf{\Psi}_0 = \mathbf{\Gamma}_0^{-1}$ the *real-world mixing matrix*.

Analogous representation holds for data generated by the simulation model:

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

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

where $\boldsymbol{\varepsilon}_{j,t}(\theta_i)$ and $\mathbf{u}_{jt}(\theta_i)$ are the model's shocks and the reduced-form residuals, respectively. We call the impact matrix $\mathbf{\Psi}_{j,0}(\theta_i) = \mathbf{\Gamma}_{j,0}^{-1}(\theta_i)$ the *model mixing matrix* associated to the j-th Monte Carlo run of the i-th CoP of the simulated model. As well known in the SVAR analysis, the mixing matrix is key for identification. In the proposed method, we will rely on a statistical identification approach.

4 Calibration and validation protocol

We now enter in the core of our calibration-validation procedure. Moving from the VAR representability of our data generating processes, 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) 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.

We now state the assumptions and the theoretical background underlying ICA, which 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, e.g., 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 $\mathbf{u}_t = \boldsymbol{\Psi}_0 \boldsymbol{\varepsilon}_t$, ICA recovers (up to some indeterminacy, see below) $\boldsymbol{\Psi}_0$ and $\boldsymbol{\varepsilon}_t$ from realisations of $\mathbf{u}_t$ on the basis of the following assumption (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.

Notice that the same assumption holds for $\boldsymbol{\varepsilon}_{jt}(\theta_i)$ and $\boldsymbol{\Psi}_{j,0}(\theta_i)$. The indeterminacy relates to the fact that $\boldsymbol{\Psi}_0$ is identified by ICA 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. This means that the order and the scale of the shocks are not identified. Our choice of the minimum-distance index allows us to tackle this issue.[2]

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

[2]It is customary to normalize the SVAR models so that the structural shocks have unit standard deviations, so 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. One can therefore conclude that $\boldsymbol{\Psi}_0$ is identified up to the post-multiplication of a signed permutation matrix $\mathbf{JP}$ (where $\mathbf{J}$ is sign-change matrix, i.e. a diagonal matrix with only +1 or -1 entries in the main diagonal, and $\mathbf{P}$ is a permutation matrix).

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 k-th row of the matrix $\boldsymbol{\Gamma}_0$. The non-Gaussianity is measured using negentropy. Given a continuous random vector $\mathbf{y}$ and a Gaussian vector $\mathbf{x}$ with the same covariance matrix, negentropy is defined as $J(\mathbf{y}) = H(\mathbf{x}) - H(\mathbf{y})$, where $H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) \mathrm{d}\mathbf{y}$ is the differential entropy (Shannon, 1948) and $f(\mathbf{y})$ is the probability density function. To avoid the estimation of the probability density functions, the *fastICA* algorithm exploits the following approximation of negentropy:

$$J(\mathbf{y}) \approx \left[\mathbb{E}(g(\mathbf{y})) - \mathbb{E}(g(\mathbf{z}))\right]^2, \tag{4.1}$$

where $g(\cdot)$ is a specific nonquadratic function of a random variable, i.e. $g(z) = -\exp(z^2/2)$, where $z \sim \mathcal{N}(0,1)$. The approximation devised in Equation (4.1) drastically reduces the computational time to find ICA projections (see Issoglio et al., 2021, for a discussion). Finally, we have the following estimator:

$$\widehat{\boldsymbol{\gamma}}_0 = \arg\max_{\boldsymbol{\gamma}} \mathbb{E}\left[J(\boldsymbol{\gamma}_k' \mathbf{u}_t)\right]. \tag{4.2}$$

The statistical properties of the *fastICA* estimator hold under the following assumption:

Assumption 2. *(i)* $\mathbb{E}[\mathbf{u}_t] = 0$ *and* $\mathbf{u}_t$ *has all moments up to the fourth;*

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

(iii) $g''(\cdot)$ *is bounded.*

Therefore, it can be shown that, under Assumptions 1-2, $\widehat{\gamma}_0 = \text{vec}\left(\widehat{\mathbf{\Gamma}}_0\right)$ is consistent and asymptotically normal (see Reyhani et al., 2012), where $\widehat{\mathbf{\Gamma}}_0$ is the estimator of $\mathbf{\Gamma}_0$.

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 $\boldsymbol{\gamma}_k'\mathbf{u}_t$ 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 $f\left(\boldsymbol{\gamma}_1'\mathbf{u}_t\right) \cdot \ldots \cdot f\left(\boldsymbol{\gamma}_K'\mathbf{u}_t\right)$. 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.

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 $\widehat{\mathbf{\Psi}}_0$ and the true value $\mathbf{\Psi}_0$ 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 permutations and changes of signs of the columns of the model mixing matrix, by keeping the real-world mixing matrix as reference matrix.[3] 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 $D_{ji} := D\left(\widehat{\mathbf{\Psi}}_{j,0}\left(\theta_i\right), \widehat{\mathbf{\Psi}}_0\right)$.

[3] 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 $\widehat{\mathbf{\Psi}}_0$.

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

$$D_{ji} := \frac{1}{\sqrt{K-1}} \inf_{\mathbf{C}_{ji} \in \mathcal{C}} \left\| \mathbf{C}_{ji} \widehat{\mathbf{\Psi}}_{j,0}^{-1}(\theta_i) \widehat{\mathbf{\Psi}}_0 - \mathbf{I}_K \right\|_F \tag{4.3}$$

where

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

$\mathcal{G}$ 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, $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ is the estimator of the model mixing matrix $\mathbf{\Psi}_{j,0}(\theta_i)$, $\widehat{\mathbf{\Psi}}_0$ is the estimator of the real-world mixing matrix $\mathbf{\Psi}_0$ (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 $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ is close to $\widehat{\mathbf{\Psi}}_0$.

Since it implies the minimization over all choices $\mathbf{C}_{ji} \in \mathcal{C}$, 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 $\mathbf{P}$ of equation (4.3) 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 which 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) which delivers 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 $i = 1, \ldots, m_0$. For each CoP i: (i) we run n Monte Carlo simulations $\mathbf{z}_{jt}(\theta_i)$ ($j = 1, \ldots, n$); (ii) we derive the model mixing matrix $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ (for each Monte Carlo j); and (iii) we compute the MDI between $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)$ and the real-world mixing matrix $\widehat{\mathbf{\Psi}}_0$ (for each j).

Let $\overline{D}_i := \mathbb{E}_{\widehat{\mathbf{\Psi}}_0(\theta_i)} D\left(\widehat{\mathbf{\Psi}}_{j,0}(\theta_i), \widehat{\mathbf{\Psi}}_0\right)$ 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 $\overline{\mathbf{D}} := (\overline{D}_1, \ldots, \overline{D}_{m_0})'$ be the m_0-dimensional vector of such expected values for the m_0 CoPs. Let $\overline{D}_i^{(n)} := \frac{1}{n}\sum_{j=1}^n D_{ji}$ and $\overline{\mathbf{D}}^{(n)} := \left(\overline{D}_1^{(n)}, \ldots, \overline{D}_{m_0}^{(n)}\right)'$ be the sample counterparts of $\overline{D}_i$ and $\overline{\mathbf{D}}$ respectively. Defining $\mathbf{D}_j := (D_{j1}, \ldots, D_{jm_0})'$, 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

$$\mathcal{M}^\star := \left\{ h \in \mathcal{M}^0 : \overline{D}_h = \min_{i \in \mathcal{M}^0} \overline{D}_i \right\} \tag{4.4}$$

be the set of parameters minimizing the distance $\overline{D}_i$. For $i \in \mathcal{M}^0$, the estimator $\widehat{i}^{(n)}$ is the value that minimizes the sample average distance $\overline{D}_i^{(n)}$. Note that $\widehat{i}^{(n)}$ 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 $\sigma_i^2 := \mathbb{V}_{\widehat{\mathbf{\Psi}}_0(\theta_i)} D\left(\widehat{\mathbf{\Psi}}_{j,0}(\theta_i), \widehat{\mathbf{\Psi}}_0\right)$, following Seri et al. (2021). To do that, we need the following assumptions.

Assumption 3. *For $j = 1, \ldots, n$ the vectors $\mathbf{D}_j$ are independent and identically distributed. For each $i \in \mathcal{M}^0$, the distances D_{ji} are independent. The mean $\mathbb{E}D_{ji}$ exists*

and is finite for each $i \in \mathcal{M}^0$.

Assumption 3 guarantees consistency and measurability of $\widehat{i}^{(n)}$. 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).

Assumption 4. *The variance* σ_i^2 *is finite for any* $i \in \mathcal{M}^0$.

While Assumption 3 is fulfilled by the construction of the simulation model, Assumption 4 can be 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 $\mathsf{H}_{1,\mathcal{M}}$:

$$\mathsf{H}_{0,\mathcal{M}} : \overline{D}_i = \overline{D}_h, \forall i, h \in \mathcal{M}; \tag{4.5}$$

$$\mathsf{H}_{1,\mathcal{M}} : \exists i, h \in \mathcal{M} \text{ such that } \overline{D}_i \neq \overline{D}_h. \tag{4.6}$$

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 $i = 1, \ldots, m_0 - 1$, 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:

$$\widehat{p}_{e_{\mathcal{M}_h}} := \max_{i \leq h} p_{\mathsf{H}_{0,\mathcal{M}_i}}, \tag{4.7}$$

for $h = 1, \ldots, m_0$.

The implementation of the MCS procedure follows an algorithm in which the main steps are iterated:

1. start from the set $\mathcal{M}^0 \coloneqq \{1, \ldots, m_0\}$ and test, with level $1 - \alpha$, that all the average distances are equal: if $\widehat{p}_{e_{\mathcal{M}_1}} > \alpha$, do not reject $\mathsf{H}_{0,\mathcal{M}}$ and the procedure is over; if $\widehat{p}_{e_{\mathcal{M}_1}} \leq \alpha$, 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 $\mathcal{M}^1 \coloneqq \{1, \ldots, m_0 - 1\}$;

3. test, with level $1-\alpha$, that all the average distances associated with $i \in \mathcal{M}^1$ are equal; again, if the $\widehat{p}_{e_{\mathcal{M}_2}} > \alpha$, do not reject the null hypothesis and the procedure is over; if the $\widehat{p}_{e_{\mathcal{M}_2}} \leq \alpha$ 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 $\widehat{\mathcal{M}}^*$.

Note that our MCS-based calibration procedure can be easily adapted to a MDI which refers not just to the mixing matrices, i.e. $\mathbf{\Psi}_0$, $\mathbf{\Psi}_{j,0}(\theta_i)$, but rather to structural moving-average matrices at different lags, $\mathbf{\Psi}_\ell$, $\mathbf{\Psi}_{j,\ell}(\theta_i)$ (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 fulfills 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 $\mathbf{\Psi}_0$ 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 $\mathbf{\Psi}_{j,0}(\theta_i)$, which contains in principle the same column/sign indeterminacy. However, the calculation of the MDI D_{ji} has delivered a unique matrix $\mathbf{C}_{ji}$ for each j and i. From $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)\,\mathbf{C}'_{ji}$, 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, that can be inferred by exploiting the Monte Carlo simulations.

We represent the shocks-variables structure via a matrix which we call "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 $< k_v, k_s >$ is 1 if and only if there is a significant impact of the shock $\varepsilon_{k_s,t}$ (or $\varepsilon_{k_s,t}(\theta_i)$) on the variable $y_{k_v,t}$ (or $z_{k_v,t}(\theta_i)$), 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 $\widehat{\mathbf{\Psi}}_0$ and $\widehat{\mathbf{\Psi}}_{j,0}(\theta_i)\,\mathbf{C}'_{ji}$, with $i \in \mathcal{M}^* \subseteq \mathcal{M}^0$. As regards $\widehat{\mathbf{\Psi}}_0$, 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 $\widehat{\mathbf{\Psi}}_0^{n^*}$ is right-multiplied by a signed permutation matrix $\mathbf{C}'_{n^*}$, where $\mathbf{C}_{n^*}$ corresponds to the *arg inf* of the MDI between $\widehat{\mathbf{\Psi}}_0^*$ and $\widehat{\mathbf{\Psi}}_0$ (consistently to the scheme we apply to the model mixing matrices).

Once obtained ICRs for both synthetic and real data we calculate the Structural Hamming Distance (SHD) 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 (VM) as follows:

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

$$\text{VM} := 1 - {}^{SHD}/_{K^2}, \tag{4.8}$$

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

If $\text{SHD} \ll K^2$, then $\text{VM} \to 1$. 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. 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.

5 The DSK model

In this section, we briefly illustrate the DSK model by Lamperti et al. (2019), 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 analyze the public costs of climate-induced financial instability (Lamperti et al., 2019), 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. (2019) 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 its 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 pro-

duced 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 Appendix A.

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. 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., 2019). 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., 2019, and references therein). The price of consumption-good firms is determined as a mark-up on the unit production cost that changes over time. We let the initial mark-up $\overline{\mu}_0$ vary between 0.2 and 0.3 (see Section A.2). The mark-down for bank deposits μ^{dep} and the mark-down on bank reserves deposited at the Central Bank μ^{res} regulate banks' profit margins. We vary μ^{dep} between 0.75 and 0.95 and μ^{res} between 0.5 and 0.9 (details in Section A.4). The firm search capabilities parameters $\zeta_{1,2}$ vary between 0.3

Table 6.1: Parameter notations and interval values

Description	Parameter	Values
Consumption-good firm initial mark-up	$\overline{\mu}_0$	$[0.2, 0.3]$
Mark-down for bank deposits	μ^{dep}	$[0.75, 0.95]$
Mark-down on the bank reserves at Central Bank	μ^{res}	$[0.5, 0.9]$
Firm search capabilities parameters	$\zeta_{1,2}$	$[0.3, 1]$
$Beta$ distribution support (innovation)	$\underline{x}$	$[-0.15, -0.05]$
	$\overline{x}$	$[0.05, 0.15]$
$Beta$ distribution support (energy)	$\underline{x}_{\text{en}}$	$[-0.1, -0.01]$
	$\overline{x}_{\text{en}}$	$[0.01, 0.1]$
Payback period (industrial)	b	$[2, 3.75]$

and 1. This feature determines whether a firm is able to access to innovation or imitation. When a firm innovates, the probability of drawing a new machinery is modelled according to a $Beta$ distribution. The parameter $\underline{x}$ is the lower bound of the support of the $Beta$ distribution for innovation 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 $\underline{x} = -0.15$, $\overline{x} = 0.15$ (Section A.1). We adopt the same reasoning for the support of the $Beta$ distribution related to green technologies. In this case $\underline{x}_{\text{en}}$ varies between -0.1 and -0.01, while $\overline{x}_{\text{en}}$ varies between 0.01 and 0.1. Finally, the range of variation of the payback period parameter b for the industrial sector, measuring the replacement of a machinery with respect to its obsolescence, goes from 2 to 3.75 (Section A.2).

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 uni-

formity 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 $O\left(N^{-1}\right)$ (while Monte Carlo techniques hold a convergence rate of $O\left(N^{-1/2}\right)$).

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.[4] We take logs of all variables except for the unemployment rate. A description of the empirical dataset is provided in Table 6.2.

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	Milion Metric Tons of CO2	Total Carbon Dioxide emissions

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.[5]

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 $\widehat{\mathbf{\Psi}}_0$ and $\widehat{\mathbf{\Psi}}_{j,0}\left(\theta_i\right)$ is achieved via *fastICA* from the estimated reduced-form residuals $\widehat{\mathbf{u}}_t$ and $\widehat{\mathbf{u}}_{jt}(\theta_i)$. On these residuals,

[4]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/.

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

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:

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

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. (2019), 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}); (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 ($\underline{x}_{\text{en}}, \overline{x}_{\text{en}}$); (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.

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 $\delta_{\mathcal{M}_k}$ ($p_{\mathsf{H}_{0,\mathcal{M}_k}}$)	MCS p-value ($\widehat{p}_{e_{\mathcal{M}_k}}$)	$\overline{D}_i^{(200)}$
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. (2019).

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
$\underline{x}_{\text{en}}$	-0.0566	-0.058
$\overline{x}_{\text{en}}$	0.0566	0.058
b	2.59	3

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 $\widehat{\mathbf{\Psi}}_0$, using bootstrap, and the significance of the entries of the matrix $\widehat{\mathbf{\Psi}}_{j,0}(\theta_{35})$, relying on the distributions of the Monte Carlo runs. Then, we infer ICR for both the simulated and the actual causal structures. The ICR_{rw} and ICR_{sim} are reported in Table 6.5. To simplify the notation, we write $\varepsilon_{k_s,t} = \varepsilon_{k_s}$, with $k_s = 1, \ldots, 7$.

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

	ICR_{rw}								ICR_{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

The analysis of the ICRs provides the following outcomes: (i) as regards the ICR_{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 ICR_{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 $\text{VM} = 1 - {}^{14}/_{49} = 0.714$, therefore, the simulated model is able to recover the 71.4% of the real-world shocks-variables structures.

It is worth noting that CoP 35 is the "best" configuration across different SVAR specifications, confirming the robustness of the calibration exercise (see Appendix C). The analysis provided in Appendix C shows also that a set of shocks hitting energy and investment, identified in the simulated model, match quite accurately their empirical counterpart.

7 Conclusion

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. (2019). 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.

References

Acid, S. and L. M. de Campos (2003). Searching for Bayesian network structures in the space of restricted acyclic partially directed graphs. *Journal of Artificial Intelligence Research 18*, 445–490.

Altissimo, F. and A. Mele (2009). Simulated Non-Parametric Estimation of Dynamic Models. *Review of Economic Studies 76*(2), 413–450.

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

Barde, S. (2020). Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion. *Journal of Economic Dynamics and Control 111*, 103795.

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

Cantore, C., V. J. Gabriel, P. Levine, J. Pearlman, and B. Yang (2013). In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pp. 441–463. Edward Elgar Publishing.

Casini, L., A. Moneta, and M. Capasso (2021). Variable definition and independent components. *Philosophy of Science 88*(5), 784–795.

Choirat, C. and R. Seri (2012). Estimation in Discrete Parameter Models. *Statistical Science 27*(2), 278–293.

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

Christiano, L. J., M. S. Eichenbaum, and M. Trabandt (2018). On DSGE models. *Journal of Economic Perspectives 32*(3), 113–40.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Cooley, T. F. (1997). Calibrated models. *Oxford Review of Economic Policy 13*(3), 55–69.

Cúrdia, V., M. Del Negro, and D. L. Greenwald (2014). Rare shocks, great recessions. *Journal of Applied Econometrics 29*(7), 1031–1052.

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

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of new Keynesian models. *Journal of Business & Economic Statistics 25*(2), 123–143.

Delli Gatti, D. and J. Grazzini (2020). Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic Agent Based Models. *Journal of Economic Behavior & Organization 178*, 875–902.

Dosi, G., G. Fagiolo, M. Napoletano, A. Roventini, and T. Treibich (2015). Fiscal and monetary policies in complex evolving economies. *Journal of Economic Dynamics and Control 52*, 166–189.

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

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

Fagiolo, G., M. Guerini, F. Lamperti, A. Moneta, and A. Roventini (2019). Validation of Agent-Based Models in Economics and Finance. In C. Beisbart and N. J. Saam (Eds.), *Computer Simulation Validation*, pp. 763–787. Cham: Springer.

Favero, C. A. (2001). *Applied Macroeconometrics*. Oxford University Press.

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

Frazier, D. T., G. M. Martin, C. P. Robert, and J. Rousseau (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*, Volume 32 of *Advances in Econometrics*, pp. 1–25. Emerald Group Publishing Limited.

Gomme, P. and P. Rupert (2007). Theory, measurement and calibration of macroeconomic models. *Journal of Monetary Economics 54*(2), 460–497.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*(S1), S85–S118.

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

Guerini, M. and A. Moneta (2017). A method for agent-based models validation. *Journal of Economic Dynamics and Control 82*, 125–141.

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

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

Hansen, L. P. and J. J. Heckman (1996). The Empirical Foundations of Calibration. *Journal of Economic Perspectives 10*(1), 87–104.

Hansen, P. R., A. Lunde, and J. M. Nason (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. *Oxford Bulletin of Economics and Statistics 80*(4), 736–754.

Herwartz, H. and M. Plödt (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. *Journal of International Money and Finance 61*, 30–44.

Hinkelmann, F., D. Murrugarra, A. S. Jarrah, and R. Laubenbacher (2011). A mathematical framework for agent based models of complex biological networks. *Bulletin of Mathematical Biology 73*(7), 1583–1602.

Hoover, K. D. (2012). Economic theory and causal inference. In U. Mäki (Ed.), *Handbook of the Philosophy of Economics*, pp. 89–113. Elsevier/North-Holland.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks 10*(3), 626–634.

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*(1984), 20110534.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent component analysis*. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. New York: J. Wiley.

Hyvärinen, A. and E. Oja (2000). Independent component analysis: algorithms and applications. *Neural Networks 13*(4-5), 411–430.

Ilmonen, P., K. Nordhausen, H. Oja, and E. Ollila (2010). A New Performance Index for ICA: Properties, Computation and Asymptotic Analysis. In V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent (Eds.), *Latent Variable Analysis and Signal Separation*, Volume 6365, pp. 229–236. Berlin: Springer. Series Title: Lecture Notes in Computer Science.

Ireland, P. N. (2004). A method for taking models to the data. *Journal of Economic dynamics and control 28*(6), 1205–1226.

Issoglio, E., P. Smith, and J. Voss (2021). On the estimation of entropy in the FastICA algorithm. *Journal of Multivariate Analysis 181*, 104689.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kristensen, D. and Y. Shin (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics 167*(1), 76–94.

Kucherenko, S., D. Albrecht, and A. Saltelli (2015). Exploring multi-dimensional spaces: a comparison of Latin Hypercube and Quasi Monte Carlo sampling techniques.

Kukacka, J. and S. Sacht (2023). Estimation of heuristic switching in behavioral macroeconomic models. *Journal of Economic Dynamics and Control 146*, 104585.

Kydland, F. E. and E. C. Prescott (1982). Time to Build and Aggregate Fluctuations. *Econometrica 50*(6), 1345.

Lamperti, F., V. Bosetti, A. Roventini, and M. Tavoni (2019). The public costs of climate-induced financial instability. *Nature Climate Change 9*(11), 829–833.

Lamperti, F., V. Bosetti, A. Roventini, M. Tavoni, and T. Treibich (2021). Three green financial policies to address climate risks. *Journal of Financial Stability 54*, 100875.

Lamperti, F., G. Dosi, M. Napoletano, A. Roventini, and A. Sapio (2020). Climate change and green transitions in an agent-based integrated assessment model. *Technological Forecasting and Social Change 153*, 119806.

Lamperti, F., A. Roventini, and A. Sani (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control 90*, 366–389.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of 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. *Carnegie-Rochester Conference Series on Public Policy 1*, 19–46.

Matteson, D. S. and R. S. Tsay (2017). Independent Component Analysis via Distance Covariance. *Journal of the American Statistical Association 112*(518), 623–637.

McCracken, M. and S. Ng (2020, March). 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., D. Entner, P. O. Hoyer, and A. Coad (2013). Causal Inference by Independent Component Analysis: Theory and Applications. *Oxford Bulletin of Economics and Statistics 75*(5), 705–730.

Moneta, A. and G. Pallante (2022). Identification of structural var models via independent component analysis: A performance evaluation study. *Journal of Economic Dynamics and Control 144*, 104530.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences 114*(7), 201609244.

Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica 57*(5), 1027–1057.

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science 87*(3), 457–477.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics 54*(7), 2048–2064.

Reyhani, N., J. Ylipaavalniemi, R. Vigário, and E. Oja (2012). Consistency and asymptotic normality of FastICA and bootstrap FastICA. *Signal processing 92*(8), 1767–1778.

Secchi, D. and R. Seri (2017). Controlling for false negatives in agent-based models: a review of power analysis in organizational research. *Computational and Mathematical Organization Theory 23*(1), 94–121.

Seri, R., M. Martinoli, D. Secchi, and S. Centorrino (2021). Model calibration and validation via confidence sets. *Econometrics and Statistics 20*, 62–86.

Seri, R. and D. Secchi (2017). How Many Times Should One Run a Computational Simulation? In B. Edmonds and R. Meyer (Eds.), *Simulating Social Complexity: A Handbook*, Understanding Complex Systems, pp. 229–251. Cham: Springer International Publishing.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal 27*(3), 379–423.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review 97*(3), 586–606.

Smith, A. A. J. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*(S1), S63–S84.

Sobol', I. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics 7*(4), 86–112.

Stern, N. and J. E. Stiglitz (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., L. E. Brown, and C. F. Aliferis (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine learning 65*(1), 31–78.

Weyant, J. (2017). Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy 11*(1), 115–137.

Windrum, P., G. Fagiolo, and A. Moneta (2007). Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation 10*(2), 19.

Appendix

A The DSK model: Detailed description

This appendix provides a description of the DSK model's structure of Lamperti et al. (2019, 2021). We start with an account of the *capital-good sector*, which determines how research, production and pricing decisions are taken. We then describe the *consumption good sector* and the *energy sector*. After having described the functioning of the *banking sector*, we outline the equations that refer to the *public sector* and to the conduct of the *monetary authorities*. Finally, we sketch the functioning of the *climate module*.

A.1 Capital-good sector

Capital goods firms produce a machine of vintage τ, whose technology is characterized by a given level of labor productivity (LP), energy efficiency (EE) and environmental friendliness (EF). These features are represented by a set of coefficients $(A^l_{i,\tau}, B^l_{i,\tau})$, where $l = \{LP, EE, EI\}$.

The coefficient $A^{LP}_{i,\tau}$ can be considered as the productivity of the machinery in the consumption good industry; $B^{LP}_{i,\tau}$ is the productivity of the process leading to the manufacturing of the capital good. Analogously, $A^{EE,EF}_{i,\tau}$ and $B^{EE,EF}_{i,\tau}$ represent the level of energy efficiency (EE) and environmentally friendliness (EF) in the production processes of both type of goods.

In order to gain market power and be selected by the competitive forces of the market, upstream firms undertake a costly process for technology improvement, both via innovation and imitation. Following Dosi et al. (2010), both activities are modelled in two steps. In the first, the dynamics of technical change randomly determines the probability of success: this is modelled by realization of Bernoulli-distributed random variables in which the level of R&D investments $INNOV_i(t)$ positively determines the probability that the innovation is successful, namely with parameter $\vartheta^{in}(t) = 1 - \exp^{-\zeta_1 INNOV_i(t)}$. Similarly, R&D investments for imitation of existing technologies, $IMIT_i(t)$, are positively correlated with the probability of successful imitation $\vartheta^{im}(t) = 1 - \exp^{-\zeta_2 IMIT_i(t)}$. In a second step, the characteristics of the machines which result from the first process are determined. The same process applies both in case of imitation and innovation. The new discovered machinery (i.e. vintage $\tau + 1$) is described as:

$$A^k_{i,\tau+1} = A^k_{i,\tau}(1 + \chi^k_{A,i}) \quad \text{for } k = LP, EE \tag{A.1}$$

$$B^k_{i,\tau+1} = B^k_{i,\tau}(1 + \chi^k_{B,i}) \quad \text{for } k = LP, EE, \tag{A.2}$$

$$A^{EF}_{i,\tau+1} = A^{EF}_{i,\tau}(1 - \chi^{EF}_{A,i}) \tag{A.3}$$

$$B^{EF}_{i,\tau+1} = B^{EF}_{i,\tau}(1 - \chi^{EF}_{B,i}), \tag{A.4}$$

where $\chi^k_{A,i}$ and $\chi^k_{B,i}$ are independent draws from $Beta(\alpha^k, \beta^k)$ distributions over the supports $[\underline{x}^k, \bar{x}^k]$, respectively for $k \in \{LP, EE, EF\}$. The higher the support of each distribution, the higher the technological opportunities. In Lamperti et al. (2019) the support is equal for all levels of k, so that $\underline{x}^k = \underline{x}$ and $\bar{x}^k = \bar{x}$.

A.2 Consumption good sector

Consumer-good firms produce a homogeneous good using their stock of machines, energy and labour under constant returns to scale. Workers consumption determines the level of demand to be satisfied and accordingly, firms adaptively plan their production quantities Q^d_j (also considering desired inventories (N^d_j) and the actual stock (N_j)) according to the expected level of demand $D^e_j = f[D_j(t-1), D_j(t-2), ..., D_j(t-h)]$:

$$Q_j(t)^d = D^e_j(t) + N^d_j(t) - N_j(t), \tag{A.5}$$

where $N_j(t) = \iota D^e_j(t),\ \iota \in [0, 1]$.

The production levels of downstream firms are constrained by the level of their capital stock (K^d). Accordingly, if the current capital, $K_j(t)$, is not sufficient to satisfy the desired level of production, they can invest and buy new machineries.

$$EI^d_j(t) = K^d_j(t) - K_j(t). \tag{A.6}$$

Firms also invest by replacing technologically obsolete machineries that is, for a given set of capital goods $\Xi_j(t)$, the vintage τ is substituted with a more productive one if

$$\frac{p^{new}}{c^{con}_j(t) - c^{new}} = \frac{p^{new}}{\frac{w(t)}{A^{LP}_{i,\tau}} + \frac{c(t)^{en}}{A^{EE}_{i,\tau}} - c^{new}_j} \leq b, \tag{A.7}$$

with p^{new} and c^{new} being the price of the machinery and its unitary cost of production, respectively. The parameter b discounts firms' "patience" on the rate of return on investments.[6]

To finance their investments, consumption good firms operate in imperfect credit markets and prioritize internal funds. If the latter are not sufficient to fully cover production plans and investments, external funds are borrowed from a bank in the form of a credit line. Given the total credit supply of a bank, the latter lends out money to firms on a pecking-order, determined by the ratio between equity and sales. In the case the credit demand is greater than the credit supply of the borrower, credit-rationing is observed.

The pricing decision of consumption good firms also follow the rule of charging a

[6]The unitary cost of production for both consumption and capital goods firms are as follows: $c^{con}_j(t) = w(t)/A^{LP}_{i,\tau} + p^e(t)/A^{EE}_{i,\tau}$. The price of any machinery is set with a fixed mark-up $\mu_1 > 0$.

markup over the unit cost of production:

$$p_j^{con}(t) = c_j^{con}(t)[1 + \mu_j(t)]. \tag{A.8}$$

The distribution of markups' levels depends on selection processes of the markets in which firms operate. In particular, it depends on the evolution firms' market share, f_j:

$$\mu_j(t) = \mu_j(t-1)\left[1 + \upsilon \frac{f_j(t-1) - f_j(t-2)}{f_j(t-2)}\right], \tag{A.9}$$

with $0 \leq \upsilon \leq 1$.

Moreover, market shares evolution is determined by a "quasi replicator" mechanism: less competitive firms do not survive the selection process and are driven out from the market.

At the end of every period, all firms' profits (net of taxes) are computed and the level of cash reserves is updated. They are calculated as follows:

$$\Pi_j(t) = S_j(t) + r^D NW_j(t-1) - c_j(t)Q_j(t) - r_j^{deb}(t)Deb_j(t), \tag{A.10}$$

with total sales $S_j(t) = p_j(t)D_j(t)$, production costs $c_j(t)Q_j(t)$, and debt costs $r_j^{deb}(t)Deb_j(t)$, where Deb denotes the stock of debt. The tax rate for firm profits is tax_p. The stock of liquid assets ($NW_j(t)$) is defined as:

$$NW_j(t) = NW_j(t-1) + (1-tr)\Pi_j(t) - CI_j(t), \tag{A.11}$$

with CI_j internal funds used by firm j to finance investment. If net wealth is negative or the market share goes to zero, a firm exits the market and it is replaced by a new entrant that charges an initial mark-up $\bar{\mu_0}$.

A.3 Energy sector

The energy sector is characterized by green plants (whose variables are labelled with ge, i.e. "green energy") and by brown plants (labelled with de, i.e. "dirty energy"). Energy is provided as input for the production of the capital and consumption goods. Demand for electricity, D_e, and aggregate energy production, Q_e, are matched from the energy portfolio of plants, since it is assumed that energy cannot be stored.

"Brown" plants use fossil fuel to produce energy with vintage-specific thermal efficiencies A_{de}^{τ}, which indicates how much energy is produced for units of employed non-renewable resources. Energy production also implies CO2 emissions, depending on the emission intensity em_{de}^{τ} (amount of emissions per unit of energy produced).[7] The average

[7]The parameter τ denotes the technology vintage.

production cost for a brown plant of vintage τ is:

$$c_{de}(\tau, t) = \frac{p_f(t)}{A^{\tau}_{de}(t)}, \tag{A.12}$$

where p_f is the price of fossil fuels, exogenously determined on an international market.

"Green" plants produce (from renewable resources) energy at a null production cost, i.e., $c_{ge}(t) = 0$. Given that the total (potential) production of green plants is K_{ge}, IM is the set of plants which should be gradually activated to meet the effective energy demand. Given the costs of production using brown and green plants, green power plants are employed first. Instead, if total green installed capacity $D_e(t)$ is greater than $K_{ge}(t)$, some dirty plants need to be activated too. The total production cost then corresponds to the sum of the production costs of the brown plants that are activated. Assuming that the absolute frequency of vintage τ plants is $g_{de}(\tau, t)$, if dirty plants are operative the total production cost is:

$$PC_e(t) = \sum_{\tau \in IM} g_{de}(\tau, t) c_{de}(\tau, t) A^{\tau}_{de}(t) \tag{A.13}$$

The energy price is computed by adding a fixed markup $\mu_e \geq 0$ to the average cost of the most expensive infra-marginal plant:

$$p_e(t) = \mu_e, \tag{A.14}$$

if $D_e(t) \leq K_{ge}(t)$, and

$$p_e(t) = \overline{c}_{de}(\tau, t) + \mu_e \tag{A.15}$$

if $D_e(t) > K_{ge}(t)$, where $\overline{c}_{de}(\tau, t) = \max_{\tau \in IM} c_{de}(\tau, t)$. This unit cost level ensures that all infra-marginal plants are obtaining positive profits.

Investment in the energy sector is due to the replacement of obsolete power plants (after η periods they cannot produce energy anymore) or for capacity expansion. Expansionary investments are undertaken if the maximum electricity production level $\overline{Q}_e(t)$ is lower than electricity demand $D_e(t)$. The amount of new expansion investments EI_e thus equals:

$$EI_e(t) = K^d_e(t) - K_e(t), \tag{A.16}$$

if $\overline{Q}_e(t) < D_e(t)$, whereas $EI_e(t) = 0$ if $\overline{Q}_e(t) \geq D_e(t)$. What type of new plants will be installed in this case? Construction costs for new dirty plants are normalized to zero, whilst the construction of new green plants of vintage τ come with a fixed cost IC^{τ}_{ge}. The decision is based upon a profitability rule according to which green plants are compared to brown counterparts in terms of lifetime costs. This means that green energy technologies are chosen whenever the fixed cost of building the cheapest green vintage is below the discounted (variable) production cost of the most efficient dirty plant. Hence, the following payback rule should be satisfied:

$$\underline{IC}_{ge} \leq b_{en} \cdot \underline{c}_{de}, \tag{A.17}$$

where b_{en} is a payback period parameter (as in Dosi et al., 2010), $\underline{IC}_{ge} = \min_\tau IC^\tau_{ge}$, and $\underline{c}_{de} = \min_\tau c^\tau_{de}$.[8] Accordingly, in case of new green capacity, the expansion investment cost amounts to

$$EC_e(t) = \underline{IC}_{ge} EI_e(t); \tag{A.18}$$

whereas it is zero if the payback rule is not met and the firm builds new dirty plants.

As it is for the capital-good sector, plants' costs and characteristics vary due to the process of technical change. Energy plants invest a fraction $v_e \in (0, 1)$ of total past sales in R&D and stochastically improve their cost structure and emission intensity through a two-step procedure. At the end of the period, the central authority computes profits in the energy sector (see eq. A.21 below) and levies taxes at the rate tax_p.

R&D expenses in the energy sector aim at improving the technology of green and dirty plants. Better green plants are characterized by lower fixed costs of installation. Instead, more technological advances for dirty plants lead to an increase in energy efficiency (A) and a reduction in carbon emissions (em). The budget for innovative activities is split between green innovations ($IN_{ge}(t) = \xi_e RD_e(t)$) and dirty ones ($IN_{de}(t) = (1 - \xi_e)RD_e(t)$). The level of these expenses positively affect the probability of having access to the first step of the innovation step. For example, for the case of green innovations we have:

$$\theta_{ge}(t) = 1 - e^{-\eta_{ge} IN_{ge}(t)} \tag{A.19}$$

with $\eta_{ge} \in (0, 1)$. A similar process underlies the access to dirty innovations.

In the second step, energy firms may be successful or not in the process of discovery. For green plants the improvement will reduce the fixed costs of constructing and installing the plant by a factor $x_{ge} \in (0, 1)$ (a random draw from a $Beta$ distribution over the support $[\chi_{en}, \bar{\chi}_{en}]$) with respect to the previous vintage:

$$IC^\tau_{ge} = IC^{\tau-1}_{ge} x_{ge}$$

For brown plants, two independent random draws x^A_{de} and x^{em}_{de} (again, from a $Beta$ distribution) will affect the thermal efficiency and the emission intensity as follows:

$$A^\tau_{de} = A^{\tau-1}_{de}(1 + x^A_{de}) \qquad em^\tau_{de} = em^{\tau-1}_{de}(1 - x^{em}_{de}) \tag{A.20}$$

Finally, the profits $\Pi_e(t)$ of the energy monopolist are calculated as follows:

$$\Pi_e(t) = p_e(t)D_e(t) - PC_e(t) - IC_e(t) - RD_e(t), \tag{A.21}$$

where $p_e(t)$ is the energy price, $D_e(t)$ the quantity produced. On the expenditure side $PC_e(t)$, $IC_e(t)$ and $RD_e(t)$ are the total, investments and R&D costs, respectively.

The energy firm then pays taxes on positive profits at the rate tax_p.

[8] Under the assumption that plants are utilized for energy production the same number of periods, equation A.17 boils down to a comparison of levelized costs of energy.

A.4 Banking Sector

Following Dosi et al. (2015), the banking sector consists of B commercial banks that collect deposits from households and workers and supply credit to firms, as well as a central bank that manages monetary policy and buys government bonds as needed. Banks vary in the number of clients they serve, their balance sheet structure, and their lending conditions. Imperfect information makes it difficult for firms to find the best lending rates, so the bank-firm network is assumed to be fixed and based on the empirical distribution of bank size.

A key decision for a financial institution is how much credit to offer to its clients. It is assumed that the supply of credit is a multiple of a bank's net worth (i.e., equity):

$$TC_b(t) = \frac{NW_b(t-1)}{\tau_{CAR}\left(1 + \frac{\beta BD_b(t-1)}{TA_b(t-1)}\right)}, \tag{A.22}$$

with $TC_b(t)$ being total credit supplied by bank b at time t, $NW_b(t-1)$ is the bank's equity at time $(t-1)$ and $TA_b(t-1)$ the value of total assets in $(t-1)$, τ_{CAR} the policy parameter that governs capital adequacy requirements, β is a behavioural parameter that allows banks to have a prudential buffer over the capital requirement the more sensitive it is to financial fragility of the balance-sheet. The magnitude of this buffer changes over the business cycle, in particular according to the ratio of "bad debt" ($BD_b(t-1)$, indicating the amount of non-performing loans in $(t-1)$) and total assets of bank b.

The interest rate on private loans $r^{deb}(t)$ is determined by two factors. The first is a mark-up (μ^{deb}), which is the same for all banks and is applied to the Central Bank interest rate $r^{cb}(t)$. The second factor is a firm-specific risk premium $r^{deb}_{k,cl}(t)$, which reflects the financial fragility of the firm cl. This risk premium is calculated by dividing firms into four credit classes ($q^{cl} = 1, 2, 3, 4$) based on their financial stability. According to Dosi et al. (2015), the risk premium charged by bank k to firm cl at time t is given by:

$$r^{deb}_{k,cl}(t) = r^{deb}(t)(1 + (q^{cl} - 1)kconst), \tag{A.23}$$

where $r^{deb}(t)$ is the base loan interest rate defined as $r^{deb}(t) = r^{cb}(t)(1 + \mu^{deb})$. Firm deposits are paid at an interest rate of $r^{dep}(t)$, bonds are rewarded at a rate of $r^{bonds}(t)$, and deposits are paid at the central bank facility cost $r^{res}(t)$. The entire interest rate structure is such that the following inequalities hold:

$$r^{dep}(t) < r^{res}(t) < r^{cb}(t) < r^{bonds}(t) < r^{deb}(t). \tag{A.24}$$

It is important to note that $r^{dep}(t) = r(t)(1 - \mu^{dep})$ and $r^{res}(t) = r(t)(1 - \mu^{res})$. The main refinancing operation rate is set by the central bank according to a dual-mandate Taylor rule:

$$r^{cb}(t) = r^T + \gamma\pi(\pi(t) - \pi^T) + \gamma_U(U(t) - U^T), \quad \text{with} \quad \gamma_\pi > 1, \quad \gamma_U \geq 0 \tag{A.25}$$

where the terms in parentheses represent the gaps in the inflation rate and unemployment rate, while the parameters (γ_π, γ_U) measure the central bank's aggressiveness in pursuing each objective.

A.5 The public sector

The public sector collects taxes on incomes generated by firm profits and household wages, while government expenditures are in the form of unemployment subsidies, proportional to the current level of market wage. Institutional, market-related and macroeconomic factors affect wage levels. Accordingly, they ultimately depend on the inflation gap, average productivity, and unemployment rate, as follows:

$$\frac{\Delta w(t)}{w(t-1)} = \pi^T + \psi_1[\pi(t-1) - \pi^T] + \psi_2 \frac{\Delta \overline{AB}(t)}{\overline{AB}(t-1)} - \psi_3 \frac{\Delta U(t)}{U(t-1)}, \tag{A.26}$$

where $\overline{AB}$ stands for the economy-wide average productivity and $\psi_1, \psi_2, \psi_3 > 0$.

The sum of all unemployment subsidies adds up to the level of Government expenditures $G(t) = w^u(t)[L^S - L^D(t)]$. Since workers consume all their income, aggregate consumption is determined by the sum of all incomes, both from employed and unemployed: $C(t) = w(t)L^D(t) + G(t)$.

The tax rate is fixed at tr. Public expenditures also comprises the bank bailout costs. Public deficit is calculated accordingly and is set to be equal to $Def(t) = Debt^{cost}(t) + G^{bailout}(t) + G(t) - Tax(t)$. Whenever the deficit is positive, the Government issues bonds that are acquired by banks according to their size.

All the aggregate variables are then the result of microeconomic behavior and interaction. Since there are not intermediate goods, aggregate production is the sum of firms' value added. National accounting entities are also met: the value of total production corresponds to the sum of aggregate consumption, investment and change in inventories ($\Delta N(t)$):

$$\sum_{i=1}^{F_1} Q_i(t) + \sum_{j=1}^{F_2} Q_j(t) = Y(t) = C(t) + I(t) + \Delta N(t). \tag{A.27}$$

A.6 Climate module

The relationship between the economy and climate change is straightforward and relies on the well-established linear relationship between cumulative emissions and temperature increases. Economic losses resulting from changes in temperature are firm-specific, via damages to labour or capital production factors:

$$\frac{\Delta TA}{\Delta CE} = \lambda_{CCR}, \tag{A.28}$$

with Δ being the yearly variations and λ_{CCR} the carbon-climate response.

The distribution of climate damage shocks depends on the temperature level. Indeed, following Nordhaus (2017), the damages function $\Omega(t)$ is a quadratic function of temperature levels,

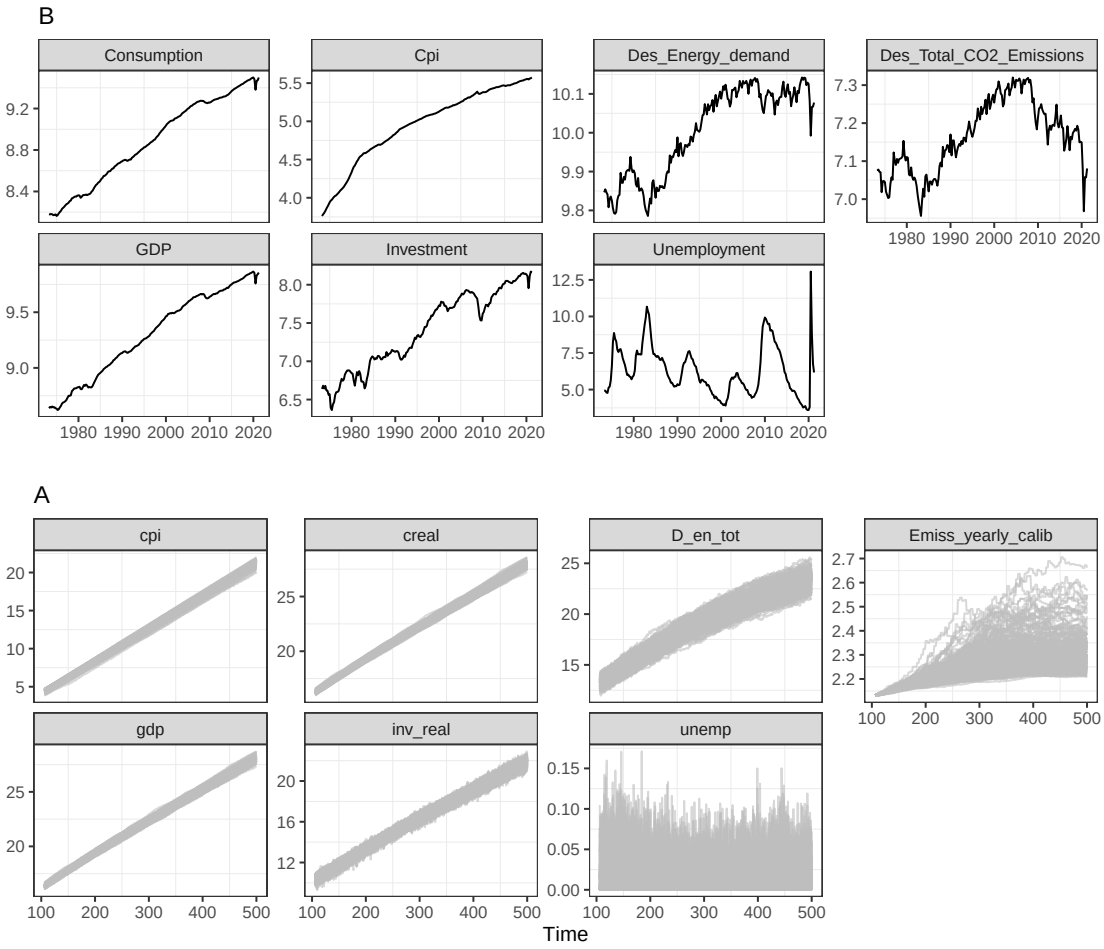
$$\Omega(t) = \frac{1}{1 + c_1 TA(t) + c_2 TA(t)^2}, \tag{A.29}$$

with $c_1, c_2 \in [0; 1]$ first and second-order response parameters, respectively. However, for the sake of the calibration and validation exercise we focus on the baseline version of the model where climate-induced shocks are not hitting the agents operating in the economy, with CO2 emissions being the only economic feedback that affects the environment.

B Behaviour of the validated simulated variables

Figure B.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.

Figure B.1: Plots of the real-world vs simulated variables obtained with CoP 35.



C 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, that we call SVAR_1, SVAR_2, SVAR_3 and SVAR_4. The composition of these SVAR models is reported in Table C.1.

Table C.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 C.2 (for readability, we show only the first ten and the last ten eliminated CoPs).

Table C.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	$e_{\mathcal{M}_k}$	$\widehat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\widehat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\widehat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$	$e_{\mathcal{M}_k}$	$\widehat{p}_{e_{\mathcal{M}_k}}$	$\overline{D}_i^{(200)}$
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 C.3, we devise the validation measures and the shocks-variables structures common to both simulated and actual data, for all the SVAR specifications.

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

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

D Moving-average representation and impulse response functions

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

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

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

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

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

where $\mathbf{\Psi}_\ell = \mathbf{\Phi}_\ell \mathbf{\Gamma}_0^{-1}$ and, in particular, $\mathbf{\Psi}_0 = \mathbf{\Gamma}_0^{-1}$. The entries of the matrices $\mathbf{\Psi}_\ell$, for $\ell = 0, \dots, H$, 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 $\mathbf{\Psi}_\ell$.

If the $\mathbf{y}_t$ contains processes with unit roots, although the VAR model does not admit a Wold representation, the matrices $\mathbf{\Phi}_\ell \mathbf{\Gamma}_0^{-1} = (\sum_{d=1}^{\ell} \mathbf{\Phi}_{\ell-d} \mathbf{A}_d) \mathbf{\Gamma}_0^{-1}$ 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:

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

The impulse response functions for VAR_2, obtained using real-world data, are displayed in Figure D.1, while the impulse response functions for VAR_2, estimated using the data simulated from the model, are shown in Figure D.2.

Figure D.1: Plots of the real-world IRFs of VAR_2.

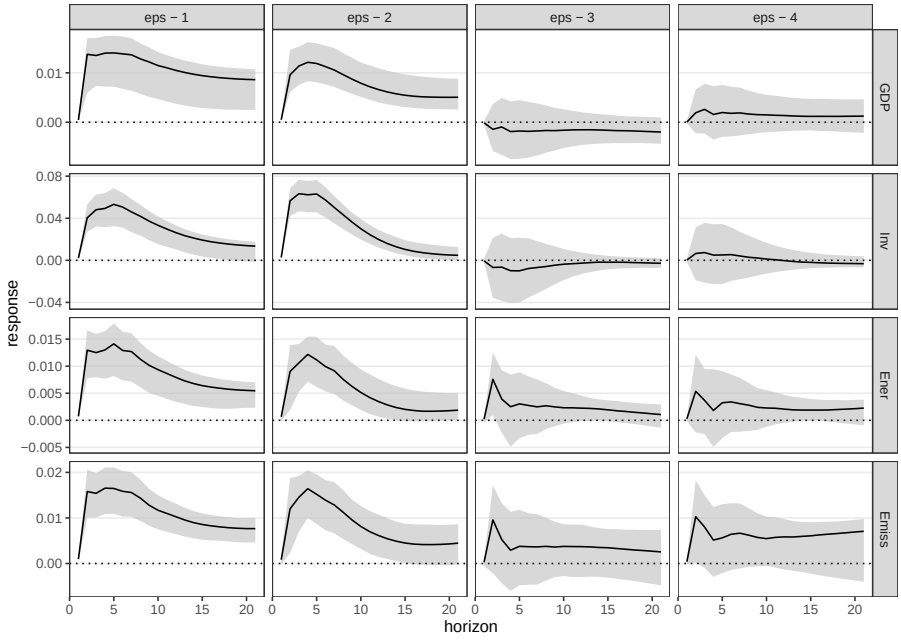
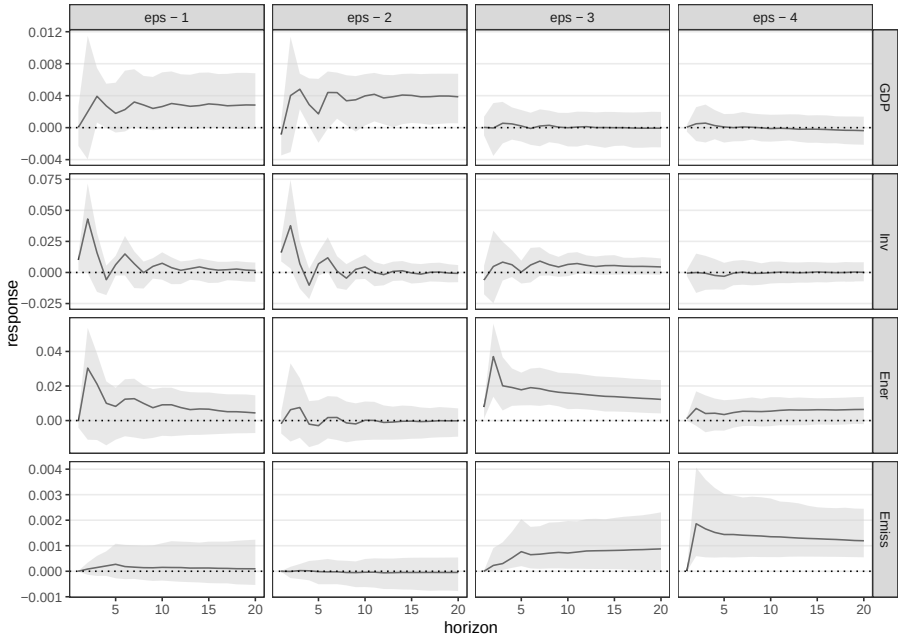


Figure D.2: Plots of the simulated IRFs of VAR_2 obtained with CoP 35.



E R Codes and Data

The R codes and data to replicate the results reported in this article are available at the following repository:

https://www.dropbox.com/sh/953wx8zedosk7bq/AADKHrbthlVCNw30LK4iYnRNa?dl=0