LEM
WORKING PAPER SERIES

A Method for Agent-Based Models Validation

Mattia Guerini °
Alessio Moneta °

° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

A Method for Agent-Based Models Validation

Mattia Guerini[*][a] and Alessio Moneta[†][a]

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

April 11, 2016

Abstract

This paper proposes a new method for empirically validate simulation models that generate artificial time series data comparable with real-world data. The approach is based on comparing structures of vector autoregression models which are estimated from both artificial and real-world data by means of causal search algorithms. This relatively simple procedure is able to tackle both the problem of confronting theoretical simulation models with the data and the problem of comparing different models in terms of their empirical reliability. Moreover the paper provides an application of the validation procedure to the Dosi et al. (2015) macro-model.

Keywords: Models validation; Agent-Based models; Causality; Structural Vector Autoregressions.
JEL classification: C32, C52, E37.

---

[*]*Corresponding Author:* m.guerini@sssup.it
[†]amoneta@sssup.it

Although demonstrative simulation models are useful,

not least at performing *"what if"* exercises of exploration of different models,

policy analysis requires validated, descriptive simulation models.

Marks (2013)

1 Introduction

Economics, as any scientific discipline intended to inform policy, has inevitably addressed questions related to identification and measurement of causes and effects. This paper, by identifying and comparing causal structures, proposes a method which improves the empirical reliability of policy-oriented simulation models.

The foundation of the *Econometric Society* in 1930 paved the way of a rigorous and formal approach to the analysis of causality, which, as Heckman (2000) points out, constituted the major contribution of econometrics.[1] In the post World War II period causal claims were introduced in macroeconomics by means of aggregate, mechanic and dynamic models in which the *ex-ante* use of economic theory was pivotal. Under this approach the causal process used to be partitioned in a deterministic component and a random component. The former was meant to reflect the causal relations dictated by economic theory. The condition for it to be considered "valid" was to have the random component satisfying the standard Gauss-Markov statistical properties. Such a methodology goes under the name of *Cowles Commission* or *Simultaneous Equations Model* (SEM) approach. The most prominent proposers were Haavelmo (1944) and Koopmans (1950).

This approach has been strongly criticized by Lucas (1976) and Sims (1980) on theoretical and methodological grounds respectively: the former insisted on the fact that individuals endowed with rational expectations would have anticipated the policy interventions supported by SEMs and their behaviour would have brought to opposite results with respect to the ones predicted by SEMs; the latter instead stressed the fact that in the *Cowles Commission* approach the distinction between endogenous and exogenous variable was stated ad-hoc, in order to ensure system identifiability.

Taking as starting points the Lucas (1976) and Sims (1980) critiques, Kydland and Prescott

---

[1]As Hoover (2004) has shown, however, causal language has not always been explicit in economics and in the sciences in general. In the first half of the twentieth century, under the influence of Karl Pearson, Ernst Mach and Bertrand Russell, many research scientists endeavoured to eschew causal concepts in order to privilege talk of functional and statistical dependencies (Illari et al., 2011). Causal talk revived in the second half of the last century (Hoover, 2004). See also Granger (1980)

([1982](#)) paved the way toward a new class of models, becoming the founding fathers of the stream of literature that goes under the name of *Real Business Cycle* (RBC) theory and which then evolved in what today is known as the *Dynamic Stochastic General Equilibrium* (DSGE) approach. These types of models are nowadays the most widely used to draw and to evaluate policy claims because they bear the advantage of simultaneously addressing two critical issues about causal structures. On the one hand, under the acceptance of the rational expectation hypothesis, the structure modelled by the RBC/DSGE approach remains invariant under policy intervention because it takes into account the forward-looking behaviour of the economic agents. On the other hand, the theoretical structure has an empirical counterpart in which the distinction between endogenous and exogenous variables is eschewed. The empirical counterpart is represented by a *Structural Vector Autoregressive* (SVAR) model.[2] But the RBC/DSGE approach is not exempt from problematic issues: structural stability is grounded on individual behaviour, but assuming a representative agent which neglects or even denies any form of interaction. Moreover, the identification of the empirical structure in the SVAR model is typically achieved by imposing restrictions derived from the theoretical model, which are therefore not subjected to any severe test. (See Fagiolo and Roventini ([2012](#)) for a detailed criticism on similar issues).

An alternative approach to the problem of representing macroeconomic causal structures, in which it is possible to run reliable policy experiments, is to build a class of models which better reflect the existing economic mechanisms, including the microeconomic interactions. This is the aim of the *Agent-Based Model* (ABM) approach, also known as the *Agent-Based Computational Economics* (ACE) approach, in which the macroeconomic structure is analysed as an emerging property from the interaction between heterogeneous and bounded rational economic actors. This modelling strategy has been applied to economic theory only during the last three decades, but it rapidly gained a significant success and in the very last years has begun to be perceived as a new valuable paradigm, able to provide a viable alternative to the DSGE framework.[3] ABM is a useful and flexible tool for performing rich policy experiments and for evaluating their implications; among the main advantages of the ABM strategy, stands the possibility of analysing endogenously generated booms and busts and also to study the reaction of the

---

[2]See, however, Canova and Sala ([2009](#)) and Fukac and Pagan ([2006](#)) for cautionary notes about the existence of the empirical counterpart of a DSGE model.

[3]The fast success of ABM might be due both to the huge improvements in computational power and recently, to their ability of explaining and revealing intriguing aspects of the world financial and economic crisis; DSGE models instead, has proven to be of few help when facing the crisis. See Trichet ([2010](#)) for clarification on this last point.

economy to different stimuli, also applied in periods of distress and not only around a fictitious locally stable steady state of the economy.

But ABMs pose a serious methodological problem because of their unclear relationship with the empirical evidence. This paper aims to address this issue. The difficulties of the ABM approach, which represent the counterpart of its flexibility, are perceived both in the model-data confrontation and in the comparison of different models investigating the same piece of evidence. The quality of ABMs has been up to now evaluated according to the ex-post ability in reproducing a number of stylized facts even if other validation procedures are available (see Fagiolo et al., 2007). We argue that such an evaluation strategy is not rigorous enough. Indeed the reproduction, no matter how robust, of a set of statistical properties of the data by a model is a quite weak form of validation, since, in general, given a set of statistical dependencies there are possibly many causal structures which may have generated them. Thus models which incorporate different causal structures, on which diverse and even opposite practical policy suggestions can be grounded, may well replicate the same empirical facts.[4]

The present work proposes a procedure to validate a simulation model which proceeds by first estimating both the causal structure incorporated in the model (using the data artificially generated by the model) and the causal structure underlying the real-world data. Secondly, it compares the two inferred causal structures. In this manner the proposed procedure offers a solution to both the issue of confronting an ABM to empirical data and the issue of confronting different simulation models. Indeed causal structures inferred from different simulation data, generated by different models can be compared in the same way. A good matching between the causal structure incorporated in the ABM and the causal structure underlying the real-world data provides a more rigorous empirical support to the policy statements drawn from the ABM, if compared with the support coming from mere replication of statistical evidence. Other validation procedures have been recently proposed based on information criteria by Lamperti (2015) and Barde (2015); other researchers such as Grazzini (2012), Lux (2012), Recchioni et al. (2015), Gilli and Winker (2003) have focused on estimation, or on the analysis of the emergent properties stemming from ABMs (see Grazzini and Richiardi, 2015); there has also been interest for parameter space exploration (see Salle and Yildizoglu, 2014). The flourishing of all these complementary approaches devoted to the solution of such interrelated issues can be seen an indicator of their relevance and a signal of the vitality of the agent-based community.

The paper is organised as follows. Section 2 reviews the different strands of literature on

---

[4]At the root of this underdetermination problem is the fact that statistical relationships are in general symmetric, while this is not necessarily the case for causal relationships.

which our method is built upon; the validation algorithm is presented extensively in Section 3; Section 4 provides a first application of the method to the *"Schumpeter meeting Keynes"* model proposed by Dosi et al. (2015). Section 5 concludes.

2 Background literature

DSGE models are confronted to the data in two ways. The first and traditional approach is through calibration, in which the parameters of the model are chosen from pre-existing microeconomic studies or in order to replicate the statistical properties of aggregate variables (Kydland and Prescott, 1996). The second approach is through the estimation of a VAR model built to represent the empirical counterpart of the DSGE model. Having estimated a VAR, one can identify a SVAR and confront its impulse response functions with the responses to policy shocks derived from the DSGE model. Alternatively, as proposed by Ireland (2004), one can augment the DSGE model with the VAR residuals and estimate a hybrid model via maximum likelihood (for a criticism see Juselius, 2011).

Calibration and replication of statistical properties of data is practised in the ACE community as well. To our knowledge, however, the models by Bianchi et al. (2007) and Bianchi et al. (2008) are the unique medium-scale agent-based macro-models in which parameters are estimated *ex-ante* and calibrated *ex-post* in order to replicate statistical properties of observed data.[5]

Although calibration and replication of statistical properties of data make some step in taking the model to the data, we claim that this is not enough for the reliability of the policy implications derived from the model, since two models with alternative policy implications may well be both calibrated in order to replicate certain statistical properties of observed data. Reliability can be improved only through a validation exercise designed to provide evidence that the modelled data generating mechanism is an adequate representation of the real-data generating mechanism.

We are aware that the economic system has the characteristics of a complex system in which somehow stable macroscopic aggregate properties emerge from intricate connections at the microscopic level. But we further believe that representing into a unique model every single micro-mechanism at work in a complex economy and showing its good match with data at different levels of aggregation is a very difficult task. A reduction in the complexity of the

---

[5]There are, however, several small-scale ACE financial models which are instead calibrated or estimated such as the ones by Alfarano et al. (2005, 2006, 2007).

issue may be necessary, and hence in what follows we will analyse only the relations between macro variables.[6] Our strategy is indeed to focus only on representing causal structures among aggregate variables of the ABM and test whether they significantly differ from the causal structures that can be found in the real world from observed aggregate variables, without further considerations of the micro properties. In other words, we compare a macro-reduced version of the modelled generated mechanism with a macro-reduced version of the real-data generating mechanism.[7]

Our procedure will separately identify the causal structures of the two different data generating processes at their aggregate level, and then will compare the results of the estimations: if the causal structures are similar, then the model is a good characterization of the causal structure of the real-world data generating process and we will consider it as "valid". The identification method is the same for both processes: we will estimate a SVAR model using both observed and simulated aggregate data. This model, being a model with well known properties, provides us enough power and flexibility to compare the explanatory performances of ABM with that of real-world data. But a crucial feature in the SVAR estimation is the identification procedure, which we describe in the next subsections.

2.1 SVAR identification: an open issue

Starting from a multiple time series dataset composed of K variables collected for T periods we can denote by $\mathbf{Y_t} = (Y_{1t}, \ldots, Y_{Kt})'$ the values of these variables at a particular time t. A simple, but useful way of representing the data generating process, is to model the value of each variable Y_{kt} as a linear combination of the previous values of all the variables as well as their contemporaneous values:

$$\mathbf{Y_t} = \mathbf{B}\mathbf{Y_t} + \mathbf{\Gamma_1}\mathbf{Y_{t-1}} + \cdots + \mathbf{\Gamma_p}\mathbf{Y_{t-p}} + \boldsymbol{\varepsilon_t} \tag{1}$$

where the diagonal elements of the matrix $\mathbf{B}$ are set equal to zero by definition and where $\boldsymbol{\varepsilon_t}$ represents a vector of error terms which we will assume to be mutually statistically independent. Therefore the covariance matrix $\mathbf{\Sigma_{\varepsilon}} = \mathbf{E}\left[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'}\right]$ is diagonal. The SVAR form of this model can

---

[6]Cfr. Haldane (2012).

[7]Possible developments of the same method may allow to compare a micro-macro version of the modelled generated mechanism with the real-data generating mechanism, using observations at different levels of aggregation.

also be written as

$$\mathbf{\Gamma_0 Y_t = \Gamma_1 Y_{t-1} + \cdots + \Gamma_p Y_{t-p} + \varepsilon_t} \tag{2}$$

where $\mathbf{\Gamma_0 = I - B}$. The problem with equations (1) and (2) is that they cannot be directly estimated without biases, being the contemporaneous variables endogenous. What is typically done in the literature is to estimate the reduced form VAR model:

$$\begin{aligned} \mathbf{Y_t} &= \mathbf{\Gamma_0^{-1}\Gamma_1 Y_{t-1} + \cdots + \Gamma_0^{-1}\Gamma_p Y_{t-p} + \Gamma_0^{-1}\varepsilon_t} \\ &= \mathbf{A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t}. \end{aligned} \tag{3}$$

where $\mathbf{u_t = \Gamma_0^{-1}\varepsilon_t}$ is a zero-mean white noise process with a covariance matrix $\mathbf{\Sigma_u = E\left[u_t u_t'\right]}$ that in general is not diagonal.

The problem is that, even if the parameters contained into $\mathbf{A_i}$, for $i = 1,\ldots,p$ can be estimated from equation (3) without incurring in any particular issue, their knowledge is not sufficient for the recovery of the structural parameter contained in $\mathbf{B}$ and in $\mathbf{\Gamma_i}$, for $i = 1,\ldots,p$ of equation (1), making impossible the inference of any causal and/or policy claim. To do such claims we need to recover the matrix $\mathbf{\Gamma_0}$ that contains the contemporaneous causal effects. But the problem is that any invertible unit-diagonal matrix might be compatible with the coefficients estimated from the VAR in equation (3).

The problem of finding the appropriate $\mathbf{\Gamma_0}$ (and hence also finding the matrices $\mathbf{\Gamma_1,\ldots,\Gamma_p}$) is called the *identification problem* and it is usually performed by imposing restrictions on the $\mathbf{\Gamma_0}$ matrix using a Cholesky factorization of the estimated covariance matrix $\mathbf{\Sigma_u}$. But this approach should only be employed when the recursive ordering implied by the identification scheme is firmly supported by theoretical consideration. A class of alternative identification procedures derives from the seminal papers by Bernanke (1986) and Blanchard (1989) and imposes zero restrictions which are based on economic considerations about contemporaneous interactions. An alternative identification strategy descends from Shapiro and Watson (1988) and Blanchard and Quah (1989) by assuming that certain economic shocks (e.g. supply shocks) have long-run effects on some variables but do not influence in the long-run the level of other variables, while other shocks (e.g. demand shocks) have only short-run effects on all the variables. Unfortunately these identification strategies are grounded on some level of theoretical apriorism which does not completely solves the critique put forward by Sims (1980).

A relatively recent approach for solving the identification issue of a SVAR model in a more agnostic and data-driven fashion, allowing to avoid as much as possible subjective choices and

theory driven considerations, has been put forward by Swanson and Granger (1997), Bessler and Lee (2002), Demiralp and Hoover (2003), Moneta (2008) and Moneta et al. (2011) and is based on graphical causal models (see Pearl, 2000; Spirtes et al., 2000).

2.2 Graphical causal models and SVAR identification

A *Causal Graph* $\mathcal{G}$ is a model that consists of a set $\mathcal{V}$ of vertices (nodes) and a set $\mathcal{E}$ of edges (links) and might be written concisely as $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$. It is aimed at representing and analysing specific features of the data-generating process underlying the set of observed variables. The vertices of such a graph correspond to random variables and the edges denote causal relationships among them. In what follows we focus on the simple case of *Directed Acyclic Graphs* (DAG) in which all the edges are directed and causal loops are not allowed.

The identification procedure based on graphical causal models consists of three steps: *(i)* estimating the reduced form VAR of equation (3), *(ii)* applying a search algorithm to the estimated residuals $\mathbf{u_t}$ to obtain the matrix $\mathbf{\Gamma_0}$ (cfr. equation 2) and *(iii)* recovering the other matrices $\mathbf{\Gamma_i}$ $(i = 1, \ldots, p)$ of the SVAR model.

The critical part of the procedure is the second step, in which an algorithm is applied in order to uncover the causal dependencies among the residuals $\mathbf{u_t}$. The literature on causal search models has developed a plethora of algorithms which differ among each other for the assumptions on which they are based and the computational properties. Assumptions typically concern the form of the causal structure (e.g. cyclic or acyclic), the presence or exclusion of latent variables (i.e. causally not sufficient or causal sufficient structures), rules of inference (more on that below), and, finally, statistical properties of the residuals (e.g. normality or linearity) which allow the application of specific tests of conditional independence.

The algorithm presented in Appendix A is the PC algorithm originally developed by Spirtes et al. (2000). In this algorithm causal loops are not allowed. Indeed it is assumed that the causal generating mechanism can be modelled by a DAG. In the SVAR framework this amounts to exclude feedbacks in the contemporaneous causal structure, while feedbacks over time are of course conceivable (e.g. X_t causes Y_{t+1} which in turn causes X_{t+2}). The PC algorithm also assumes causal sufficiency, i.e. there is no unmeasured variable which simultaneously affect two or more observed variables. Rules of inference are conditions which permit deriving causal relationships starting from tests of conditional independence. The PC, and similar algorithms

of the same class, hinge on two rules of inference (see Spirtes et al., 2000):

Condition 1. (Causal Markov Condition) *Any variable in the causal graph $\mathcal{G}$ is conditionally independent of its graphical nondescendants (i.e. non-effects) — except its graphical parents — given its graphical parents (i.e. direct causes).*

Condition 2. (Faithfulness Condition) *Let $\mathcal{G}$ be a causal graph and $\mathcal{P}$ be a probability distribution associated with the vertices of $\mathcal{G}$. Then every conditional independence relation true in $\mathcal{P}$ is entailed by the Causal Markov Condition applied to $\mathcal{G}$.*

The PC algorithm, as many other of the class of *constraint-based search* algorithms, needs as input knowledge of the conditional independence relationships among the variables. There are many possibilities of testing conditional independence which in principle are all compatible with the PC algorithm. If the probability distribution underlying the data is Gaussian, zero partial correlation implies conditional independence. Then a typical procedure is to test for Gaussianity and in case this is not rejected, one can test for zero partial correlations. In many statistical packages the default option is to test zero partial correlation through the Fisher-z-transformation, as proposed by (Spirtes et al., 2000). An alternative option, suited for the SVAR framework, is to test zero partial correlations among the VAR residuals through a Wald test that exploits the asymptotic normality of the covariance matrix of the maximum-likelihood estimated VAR residuals (for details see Moneta, 2008). If Gaussianity is rejected or one is not willing to make distributional assumptions, one way to proceed is to rely on nonparametric tests of conditional independence, which, however, present the well-known problem of dimensionality (cfr. Chlass and Moneta, 2010).

The PC algorithms follows this scheme:

  i. Create a complete graph on the variables $(X_1, \ldots, X_k)$;

 ii. Apply tests for conditional independence in order to prune unnecessary edges;

iii. Apply tests for conditional independence in order to direct remaining edges.

There are other algorithms in the literature that, following a similar scheme, allows for feedback loops or the possibility of latent variables (e.g. the CCD or FCI algorithm; cfr. Spirtes et al., 2000).

Usually, the output of the causal search is in general not a unique graph $\mathcal{G}$, but a set of *Markov equivalent* graphs which represent all the possible data generating processes consistent

with the underlying probability $\mathcal{P}$. Hence the information obtained from this approach is generally not sufficient to provide full identification of the SVAR model requiring again a certain level of a priori theoretical knowledge. Moreover, if the distribution of the residuals is non-Gaussian, it is necessary to apply tests of conditional independence which are different from tests of zero partial correlation. However, Moneta et al. (2013) have shown that if the VAR residuals are non-Gaussian, one can exploit higher-order statistics of the data and apply *Independent Component Analysis* (ICA) (see Comon, 1994; Hyvarinen et al., 2001) in order to fully identify the SVAR model.

2.3 Independent component analysis and SVAR identification

Let us recall the fact that VAR disturbances $\mathbf{u_t}$ and structural shocks $\boldsymbol{\varepsilon_t}$ are connected via

$$\mathbf{u_t} = \boldsymbol{\Gamma_0^{-1}} \boldsymbol{\varepsilon_t}. \tag{4}$$

In this framework the VAR residuals are interpreted as generated by a linear combination of non-Gaussian and independent structural shocks via the *mixing matrix* $\boldsymbol{\Gamma_0^{-1}}$. Independent Component Analysis applied to equation (4) allows the estimation of the mixing matrix $\boldsymbol{\Gamma_0^{-1}}$ and the independent components $\boldsymbol{\varepsilon_t}$ by finding linear combinations of $\mathbf{u_t}$ whose mutual statistical dependence is, according to some given measure, minimized. Some points should be noticed: *(i)* while the assumptions of mutual independence of the structural shocks is usually not necessary in a SVAR framework (orthogonality is usually sufficient), such an assumption is necessary to apply ICA; *(ii)* ICA does not require any specific distribution of the residuals $\mathbf{u_t}$ but only requires that they are non-Gaussian (with the possibility of at maximum one Gaussian element); *(iii)* the ICA-based approach for causal search does not require the faithfulness condition; *(iv)* in non-Gaussian settings while conditional independence implies zero partial correlation, the converse does not hold in general.

The application of ICA to the estimated VAR residuals allows identifying the rows of the matrix $\boldsymbol{\Gamma_0}$, but not their order, sign and scale (for details see Hyvarinen et al., 2001). In order to obtain the correct matrix $\boldsymbol{\Gamma_0}$, that is the matrix incorporating the contemporaneous causal structure and such that $\boldsymbol{\Gamma_0} = \mathbf{I} - \mathbf{B}$ in equation (1), we further assume that the VAR residuals can be represented as a *Linear Non-Gaussian Acyclic Model* (LiNGAM) so that the contemporaneous causal structure can be represented as a DAG. On the basis of this assumption, it is possible to apply the causal search algorithm presented in Appendix B (VAR-LiNGAM),

which draws on the original contributions of Shimizu et al. (2006) and Hyvarinen et al. (2010) (for an application to economics see Moneta et al., 2013). The basic idea by which the VAR-LiNGAM algorithm solves the order indeterminacy is that if the underlying causal structure is acyclic, there must be only one row-permutation of the ICA-estimated rows of $\mathbf{\Gamma_0}$ such that all the entries of the main diagonal are different from zero. Hence, the algorithm applies a search procedure to find such a permutation (for details see Appendix B, step C). The scaling indeterminacy is solved by normalizing the elements of the ICA-estimated matrix and rightly row-permuted $\mathbf{\Gamma_0}$, such that the main diagonal is one (and the main diagonal of $\mathbf{B}$ is zero, as dictated by equation (1).

In the literature there are alternative ICA-based search algorithms, which relax the assumption of acyclicity and causal sufficiency: see for example the algorithms proposed by Lacerda et al. (2008) and Hoyer et al. (2008), which respectively allow for feedback loops and for latent variables. However, since this is a first application of this type of framework to validation of simulated model, we decided to keep the analysis as simple as possible so that in future works we might relax assumptions, and understand which are the most critical for validation concerns.

3 The validation method

In this section we describe our validation procedure which is composed of five different steps as shown in figure (1). In the first step we apply some simple transformations that allow the empirical and the artificial data to be directly comparable; in the second step we analyse the emergent properties of the series produced by the simulated model; in the third step we estimate the reduced-form VAR model; in the fourth step we identify the structural form of the model by means of some causal search algorithm; in the last step we compare the two estimated causal structures according to some distance measure.

3.1 Dataset uniformity

Our method starts by selecting, in the model under validation inquiry, K variables of interest $(v_1, \ldots, v_K)$. We then collect a dataset that corresponds to the actual realization of these variables in the real world (we call this dataset RW-data) as well as a dataset for the realizations of M Monte Carlo simulations of the agent-based model (we call this one AB-data). We thus obtain two preliminary datasets $\mathcal{V}_{RW}$ and $\mathcal{V}_{AB}$ which might be of different dimensions. In

Figure 1: The five steps of the validation method.
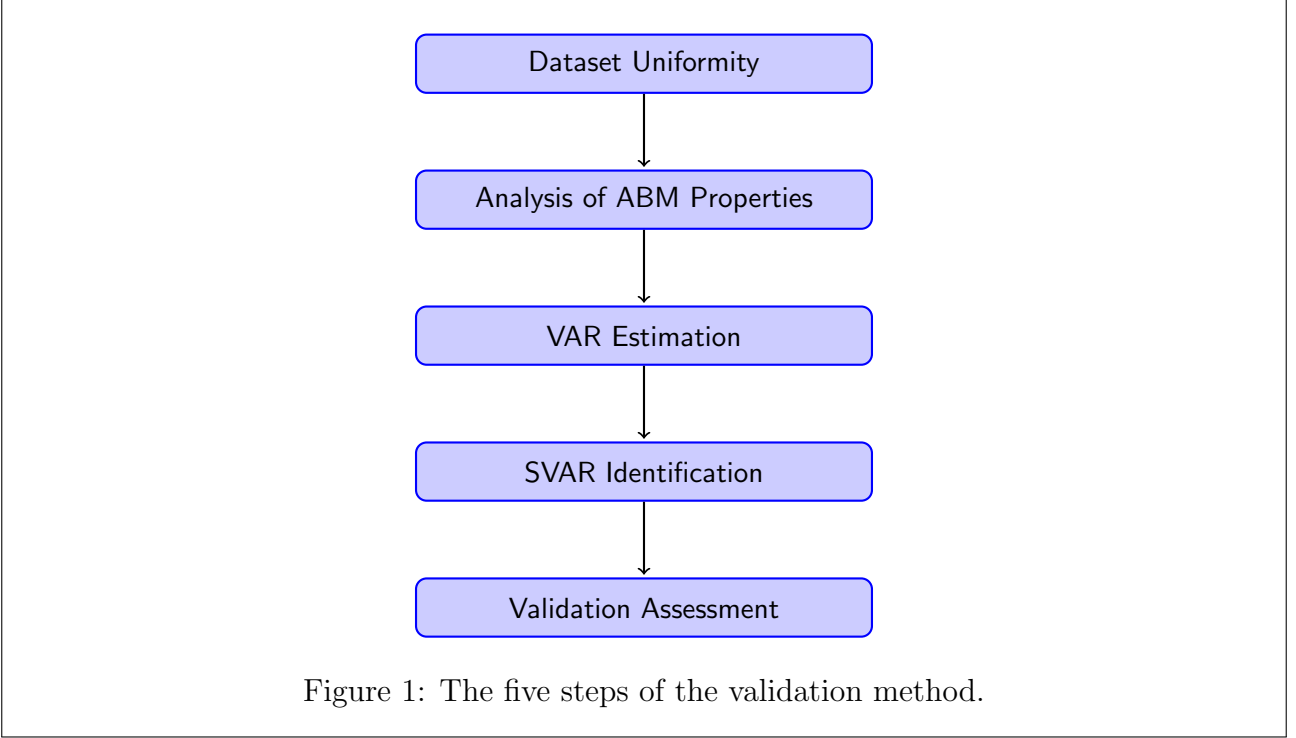
general we will have

$$
\begin{cases}
dim(\mathcal{V}_{RW}) & = 1 \times K \times T_{RW} \\
dim(\mathcal{V}_{AB}) & = M \times K \times T_{AB}
\end{cases}
$$

meaning that for the real world we observe only 1 realization of the the K variables of interest for a period of length T_{RW} while for the simulated data (for which we can possibly have an infinity of observations) we will have M Monte Carlo realizations, of the same K variables, for a period of length T_{AB}; it often holds true that $T_{AB} >> T_{RW}$, so that the two datasets are not perfectly matchable.

The large availability of realizations in the simulated data is in fact an advantage and not an issue, since this allows a pairwise comparison of each run of the Monte Carlo simulation with the unique empirical realization. But the presence of different lengths in the time series might generate issues in two directions: *(i)* using the whole length of the AB-data time series creates the risk of capturing the effects present in the transient period, which does not represent the true dynamic entailed by the model, but is only due to the choice of the initial conditions, *(ii)* lag selection might be affected due to unobserved persistence in some of the modelled variables. Therefore we remove an initial subset of length $T_{AB} - T_{RW}$ from each of the M artificial datasets

(as shown in figure 2) in order to force each pair of datasets to have the same dimensions:

$$dim(\mathcal{V}_{RW}) = dim(\mathcal{V}_{AB}(i)) = 1 \times K \times T_{RW} \quad \text{for } i = 1, \ldots, M.$$
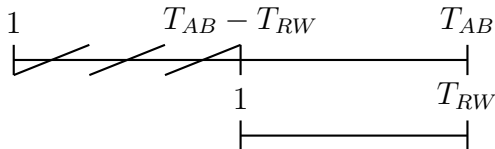

Figure 2: Time window selection. The first periods of the AB-data are canceled with the objective of homogenising the series.

Moreover the order of magnitude of RW-data and AB-data are typically different; this is not perceived by the ABM community as an issue, being the concern of a large number of ABMs the replication of stylized facts (distributions, variations, statistical properties but not levels). But in our approach this might create comparability issue. We will see that in our application it is sufficient to take a logarithmic transformation in order to smooth out this scaling issue, and we speculate that in many applications any monotonic transformation might be applied.

3.2   Analysis of ABM properties

Some considerations about two underlying assumptions are needed. For the model to be a good proxy of the data generating process, we require that it is in a statistical equilibrium state in which the properties of the analysed series are constant. In particular we require that the series, or a transformation of them (e.g. first differences), have distributional properties that are time-independent; secondly we require that the series are ergodic, meaning that the observed time series are a random sample of a multivariate stochastic process.

In a context where the series have been generated by a simulation model, which provides M Monte Carlo realizations, these two assumptions can be tested directly (see Grazzini, 2012). Indeed if we consider all the M time series realizations of a variable of interest k we will collect a matrix with dimensions $M \times T$ containing all the generated data $Y_{k,t}^m$, as represented in figure (3). We call here *ensembles* the column vectors of such a matrix; therefore each ensemble contains the M observations $Y_{k,t}^{(\cdot)}$ in which the time dimension is fixed. We instead define *samples* all the row vectors of the same matrix, each of which contains the T observations $Y_{k,(\cdot)}^m$ in which the

Monte Carlo dimension is fixed. Let's denote by $F_t(Y_k)$ the empirical cumulative distribution function of an ensemble and by $F_m(Y_k)$ the empirical cumulative distribution function of a sample. Testing for statistical equilibrium and for ergodicity reduces to test (via Kolmogorov-Smirnov test) for the following conditions:

$$F_i(Y_k) = F_j(Y_k), \quad \text{for } i,j = 1,\ldots,T \quad i \neq j \tag{5}$$

$$F_i(Y_k) = F_j(Y_k), \quad \text{for } i = 1,\ldots,T \quad j = 1,\ldots,M \tag{6}$$

Therefore we perform two tests as represented in figure (3): we recursively run tests of pairwise equality of distributions and we present the percentage of non-rejection of such tests. Rejecting the test would imply that the distribution under investigation are different from each other. Our two assumptions will result to be supported by the data if we obtain high percentages of non-rejection.

$$Y^k = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,T} \\ y_{2,1} & \cdots & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{M,1} & y_{M,2} & \cdots & y_{M,T} \end{pmatrix} \qquad Y^k = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,T} \\ y_{2,1} & \cdots & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{M,1} & y_{M,2} & \cdots & y_{M,T} \end{pmatrix}$$

Figure 3: The elements of comparison when testing for statistical equilibrium (left) and for ergodicity (right).

3.3 VAR estimation

Following the Box and Jenkins (1970) methodology, the first task when any time series model has to be estimated, is the lag selection: in our case this means choosing the two values p_{RW} and p_{AB}^i, for $i = 1,\ldots,M$ according to some information criterion like the BIC the HQC or the AIC. Two cases might anyway emerge from the data:

1. $p_{RW} - p_{AB}^i = 0$ which means that our estimations based on the artificial dataset and on the real-world dataset are perfectly comparable;

2. $p_{RW} - p_{AB}^i \neq 0$ which means that one of the two dataset presents at least one effect which

is not present in the other; we keep this fact into account when computing the similarity measure.

Once the lag selection has been performed, our procedure estimates the VAR as explicited in equation (3) via OLS and also in his VECM form, via maximum likelihood estimation (see Lutkepohl, 1993) using the Johansen and Juselius (1990) procedure.

3.4 SVAR identification

In this step we extract the vectors of residuals $(u_1, \ldots, u_K)$ from the estimation of the VAR and analyse their statistical properties and their distributions. We test for normality applying the Shapiro-Wilk and the Jarque-Bera statistics. Then according to the outcome of the tests, we select the appropriate causal search algorithm to be adopted for the identification strategy. Two algorithms, the PC (to be adopted for the Gaussian case) and the $VAR\text{-}LiNGAM$ (for the non-Gaussian case) are presented extensively in the appendices A and B. At the end of the identification procedure, we have estimated our structural matrices $\mathbf{\Gamma_i^{RW}}$ for $i = 0, \ldots, p_{RW}$ and $\mathbf{\Gamma_i^{AB,m}}$, for $i = 0, \ldots, p_{AB}$ and for $m = 1, \ldots, M$.

3.5 Validation assessment

The last step consists in the comparison of the causal effects entailed by the $SVAR_{RW}$ and the $SVAR_{AB}$ models. This will tell us how many of the real-world estimated causal effects are captured also by the agent-based model under validation inquiry.

In order to compare the causal effects we will use the similarity measure $\mathbf{\Omega}$, which we construct, as already anticipated, starting from the estimates of the SVAR.

Let's denote $\gamma_{i,jk}^{RW}$ the (j,k) element of $\mathbf{\Gamma_i^{RW}}$ for $i = 0, \ldots, p_{RW}$ and $\gamma_{i,jk}^{AB}$ the (j,k) element of $\mathbf{\Gamma_i^{AB,m}}$ for $i = 0, \ldots, p_{AB}$ and for $m = 1, \ldots, M$. We define $p_{max} = max\{p_{RW}, p_{AB}\}$ and then we set

$$
\begin{cases}
\mathbf{\Gamma_i^{RW}} = \mathbf{0} & \text{for } p_{RW} < i \leq p_{max} \quad \text{if} \quad p_{RW} < p_{max} = p_{AB} \\
\mathbf{\Gamma_i^{AB}} = \mathbf{0} & \text{for } p_{AB} < i \leq p_{max} \quad \text{if} \quad p_{AB} < p_{max} = p_{RW}
\end{cases}
$$

This allows us to penalize the value obtained by the similarity measure for the fact that the causal effects are completely mismatched after a certain lag. Then we build the indicator

function:

$$\omega_{i,jk} = \begin{cases} 1 & \text{if} \quad sign(\gamma_{i,jk}^{RW}) = sign(\gamma_{i,jk}^{AB}) \\ 0 & \text{if} \quad sign(\gamma_{i,jk}^{RW}) \neq sign(\gamma_{i,jk}^{AB}) \end{cases} \tag{7}$$

where $i = 0, \ldots, p_{max}$ is the index for the ith-order matrix containing the causal effects, while j and k are the row and column indexes of these matrices. The similarity measure is then defined as

$$\boldsymbol{\Omega} = \frac{\left(\sum_{i=1}^{p_{max}} \sum_{j=1}^{K} \sum_{k=1}^{K} \omega_{i,jk} \right)}{K^2 p_{max}}. \tag{8}$$

Our similarity measure is bounded between $[0, 1]$ allowing us to have an easily interpretable index that represents the ability of the agent-based model to recover, at the aggregate macro-level, the same causal relationsips estimated in the RW-data.

4 Application to the "Schumpeter Meeting Keynes" model

The idea of building a simulation macroeconomic laboratory performing policy exercises dates back to Lucas (1976) and Kydland and Prescott (1982) but one problem of this approach has always been the external validity. Our methodology can be interpreted as a test for external validity for any simulation model and it might be of particular interest for researchers engaged in practical policy matters and policy makers who should take decisions according to models which must be reliable and valid. We already argued that a policy reliable model is one that is able not only to replicate a list of stylized facts, but also to represent the real-world causal structure as much accurately as possible. We want to test here the validity of the agent-based laboratory by Dosi et al. (2015), a model that builds upon a series of previous papers (see Dosi et al., 2013, 2010) and that has attracted great attention in the recent literature.

The Dosi et al. (2015) model aims at investigating the implications of demand and supply public policies in a model that *"bridges Keynesian theories of demand-generation and Schumpeterian theories of technology-fuelled economic growth"*. The model by itself is able to reproduce a long list of stylized facts and in particular is able to reproduce a cross correlation table close to the one usually computed with the US observed data.

4.1 The dataset

Since the model under validation inquiry is dedicated to the analysis of the real side of an economic system and to the analysis of fiscal and monetary policies, the $K = 6$ variables of major interest that we will consider in our system of equation are: aggregate consumption (C), gross private investments (I), unemployment rate (U), gross domestic product (Y), current price index (P) and effective federal funds rate (R). The RW-data refer to the United States and are collected from the Federal Reserve Economic Database (FRED); we decided to cover the period from January 1959 to April 2014 with a quarterly frequency, implying a time series length $T = 222$; this is a typical selection when analysing US business cycle. All the variables are plotted in figure 4.

Figure 4: *Left columns:* time series of RW-data. *Right columns:* time series of a typical AB-data.
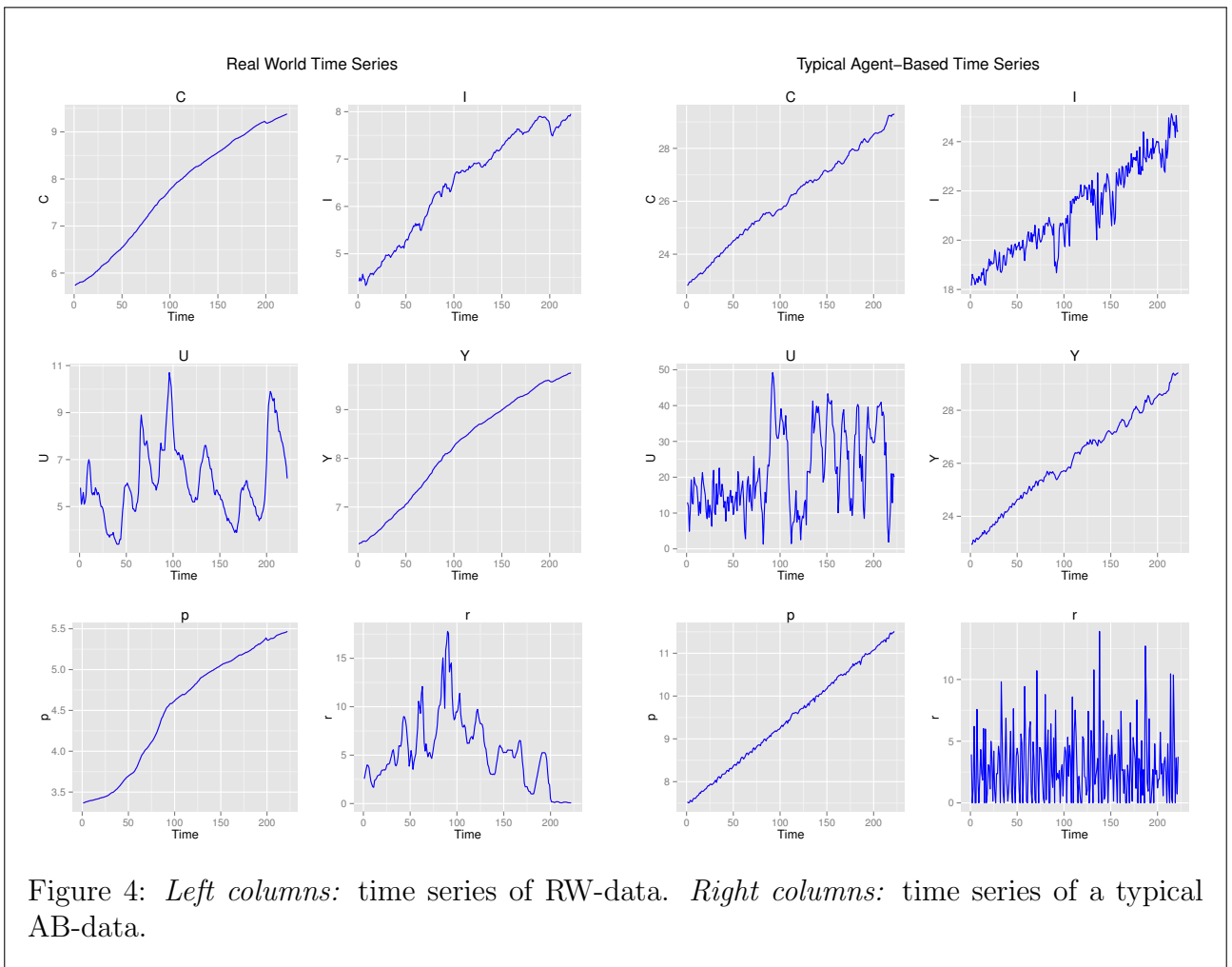
To fulfil the dataset uniformity requirement for the AB-data, we collect the last T time observations, getting rid of possible transients and we consider $M = 100$ Monte Carlo simulations;

each of them will be pairwise compared to the unique realization of the RW-data. Finally we take logs of the C, I, Y, P variables and we uniform U and R by expressing them in percentage terms.[8]

We then check whether the assumptions we require for estimation of our AB model are too stringent or they are supported by the data; we perform the statistical equilibrium and the ergodicity tests as described in the section 3.2 and in figure (3); this kind of test is in line also with the analysis proposed by Grazzini (2012).

	Equilibrium	Ergodicity
C	0.9538	0.9479
I	0.9634	0.9564
U	0.9608	0.9396
Y	0.9532	0.9513
P	0.9560	0.9055
R	0.9609	0.9716

Table 1: Percentages of non-rejection of statistical equilibrium and ergodicity.

Table (1) presents the percentage of non-rejection of the Kolmogorov-Smirnov test of each pairwise comparison, for each stationarised series. The results come from $\frac{T \times (T-1)}{2} = 24310$ and $T \times M = 22100$ pairwise comparisons for the statistical equilibrium and for the ergodicity tests respectively. For all the series we have values higher than 90% and this allows us to conclude that the assumptions about the model having reached a statistical equilibrium and producing ergodic series are reasonable.

4.2 Estimation and validation results

The augmented Dickey-Fuller test does not reject the null hypotheses of unit root in all the real-world time series. For AB-data, the evidence for ubiquity of unit root is weaker since for I, U and R we can reject at the 5% level the presence of unit root (see table 2). This does not create any difficulty to our causal search procedure and it is only a stylized fact not replicated by the model.

We then estimate the model as a vector error correction model (VECM) with cointegrating relationships, without taking first difference of any variable, using the Johansen and Juselius (1990) procedure, which is based on a maximum-likelihood estimation with normal errors, but is robust also to non-Gaussian disturbances. For sake of completeness we also check the robustness

---

[8]Three Monte Carlo simulations, $m = \{55, 61, 89\}$, are excluded from the original dataset because in (at least) one period of the model, investment goes to 0, implying we cannot take logarithm.

(a) RW-data

Variable	ADF p-value for levels	ADF p-value for 1st-differences	Critical level
C	0.99	0.02	0.05
I	0.89	0.01	0.05
U	0.06	0.01	0.05
Y	0.99	0.01	0.05
p	0.92	0.19	0.05
r	0.22	0.01	0.05

(b) AB-data

Variable	ADF p-value for levels	ADF p-value for 1st-differences	Critical level
C	0.43	0.01	0.05
I	0.01	0.01	0.05
U	0.01	0.01	0.05
Y	0.19	0.01	0.05
p	0.63	0.01	0.05
r	0.01	0.01	0.05

Table 2: Augmented Dickey-Fuller Test.

of the results by estimating the VAR in level via OLS. We select the number of lags according to the Bayes-Schwarz Information Criterion (BIC) and the number of cointegrating relationships following the Johansen procedure. For the real-world dataset the suggestion is that of using 3 lags and 2 cointegrating relationships while in the artificial datasets (typically) we should use 3 lags and 3 cointegrating relationships. This implies that we do not have any comparability issue for what concerns the number of lags. With respect to the cointegrating relations, it is again a stylized fact not matched by the model which does not create any estimation issue, since we are interested in structural form of the model and cointegration is only a reduced form property.

The empirical distributions of the VAR residuals $(u_{C,t}, \ldots, u_{R,t})$ are represented in figure (5) both for RW-data and for a typical Monte Carlo realization of the AB-data simulation; moreover table (3) collects the results of the Shapiro-Wilk and the Jarque-Bera tests for normality; for all the variables, the residuals from the real-world data and all but one residual (the unemployment) from the artificial data, the tests rejects the null hypothesis of normality.

We conclude that the residuals $\mathbf{u_t}$ are non-Gaussian and this result leads us toward the identification of the SVAR via the LiNGAM algorithm.[9]

After having completed the estimation, we compute the similarity measure as defined in

---

[9]The VAR-LiNGAM algorithm is consistent even if one variable is normally distributed and therefore even the unemployment residuals quasi-Gaussianity does not add complications to our identification and structural estimation procedure.
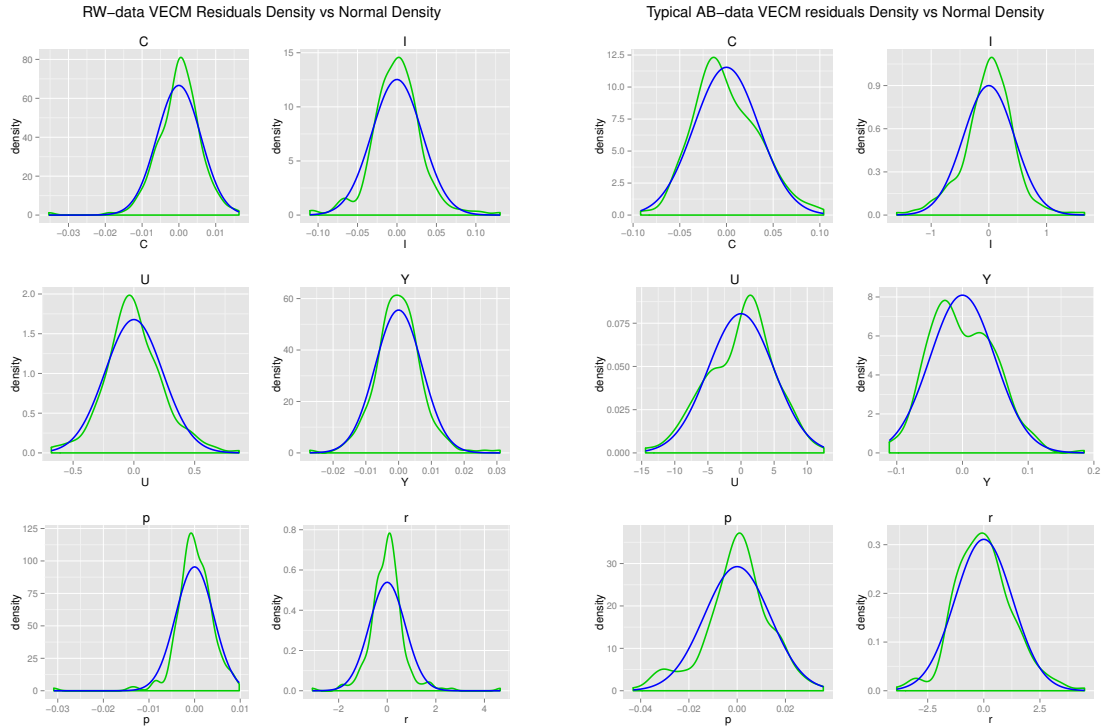
Figure 5: *Left columns:* RW-data VECM residuals distribution (green) and normal distribution (blue). *Right columns:* typical AB-data VECM residuals distribution (green) and normal distribution (blue).

equation (8). The results suggest that when we estimate the system using a OLS-VAR strategy the *Schumpeter meeting Keynes* model is able to reproduce, on a Monte Carlo average, the 78.92% of the causal relations entailed in the real-world dataset (the similarity drops to 64.9% after accounting only for bootstrapped significant parameters); on the other side, if in the first step we estimate a VECM by means of maximum likelihood, the similarity measure marks 73.85% (raising to 79.89% when considering only bootstrapped significant parameters). The results are reported also in table (4), containing not only the means but also standard deviations across Monte Carlo. Being the dispersion index quite low, we can conclude that neither very negative nor very positive outliers are present. Therefore a large fraction of simulations entail the same bulk of causal relations.

(a) RW-data

	Shapiro-Wilk Test	Shapiro-Wilk p-value	Jarque-Bera Test	Jarque-Bera p-value
C	0.95	0.00	271.48	0.00
I	0.96	0.00	54.96	0.00
U	0.98	0.02	12.16	0.00
Y	0.97	0.00	61.16	0.00
P	0.88	0.00	1710.58	0.00
R	0.89	0.00	698.46	0.00

(b) AB-data

	Shapiro-Wilk Test	Shapiro-Wilk p-value	Jarque-Bera Test	Jarque-Bera p-value
C	0.99	0.03	6.01	0.05
I	0.97	0.00	29.64	0.00
U	0.99	0.08	3.23	0.20
Y	0.98	0.02	4.72	0.09
P	0.98	0.00	10.98	0.00
R	0.99	0.05	7.30	0.03

Table 3: Normality test on the VECM residuals.

Estimation Method	μ	σ
VAR-OLS (all parameters)	0.7892	0.0517
VECM-ML (all parameters)	0.7385	0.0628
VAR-OLS (significant parameters)	0.6490	0.1030
VECM-ML (significant parameters)	0.7989	0.0689

Table 4: Mean and standard deviation of the similarity measure.

5   Conclusions

In this paper we have presented a new method for validating policy-oriented Agent-Based macroeconomic models able to generate artificial time series comparable with the aggregate time series computed by statistical offices, central banks and institutional organizations. The approach is based on comparing Structural Vector Autoregressive models which are estimated from both artificial and real-world data by means of causal search algorithms. In the paper we also have presented a first application of our method to the Dosi et al. (2015) model. We have calculated that by using the simulated data and according to the proposed similarity measure, the model is able to resemble between 65% and 80% of the causal relations entailed by a SVAR estimated on real-world data. We guess that this is a positive result for the *Schumpeter meeting Keynes* model but in order to reinforce this claim, we should need to compare this results with those coming from other models. In our opinion, this paper sets a new benchmark on which members of the agent-based community might build upon. Convinced about the fact that the validation issue cannot be settled out in an ultimate manner, other approaches for model validity can emerge and might bring complementary evidence to ours. Indeed a possible strategy, for researchers willing to bring their agent-based models to the audience of policymakers, is that of applying a plurality of methods.

References

ALFARANO, SIMONE, THOMAS LUX, and FRIEDRICH WAGNER (2005). "Estimation of agent-based models: the case of an asymmetric herding model". In: *Computational Economics* 26.1.

ALFARANO, SIMONE, THOMAS LUX, and FRIEDRICH WAGNER (2006). "Estimation of a simple agent-based model of financial markets: An application to australian stock and foreign exchange data". In: *Physica A: Statistical Mechanics and its Applications* 370.1, pp. 38–42.

ALFARANO, SIMONE, THOMAS LUX, and FRIEDRICH WAGNER (2007). "Empirical validation of stochastic models of interacting agents". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 55.2, pp. 183–187.

BARDE, SYLVAIN (2015). "Direct calibration and comparison of agent-based herding models of financial markets". In: *Studies in Economics, University of Kent.*

BERNANKE, BEN S. (1986). "Alternative Explanations of the Money-Income Correlation". In: *NBER Working Paper Series* 1842.

BESSLER, DAVID A. and SEONGPYO LEE (2002). "Money and prices: U.S. Data 1869-1914 (A study with directed graphs)". In: *Empirical Economics* 27, pp. 427–446.

BIANCHI, CARLO, PASQUALE CIRILLO, MAURO GALLEGATI, and PIETRO VAGLIASINDI (2007). "Validating and Calibrating Agent-Based Models: A Case Study". In: *Computational Economics* 30, pp. 245–264.

BIANCHI, CARLO, PASQUALE CIRILLO, MAURO GALLEGATI, and PIETRO VAGLIASINDI (2008). "Validation in agent-based models: An investigation on the CATS model". In: *Journal of Economic Behavior & Organization* 67, pp. 947–964.

BLANCHARD, OLIVIER J. (1989). "A Traditional Interpretation of Macroeconomic Fluctuations". In: *The American Economic Review* 79.5, pp. 1146–1164.

BLANCHARD, OLIVIER J. and DANNY QUAH (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances". In: *The American Economic Review* 79.4, pp. 655–673.

BOX, GEORGE and GWILYM JENKINS (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day.

Canova, Fabio and Luca Sala (2009). "Back to Square One: Identification Issues in DSGE Models". In: *Journal of Monetary Economics* 56, pp. 431–449.

Chlass, Nadine and Alessio Moneta (2010). "Can Graphical Causal Inference Be Extended to Nonlinear Settings? An Assessment of Conditional Independence Tests". In: *EPSA Epistemology and Methodology of Science*. Ed. by M. Suarez M. Dorato M. Redei. Springer Verlag, pp. 63–72.

Comon, Pierre (1994). "Independent component analysis, a new concept?" In: *Signal Processing* 36.3, pp. 287–314.

Demiralp, Selva and Kevin D. Hoover (2003). "Searching for the Causal Structure of a Vector Autoregression". In: *Oxford Bullettin of Economics and Statistics* 65, pp. 745–767.

Dosi, Giovanni, Giogio Fagiolo, Mauro Napoletano, and Andrea Roventini (2013). "Income distribution, credit and fiscal policies in an agent-based Keynesian model". In: *Journal of Economic Dynamics and Control* 37.8, pp. 1598–1625.

Dosi, Giovanni, Giorgio Fagiolo, Mauro Napoletano, Andrea Roventini, and Tania Treibich (2015). "Fiscal and Monetary Policies in Complex Evolving Economies". In: *Journal of Economic Dynamics and Control* 52, pp. 166–189.

Dosi, Giovanni, Giorgio Fagiolo, and Andrea Roventini (2010). "Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles". In: *Journal of Economic Dynamics and Control* 34.9, pp. 1748–1767.

Fagiolo, Giorgio, Alessio Moneta, and Paul Windrum (2007). "A Critical Guide to Empirical Validation of Agent-Based Models in Economics and Finance: Methodologies, Procedures and Open Problems". In: *Computational Economics* 30.3, pp. 195–226.

Fagiolo, Giorgio and Andrea Roventini (2012). "Macroeconomic policy in DSGE and agent-based models". In: *Revue de l'OFCE* 124, pp. 67–116.

Fukac, Martin and Adrian Pagan (2006). "Issues in Adopting DSGE Models for Use in the Policy Process". In: *CAMA Working Paper Series* 10.

Gilli, Manfred and Peter Winker (2003). "Global optimization heuristic for estimating agent based models". In: *Computational Statistics & Data Analysis* 42.3, pp. 299–312.

GRANGER, CLIVE W. J. (1980). "Testing for causality: A personal viewpoint". In: *Journal of Economic Dynamics and Control* 2, pp. 329–352.

GRAZZINI, JAKOB (2012). "Analysis of the emergent properties: Stationarity and ergodicity". In: *Journal of Artificial Societies and Simulation* 15.2.

GRAZZINI, JAKOB and MATTEO RICHIARDI (2015). "Estimation of ergodic agent-based models by simulated minimum distance." In: *Journal of Economic Dynamics and Control* 51, pp. 148–165.

HAAVELMO, TRYGVE (1944). "The Probability Approach in Econometrics". In: *Econometrica* 12.Supplement.

HALDANE, ANDREW (2012). *The dog and the frisbee.* Tech. rep. Speech at the Federal Reserve Bank of Kansas City 366th economic policy symposium.

HECKMAN, JAMES (2000). "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective". In: *The Quarterly Journal of Economics* 115.1, pp. 45–97.

HOOVER, KEVIN D. (2004). "Lost Causes". In: *Journal of the History of Economic Thought* 26.2, pp. 149–164.

HOYER, PATRIK O., SHOHEI SHIMIZU, ANTTI J. KERMINEN, and MARKUS PALVIAINEN (2008). "Estimation of Causal Effects Using Linear Non-Gaussian Causal Models with Hidden Variables". In: *International Journal of Approximate Reasoning* 49, pp. 362–278.

HYVARINEN, AAPO, JUHA KARHUNEN, and ERKKI OJA (2001). *Indepentent Component Analysis.* Wiley.

HYVARINEN, AAPO, KUN ZHANG, SHOHEI SHIMIZU, and PATRICK O. HOYER (2010). "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity". In: *Journal of Machine Learning Research* 11, pp. 1709–1731.

ILLARI, PHYLLIS MCKAY, FEDERICA RUSSO, and JON WILLIAMSON (2011). "Why look at Causality in the Sciences? A manifesto". In: *Causality in the Sciences.* Oxford University Press.

IRELAND, PETER N (2004). "A method for taking models to the data". In: *Journal of Economic Dynamics and Control* 28.6, pp. 1205–1226.

JOHANSEN, SOREN and KATERINA JUSELIUS (1990). "Maximum Likelihood Estimation and Inference on Cointegration. With Application to the Demand for Money". In: *Oxford Bullettin of Economics and Statistics* 52, pp. 169–210.

JUSELIUS, KATARINA (2011). "On the role of theory and evidence in macroeconomics". In: *Elgar Companion to Recent Economic Methodology*. Edward Elgar, pp. 404–436.

KOOPMANS, T. C. (1950). "Statistical Inference in Dynamic Economic Models". In: *Cowles Commission Monograph. John Wiley & Sons* 10.

KYDLAND, FINN E. and EDWARD C. PRESCOTT (1982). "Time to Build and Aggregate Fluctuations". In: *Econometrica* 50.6, pp. 1345–1370.

KYDLAND, FINN E and EDWARD C PRESCOTT (1996). "The computational experiment: an econometric tool". In: *The Journal of Economic Perspectives*, pp. 69–85.

LACERDA, GUSTAVO, PETER SPIRTES, PATRICK O. HOYER, and JOSEPH RAMSEY (2008). "Discovering Cyclic Causal Models by Independent Components Analysis". In: *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*.

LAMPERTI, FRANCESCO (2015). "An Information Theoretic Criterion for Empirical Validation of Time Series Models". In: *LEM Working Paper*.

LUCAS, ROBERT (1976). "Econometric Policy Evaluation: A Critique". In: *Carnegie-Rochester Conference Series on Public Policy* 1.1, pp. 19–46.

LUTKEPOHL, HELMUT (1993). *Introduction to Multiple Time Serie Analysis*. Springer-Verlag.

LUX, THOMAS (2012). "Estimation of an agent-based model of investor sentiment formation in financial markets". In: *Journal of Economic Dynamics and Control* 36.8, pp. 1284–1302.

MARKS, ROBERT (2013). "Validation and Model Selection: three similarity measures compared". In: *Complexity Economics* 2.1.

MONETA, ALESSIO (2008). "Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis". In: *Empirical Economics* 35.2, pp. 275–300.

MONETA, ALESSIO, NADINE CHLASS, DORIS ENTNER, and PATRIK HOYER (2011). "Causal Search in Structural Vector Autoregressive Models". In: *Journal of Machine Learning Research* 12, pp. 95–118.

MONETA, ALESSIO, DORIS ENTNER, PATRICK HOYER, and ALEX COAD (2013). "Causal Inference by Independent Component Analysis: Theory and Applications". In: *Oxford Bullettin of Economics and Statistics* 75.5, pp. 705–730.

PEARL, JUDEA (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

RECCHIONI, MARIA CRISTINA, GABRIELE TEDESCHI, and MAURO GALLEGATI (2015). "A calibration procedure for analyzing stock price dynamics in an agent-based framework". In: *Journal of Economic Dynamics and Control* 60.1, pp. 1–25.

SALLE, ISABELLE and MURAT YILDIZOGLU (2014). "Efficient Sampling and Meta-Modeling for Computational Economic Models". In: *Computational Economics* 44.4, pp. 507–536.

SHAPIRO, MATTHEW D. and MARK W. WATSON (1988). "Sources of Business Cycle Fluctuations". In: *Cowles Foundation Discussion Papers* 870.

SHIMIZU, SHOHEI, PATRIK O. HOYER, AAPO HYVARINEN, and ANTTI J. KERMINEN (2006). "A Linear Non-Gaussian Acyclic Model for Causal Discovery". In: *Journal of Machine Learning Research* 7, pp. 2003–2030.

SIMS, CHRISTOPHER (1980). "Macroeconomics and Reality". In: *Econometrica* 48.1, pp. 1–48.

SPIRTES, PETER, CLARK GLYMUR, and RICHARD SCHEINES (2000). *Causation, Prediction, and Search*. MIT Press.

SWANSON, NORMAN R. and CLIVE W. J. GRANGER (1997). "Impulse response function based on a causal approach to residual orthogonalization in vector autoregressions". In: *Journal of the American Statistical Association* 92.437, pp. 357–367.

TRICHET, JEAN CLAUDE (2010). *Reflections on the nature of monetary policy non-standard measures and finance theory*. Tech. rep.

Appendix A - PC Algorithm

A. Connect everything

Form the complete undirected graph $\mathcal{G}$ on the vertex set $(u_{1t}, \dots, u_{Kt})$ so that each vertex is connected to any other vertex by an undirected edge.

B. Cut some edges

$n = 0$

REPEAT:

    REPEAT:

        select an ordered pair of variables u_{ht} and u_{it} that are adjacent in $\mathcal{G}$ such that the number

        of variables adjacent to u_{ht} is equal or greater than $n+1$. Select a set $\mathcal{S}$ of n variables

        adjacent to u_{ht} suchthat $u_{it} \notin S$. If $u_{ht} \perp u_{it} | \mathcal{S}$ delete edge $u_{ht} - u_{it}$ from $\mathcal{G}$.

    UNTIL all ordered pairs of adjacent variables u_{ht} and u_{it} such that the number of variables

    adjacent to u_{ht} is equal or greater than $n+1$ and all sets $\mathcal{S}$ of n variables adjacent to u_{ht}

    such that $u_{it} \notin \mathcal{S}$ have been checked to see if $u_{ht} \perp u_{it} | \mathcal{S}$;

    $n = n + 1$;

UNTIL for each ordered pair of adjacent variables u_{ht}, u_{it}, the number of adjacent variables to u_{ht} is less than $n+1$.

C. Build colliders

For each triple of vertices u_{ht}, u_{it}, u_{jt} such that the pair u_{ht}, u_{it} and the pair u_{it}, u_{jt} are each adjacent in $\mathcal{G}$ but the pair u_{ht}, u_{jt} is not adjacent in $\mathcal{G}$, orient $u_{ht} - u_{it} - u_{jt}$ as $u_{ht} \rightarrow u_{it} \leftarrow u_{jt}$ if and only if u_{it} does not belong to any set of variables $\mathcal{S}$ such that $u_{ht} \perp u_{jt} | \mathcal{S}$.

D. Direct some other edges

REPEAT:

    if $u_{at} \rightarrow u_{bt}$, u_{bt} and u_{ct} are adjacent, u_{at} and u_{ct} are not adjacent and u_{bt} belongs to every set

    $\mathcal{S}$ such that $u_{at} \perp u_{ct} | \mathcal{S}$, then orient $u_{bt} - u_{ct}$ as $u_{bt} \rightarrow u_{ct}$; if there is a directed path from

    u_{at} to u_{bt} and an edge between u_{at} and u_{bt}, then orient $u_{at} - u_{bt}$ as $u_{at} \rightarrow u_{bt}$;

UNTIL no more edges can be oriented.

Appendix B - VAR-LiNGAM Algorithm

A. Estimate the reduced form VAR model of equation (3) obtaining estimates $\hat{\mathbf{A}}_{\mathbf{i}}$ of the matrices $\mathbf{A}_{\mathbf{i}}$, $\forall\, \mathbf{i} = \mathbf{1}, \ldots, \mathbf{p}$. Denote by $\hat{\mathbf{U}}$ the $K \times T$ matrix of the corresponding estimated VAR error terms, that is each column of $\hat{\mathbf{U}}$ is $\hat{\mathbf{u}}_{\mathbf{t}} \equiv (\hat{u}_{1t}, \ldots, \hat{u}_{Kt})'$, $\forall\, t = 1, \ldots, T$. Check whether the u_{it} (for all rows i) indeed are non-Gaussian, and proceed only if this is so.

B. Use *FastICA* or any other suitable ICA algorithm (Hyvarinen et al., 2001) to obtain a decomposition $\hat{\mathbf{U}} = \mathbf{P}\mathbf{E}$ where $\mathbf{P}$ is $K \times K$ and $\mathbf{E}$ is $K \times T$, such that the rows of $\mathbf{E}$ are the estimated independent components of $\hat{\mathbf{U}}$. Then validate non-Gaussianity and (at least approximate) statistical independence of the components before proceeding.

C. Let $\tilde{\tilde{\mathbf{\Gamma}}}_{\mathbf{0}} = \mathbf{P}^{-1}$. Find $\tilde{\mathbf{\Gamma}}_{\mathbf{0}}$, the row-permutated version of $\tilde{\tilde{\mathbf{\Gamma}}}_{\mathbf{0}}$ which minimizes $\sum_i \frac{1}{|\tilde{\mathbf{\Gamma}}_{\mathbf{0},\mathbf{ii}}|}$ with respect to the permutation. Note that this is a linear matching problem which can be easily solved even for high K (Shimizu et al., 2006).

D. Divide each row of $\tilde{\mathbf{\Gamma}}_{\mathbf{0}}$ by its diagonal element, to obtain a matrix $\hat{\mathbf{\Gamma}}_{\mathbf{0}}$ with all ones on the diagonal.

E. Let $\tilde{\mathbf{B}} = \mathbf{I} - \hat{\mathbf{\Gamma}}_{\mathbf{0}}$.

F. Find the permutation matrix $\mathbf{Z}$ which makes $\mathbf{Z}\tilde{\mathbf{B}}\mathbf{Z}^{\mathbf{T}}$ as close as possible to lower triangular. This can be formalized as minimizing the sum of squares of the permuted upper-triangular elements, and minimized using a heuristic procedure (Shimizu et al., 2006). Set the upper-triangular elements to zero, and permute back to obtain $\hat{\mathbf{B}}$ which now contains the acyclic contemporaneous structure. (Note that it is useful to check that $\mathbf{Z}\tilde{\mathbf{B}}\mathbf{Z}^{\mathbf{T}}$ indeed is close to strictly lower-triangular).

G. $\hat{\mathbf{B}}$ now contains $K(K-1)/2$ non-zero elements, some of which may be very small (and statistically insignificant). For improved interpretation and visualization, it may be desired to prune out (set to zero) small elements at this stage, for instance using a bootstrap approach (Shimizu et al., 2006).

H. Finally, calculate estimates of $\mathbf{\Gamma}_{\mathbf{i}}$, $\forall\, \mathbf{i} = \mathbf{1}, \ldots, \mathbf{p}$ for lagged effects using $\mathbf{\Gamma}_{\mathbf{i}} = (\mathbf{I} - \hat{\mathbf{B}})\hat{\mathbf{A}}_{\mathbf{i}}$.