LABOR
LABORATORIO RICCARDO REVELLI
CENTRE FOR EMPLOYMENT STUDIES

Working Paper no. **130**

Consistent Estimation of Agent-Based Models by Simulated Minimum Distance

Jakob Grazzini
Catholic University of Milan, ITEMQ

Matteo G. Richiardi
University of Turin, and LABORatorio R. Revelli, Turin

March, 2013

Laboratorio R. Revelli, *Collegio Carlo Alberto*
Via Real Collegio, 30 - 10024 Moncalieri (TO)
Tel. +39 011 670.50.60 - Fax +39 011 670.50.61
www.laboratoriorevelli.it - labor@laboratoriorevelli.it

Consistent Estimation of Agent-Based Models by Simulated Minimum Distance

Jakob Grazzini*[a] and Matteo Richiardi †[b,c]

[a]Catholic University of Milan, ITEMQ, via Necchi 5, 20123, Milano, Italy.

[b]University of Turin, Department of Economics and Statistics, lungo Dora Siena 100 A, 10153 Torino, Italy.

[c]Collegio Carlo Alberto and LABORatorio Revelli, via Real Collegio 30, 10024 Moncalieri, Torino, Italy

March 13, 2013

Abstract

Agent-based (AB) models are considered a promising tool for macroeconomic analysis. However, until estimation of AB models become a common practice, they will not get to the center stage of macroeconomics. Two difficulties arise in the estimation of AB models: (i) the criterion function has no simple analytical expression, and (ii) the aggregate properties of the model cannot be analytically understood. The first one calls for simulation-based estimation techniques; the second requires additional statistical testing in order to ensure that the simulated quantities are consistent estimators of the theoretical quantities. The possibly high number of parameters involved and the non-linearities in the theoretical quantities used for estimation add to the complexity of the problem. As these difficulties are also shared, though to a different extent, by DSGE models, we first look at the lessons that can be learned from this literature. We identify simulated minimum distance (SMD) as a practical approach to estimation of AB models, and we discuss the conditions which ensure consistency of SMD estimators in AB models.

Keywords: Consistent Estimation, Method of Simulated Moments, Agent-based Models

JEL codes: C15 (Statistical Simulation Methods: General), C63 (Computational Techniques; Simulation Modeling)

*Email: jakob.grazzini@unicatt.it

†Email: matteo.richiardi@unito.it

1 Introduction

Agent-based (AB) models are sometimes considered as a candidate to replace or at least complement current dynamic stochastic general equilibrium (DSGE) as the standard tool for macroeconomic analysis. As Jean-Claude Trichet, at the time President of the European Central Bank, famously stated (Trichet, 2010):

> When the crisis came, the serious limitations of existing economic and financial models immediately became apparent. [...] Macro models failed to predict the crisis and seemed incapable of explaining what was happening to the economy in a convincing manner. As a policy-maker during the crisis, I found the available models of limited help. In fact, I would go further: in the face of the crisis, we felt abandoned by conventional tools.

He went further to describe what he thinks are promising directions of research:

> We need to deal better with heterogeneity across agents and the interaction among those heterogeneous agents. We need to entertain alternative motivations for economic choices. [...] Agent-based modelling dispenses with the optimisation assumption and allows for more complex interactions between agents. Such approaches are worthy of our attention.

However, a common critique addressed towards AB models is that they often remain at a theoretical level, and lack a sound empirical grounding (Gallegati and Richiardi, 2009). When present, this is often limited to some *ad-hoc* calibration of the relevant parameters. Estimation, though, is crucial for the empirical validation of a model, for comparing a model with other available models, and for policy analysis. This resembles the state of the art in dynamic stochastic general equilibrium (DSGE) modeling a few years ago, which has now moved forward toward more formal estimation. Unless AB models will become more amenable to estimation, they will never get to the center stage of research in macroeconomics, and policy makers will not rely on them for guidance.

The main difficulty in taking AB models to the data originates from their very nature: because AB models are composed by many heterogeneous agents interacting together and with the environment, aggregation is not a simple matter and it must be performed numerically, by counting

and summing up individual quantities. Therefore, the aggregate properties of an AB model remain hidden in the complexity of the relations among the different elements and the different layers (micro, macro and possibly meso) of the system. The need for a computational solution of the aggregate dynamics is the same as in DSGE models, but in DSGE models the aggregation problem can generally be solved analytically, thanks to a low (sometimes very low) level of heterogeneity. Hence some aggregate properties of the system can be analytically understood —in particular, whether the system is ergodic and stationary— although no closed form solution for the aggregate dynamics (hence for the likelihood function) can be derived.

The lack of an analytical formalization linking the behavior of the agents to the outcome of the system impedes a traditional approach to estimation, and calls for computational methods. These methods, known as simulation-based estimation techniques (Stern, 1997, 2000), have been originally developed in the econometric literature to deal with analytical models leading to criterion functions without simple analytical expression —for instance, because of integrals of large dimensions in the probability density function or in the moments. They are now increasingly used in mainstream modeling approaches, including DSGE; however, their application to AB models is still scant.

In this paper we directly address the relevant questions concerning estimation of AB models: What makes an AB model econometrically tractable? What are the necessary restrictions or assumptions that need to be imposed before an AB model can be estimated? How do they differ from those required for DSGE models? The answer to these questions depends in part on the structure of the models, and in part on the specific econometric technique employed. We will therefore spell out the general conditions under which estimation of an AB model makes sense, and then focus on the application of one specific technique, namely simulated minimum distance (SMD). As this approach to estimation is well known in the literature (Gouriéroux and Monfort, 1996), our work could be described as a mere transposition of concepts and language to a new area of application. However, the specificities of AB modeling make the exercise far from trivial. As LeBaron and Tesfatsion (2008, p. 249) put it, "the very properties that make ACE [Agent-based Computational Economics] models so interesting to study can cause empirical headaches when estimating them". They refer to the large degrees of freedom (both in terms of the number of parameters and in terms of functional forms) AB models often have due to their design flexibility, and to their nonlinear nature: this impinges on the properties of stationarity and ergodicity required

for estimation. Even when the models are indeed stationary and ergodic, the stress on out-of-equilibrium dynamics which is typical of the AB approach makes it unclear what could be estimated, and on which data.

In this work we propose no aspirine. Rather, we wish to describe the symptoms and their causes, discussing when possible the physical exercises that help keeping the headache at a manageable level. At present, they still fall short of the one-click solution available for many econometric problems; no doubt, this will sooner or later become available, but still there will be a need to increase the awareness of the empirical AB researcher.

More explicitly, we see three merits in our work. First, it provides a formalization of AB models and of their behavior that, in our view, is missing and strongly needed in the field. Second, it provides a reference and a guide for empirical work in the field of AB modeling. Third, it points out the analogies and differences between AB and DSGE models, not only in terms of methodological assumptions and solution concepts, but also in terms of their econometric consequences with respect to estimation issues. Although DSGE models can be criticized on theoretical and ultimately on empirical grounds (Fagiolo and Roventini, 2012), AB macro models share the same goals and, to some extent, face the same technical problems, including a large number of parameters and the need for a computational solution. It seems therefore sensible to look at the more mature DSGE literature to understand whether the technical solutions it has devised can also be adopted when taking AB macro models to the data.

The rest of the paper is organized as follows. We first put empirically-based AB models in retrospective, to highlight their increasing use for quantitative analysis of macro issues (section 2). We then turn to the DSGE literature, where the problem of estimating many parameters in models that need to be evaluated numerically has already been confronted with, and identify SMD as possibly the most promising method for estimation of AB models (section 3). In section 4 we present a formal characterization of AB models as recursive systems, and discuss the properties of their implied dynamics. The specific issues concerning estimation of AB models are discussed in section 5. In section 6 we spell out the conditions for consistency in minimum distance estimation, while in section 7 we show how they apply to AB models. In section 8 we summarize our arguments and suggest the directions for further research.

2 Little AB models grow big

AB models have long been considered as *theoretical exercises* aimed at investigating the (unexpected) macro effects arising from the interaction of many individuals —each following possibly simple rules of behavior— or the (unknown) individual routines/strategies underlying some observed macro phenomenum (Richiardi, 2012). As such, the typical AB model is a relatively small "toy" model, which can be used to understand relevant mechanisms of social interaction. While a seminal contribution came from the mainstream economics camp, with the work of Thomas Schelling on racial segregation which dates back to the late 1960s[1], the roots of AB models can be traced down to the evolutionary economics approach[2] and the so-called Santa Fe perspective on the study of complex systems[3]. Little data is involved with such a line of research.

A less known root of AB modelling can be identified in the dynamic microsimulation literature, and in particular in two almost forgotten works: Barbara Bergmann's microsimulation of the US economy (Bergmann, 1974) and Gunnar Eliassons microsimulation of the Swedish economy (Eliasson, 1977).[4] These models departed from standard microsimulations as they explicitly considered the interaction between the supply and demand for labor, and modeled the behavior of firms and workers in a structural sense.[5] These early attempts were aimed at providing still relatively small laboratory experiments, which were then calibrated in order to reproduce relevant features of the real economy in the two countries. The approach was not followed and this strand of the literature came to a near dead end[6], only to be resumed from a different perspective with the big macro models that have been developed following the AB methodology in the last 10 years.[7]

These models directly attempt at replacing the DSGE approach at the core of macroeconomics (Caballero, 2010) with an AB counterpart. Here, empirical validation is crucial, as the lack of

[1] Schelling (1969).

[2] Dosi and Nelson (1994).

[3] See Anderson et al. (1988); Arthur et al. (1997); Blume and Durlauf (2006). Although associated with the Santa Fe Institute, this perspective was initiated in Europe by chemists and physicists concerned with emergent structures and disequilibrium dynamics (Prigogine and Stengers, 1984; Haken, 1983).

[4] See Neugart and Richiardi (2013).

[5] By converse, in the wake of the seminal contributions by Guy Orcutt Orcutt (1957, 1961), dynamic microsimulations mainly developed as reduced form, probabilistic partial equilibrium models, with limited interaction between the micro unit of analysis and abundant use of external coordination devices in terms of alignment to exogenously identified control totals.

[6] Some followers continued Eliasson's effort, although with a more theoretically oriented twist (Ballot and Taymaz, 1997, 1999; Ballot, 2002).

[7] See Dawid et al. (2013) and the references therein.

empirical relevance is the ultimate critique that has been moved to DSGE models.[8] However, while in the last decade the literature on DSGE models has evolved from simple calibration to proper estimation, the AB macro models are still lagging behind, although some sophisticated examples of calibration are starting to close the gap (Bianchi et al., 2007, 2008; Fabretti, 2012; Cirillo and Gallegati, 2012).[9]

Our reading of the state of the art in AB macro modeling is that the literature is now ready to take off from where Bergmann and Eliasson left it: large rather than small laboratory experiments are starting to appear, but they are yet unfit for use by monetary and fiscal authorities because they have not been seriously taken to the data. In terms of the terminology of Chen et al. (2012), we have to move from stage I (the capability to grow stylized facts in a qualitative sense) to stage II (the selection of the appropriate parameter values based on sound econometric techniques).

Indeed, only a handful of examples of structural estimation of AB models exist, and they all involve rather simple models (most of them in the area of computational finance). For some of them direct estimation techniques can be used, as the models are simple enough to derive a closed form solution for the distribution of relevant statistics.[10] However, when the model has to be aggregated numerically, indirect estimation techniques must be used. Winker and Gilli (2001) and Gilli and Winker (2003) estimate respectively two and three parameters of an AB model of the foreign exchange market, by employing the method of simulated moments (MSM). Their focus is on optimization heuristics. In Winker et al. (2007) they deal with the problem of moments selection, and propose a set of statistics on exchange rate returns to estimate models of exchange rate. Amilon (2008) employs the efficient method of moments (and maximum likelihood on a simpler model) to

[8] See Colander et al. (2008); Solow (2010); Kirman (2010); Krugman (2011); Stiglitz (2011). However, Sims (1980) warns against judging macro models only in terms of fit with the data, and suggests that large scale models may fit the data well but may perform poorly out-of-sample due to non-credible identification restrictions. In a similar vein, Kocherlakota (2007) shows that a model that fits the available data perfectly may provide worse answers to policy questions than an alternative, imperfectly fitting model. In a more recent note, Sims (2006) suggests that DSGE models should be considered only as story-telling devices and that "it does not make sense to require these models to match in fine details the dynamic behavior of the accounting constructs and proxy variables that make up our data".

[9] The distinction between calibration and estimation is fuzzy and to some extent arbitrary (Hansen and Heckman, 1996). Here we draw a line according to the objective function which drives the choice of the parameter values: in calibration the goal is primarily data tracking, while estimation aims at recovering the true values of the parameters of the unknown data generating process underlying the observed data. As such, estimation is concerned with the properties of the estimators and the quantification of the uncertainty around the estimates. Classification is sometimes difficult. For instance, Fabretti (2012) adopts the same methodology as Gilli and Winker (2003) —which we consider an exercise in estimation— but falls short of providing standard deviations for her estimates. Moreover, no stationarity and ergodicity tests are provided for the moments chosen for comparing the model with the data.

[10] See Alfarano et al. (2005, 2006); Boswick et al. (2007) and the other references listed in Chen et al. (2012).

estimate an adaptive belief system with two or three agents and 15 parameters. Franke (2009) estimates, also by MSM, six parameters of an AB asset pricing model. Finally, Grazzini et al. (2012a,b) illustrate by means of a Monte Carlo analysis how the MSM works on a simple model of innovation diffusion, identifying the conditions under which the procedure leads to a small sample bias.

Common wisdom goes, the number of parameters must be kept to a minimum in order for estimation to be feasible. AB models are however not the only models possibly plagued by the curse of dimensionality — the fact that the convergence of any estimator to the true value of a smooth function defined on a space of high dimension (the parameter space) is very slow (De Marchi, 2005; Weeks, 1995). In particular, DSGE models not only feature a large number of parameters, but share with AB models an important aspect of complex systems: they include many nonlinear feedback effects. This implies that DSGE have to be simulated, rather than analytically solved. It is thus natural to turn to the DSGE estimation literature in search of guidance.

3 Estimation of DSGE models

In DSGE modeling the law of motion of the aggregate variables can be analytically derived, so that it is normally easy to understand whether the model is ergodic (which is typically the case) and/or stationary (which is typically not). In presence of non-stationarity, the model is rescaled, in order to make it stationary around a balanced growth path. Even after rescaling however, a closed form derivation of the steady state is in general not attainable. A common solution is therefore to linearize the system around the steady state.[11]

The data are then also de-trended, and the trend is often assumed to follow an exogenous stochastic process.[12] DSGE models are evaluated with respect to how well they describe the business cycle on top of this trend.[13]

[11] Linearization however is not neutral: it eliminates asymmetries, threshold effects and many other interesting phenomena. Other solution methods, which however also involve some degree of approximation, are projection algorithms and value function iteration (Fernández-Villaverde, 2010).

[12] Other frequent transformations of the data include the elimination of outliers and the selection of appropriately stable periods.

[13] In this perspective, the critique that DSGE models fail to account for abrupt changes in behavior, systemic crises, regime switches between different equilibria, *etc.* appears too severe. DSGE models are meant to say something about reality in tranquil times: no more than that. Of course one could question the usefulness of such complicated models inhabited by perfectly rational, forward-looking maximizing agents if their range of operation is restricted to situations when the present looks terribly like the past.

On the contrary, AB macro models have the ambition of describing the economy also out-of-equilibrium: this reduces the appeal of linearization around the steady state. Moreover, they solve the aggregation problem computationally, rather than by means of the representative agent assumption or other simplistic assumptions: hence, even if the attention is restricted to the steady-state, linearization brings little additional benefits (the sum of linear function is in general *not* a linear function).

However, having an (appropriately defined, see below) stable environment is a prerequisite for *any* exercise in estimation. Therefore, taking an AB model to the data also requires some data transformation.[14] The techniques used in the DSGE literature —the use of the Hodrick-Prescott filter to detrend the data for instance— are therefore valuable also for AB macro models.

Turning to the estimation methods, old vintage DSGE models (as most current AB macro models) where mainly calibrated: the values of the parameters were chosen according to some external knowledge, theoretical belief or empirical criterion. The tide has however turned in favor of a more formal estimation approach.[15]

The literature on the econometrics of DSGE is vast and increasing.[16] The main techniques used are maximum likelihood (ML), generally with the introduction of Bayesian priors, the generalized method of moments (GMM), the method of simulated moments (MSM) and indirect inference (II). We now briefly discuss each of these methods in turns, highlighting the potential for application to AB macro models.[17]

3.1 Maximum likelihood

This is the standard method for estimating DSGE models, so it deserves a little more space. Obtaining the likelihood of a DSGE model, that is the distribution of the observables conditional on the parameters, when no analytical solution is available requires a number of steps (Fernández-Villaverde, 2010). First, an approximate solution for the steady state values of the observables (also called *policy functions*) must be found, for instance by means of linearization. This solution can

[14] For instance, in the MSM this has to to with the selection of appropriate moments.

[15] An intermediate approach is to calibrate some parameters, and then estimate the others conditional on the values of the calibrated set.

[16] See Tovar (2008) and Fernández-Villaverde (2010) for an introduction.

[17] We omit a discussion of GMM, as this methodology is not feasible in an AB framework because analytical expressions for the unconditional moments as a function of the parameters are necessary.

(often, but not always) be expressed as a Markov chain —the so-called *state space representation*— where the aggregate state of the system at any time period depends only on the previous state, the current shocks, and the parameters, while the observables are a function of the current state (plus some measurement errors). Then, *filtering theory* can be applied in order to find the sequence of conditional distributions of the states given the past observations and the parameters. From this, Bayes rule gives the conditional distribution of observables given the past states. If the state space representation is linear (or linearized) and the shocks are assumed to be normally distributed, a simple Kalman filter can be applied: this directly gives the likelihood function. If however, the state space representation is not linear and the shocks are not normal, filtering becomes more complicated and must be performed by means of simulation (for instance, with the use of the so-called particle filter): the conditional distribution of the states given the past observations and the parameters is replaced by a simulated distribution.

Now, there are two (main) problems with this procedure. First, *stochastic singularity*, which arises when a small number of structural shocks is used to generate predictions about a large number of observable variables: the model then predicts a deterministic linear combination of observable variables which causes the likelihood to be 0 with probability 1.[18] Solutions to this problem involve reducing the number of observable variables on which inference is made (or using a projection from the set of observables to a smaller set of composed indicators), increasing the number of shocks, or adding measurement errors. None is problem-free, although there is a convergence in the literature toward estimating DSGE models with measurement errors.

The second problem is that ML estimation is very sensitive to model misspecification, which often leads to absurd parameter estimates. The reason is the very flat shape of the likelihood function, which might display a multitude of local optima: a slight change in the specification of the model can then lead to big differences in the estimates. In turn, the flatness of the likelihood function might be explained by the high degrees of freedom that are introduced in the model in order to increase the fit with the data, or it could be interpreted as a sign of gross misspecification of the model itself. In order to come around this problem, the standard procedure is to add Bayesian priors about the distribution of the parameters. This adds curvature to the likelihood function and

[18] Strictly speaking, stochastic singularity is a feature of linearized DSGE models, but it may also have implications for the estimation of nonlinear models depending on the extent to which they differ from their linearized counterparts.

"iron out" the bumps. The posterior distribution is then numerically obtained using Markov chain Monte Carlo (MCMC) methods, like the Metropolis-Hastings algorithm. Of course, no "Bayesian free lunch" exists (Poirier, 1998), and Bayesians prescribe that you take your pic-nic with you: if the likelihood is flat, the posterior distribution looks a lot like the prior distribution, and the model is of little help in increasing our understanding of reality. As stated by Canova and Sala (2009, p. 448), when models are under- or weakly-identied "reasonable estimates are obtained not because the model and the data are informative but because auxiliary restrictions make the likelihood of the data (or a portion of it) informative. In these situations, structural parameter estimation amounts to sophisticated calibration".

The question for our purposes is whether this complicated procedure (for which in the setting of DSGE models there exist ready-made algorithms and software) can be fruitfully applied to AB models, the problem being that the state space representation has no analytical formulation (more on this in the next section). One could think of using simulated empirical maximum likelihood to approximate the likelihood function, as for instance in Kristensen and Shin (2012). This class of estimators prescribes to simulate a conditional density, that is producing new simulated observations given the values of the parameters, and then optimize the values of the parameters in order to maximize the simulated empirical likelihood.

A simpler approach is to focus on some moments of the empirical likelihood, rather than on the entire conditional distribution. This is in essence what MSM does, and it will be the focus of the rest of the paper.

3.2 Indirect estimation

Under this heading we deal both with MSM, which prescribes to minimize the distance between the unconditional moments of the real data and those computed on the simulated data (which depend on the value of the parameters), and II (or extended method of simulated moments, EMSM), which prescribes to minimize the distance between the parameters of a meta-model estimated on the real data and those of the same meta-model estimated on the simulated data.

Often, a vector autoregression (VAR) is used as meta-model. In this case the interpretation in terms of moments is straightforward, as the coefficients of the VAR are proportional to the covariances and autocovariances of the variables included in the specification. An alternative to

matching the VAR parameters estimated on the real and the artificial data is matching the impulse responses of the DSGE model with those predicted by a VAR estimated on the real data.

Ruge-Murcia (2007) compares these techniques (and GMM) to ML estimation on a simple one-sector real business cycle model by means of Monte Carlo analysis. He finds that both GMM and MSM are more robust to misspecication and are less affected by stochastic singularity than ML. This is because "ML estimation is limited by the number of linearly independent variables while moment-based estimation is limited by the number of linearly independent moments. The latter is a weaker restriction because it is possible to find independent moments that incorporate information about more variables than those that are linearly independent." (Ruge-Murcia, 2007, p. 2600). He concludes that moment-based estimation methods (GMM and MSM) compare very favorably to the more widely used method of ML.

This adds, we believe, further interest to our focus on MSM as a natural estimation technique for AB models. Before turning to the specific issues of minimum distance (MD) estimation, to which MSM belongs, we present a formal characterization of AB models as recursive systems, which will frame our discussion on their equilibrium and out-of-equilibrium dynamics and help us to identify the theoretical properties that can be used for estimation.

4 AB models as recursive systems

AB models are recursive systems (Leombruni and Richiardi, 2005; Epstein, 2006). This is an essential feature as "[t]he elimination of simultaneous equations allows us to get results from a simulation model without having to go through a process of solution" (Bergmann, 1990).[19]

At each time t an agent i, $i \in 1 \dots n$, is fully described by some state variables $\boldsymbol{x}_{i,t} \in \Re^k$. Let the evolution of her state variables be specified by the difference equation:

$$\boldsymbol{x}_{i,t+1} = \boldsymbol{f}_i(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{-i,t}; \boldsymbol{\xi}_{i,t}; \boldsymbol{\theta}_i). \tag{1}$$

where $\boldsymbol{\xi}_{i,t}$ is a stochastic term. We assume that the behavioral rules[20] may be individual-specific

[19] One could also argue with Watts (1991) that "The world is essentially recursive: response follows stimulus, however short the lag".

[20] Here and in the following we use "behavioral rules" and similar terms in a loose sense that encompasses the actual intentional behaviors of individuals as well as other factors such as technology *etc.*

both in the functional form of the phase line $\boldsymbol{f}_i(.)$ and in the parameters θ_i, and may also depend on the state $\boldsymbol{x}_{-i}$ of all agents other than i. The set of structural equations (1) —one for each agent— specifies the *data generating process* (DGP) of the model.

At any point in time, the system is in a state $\boldsymbol{X}_t = (\boldsymbol{x}_{i,t})$ which is the matrix of all individual states. By replacing eq. (1) in the definition above, we obtain

$$\boldsymbol{X}_{t+1} = \boldsymbol{F}(\boldsymbol{X}_t, \boldsymbol{\xi}_t, \boldsymbol{\theta}). \tag{2}$$

Eq. (2) defines a Markov chain (MC), although with a very high (possibly infinite) number of states.[21] Quite interestingly, this is the same *transition equation* that is used for describing (the approximate solutions of) DSGE models, where the vector of innovations $\boldsymbol{\xi}_t$ includes the random shocks which are the engine of the dynamics (Fernández-Villaverde, 2010). However, while eq. (2) has an explicit analytical formulation in DSGE models, it remains only implicitly defined by the micro transition equations (1) in AB models.

Often, we are interested in some aggregate (observable) statistics of our economy. A vector of aggregate variables $\boldsymbol{Y}_t$ is defined as a (vectorial) function over the state of the system, that is as a projection from $\boldsymbol{X}$ to $\boldsymbol{Y}$:

$$\boldsymbol{Y}_t = \boldsymbol{G}(\boldsymbol{X}_t, \boldsymbol{\kappa}_t). \tag{3}$$

where $\boldsymbol{\kappa}_t$ are extra random terms that accounts for measurement errors and other shocks to the observables.

In DSGE modeling, this is the *measurement equation*, which together with the transition equation form the state space representation of the system. In DSGE, this also has an analytical formulation, while in AB modeling it does not. Stochastic singularity arises —in DSGE as well as in AB models— whenever the number of aggregate variables in $\boldsymbol{Y}_t$ is higher than the number of shocks in $\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = [\boldsymbol{\xi}_t \ \boldsymbol{\kappa}_t]$ is the stacked matrix of stochastic terms. This is because the variables in excess are a deterministic function of the other observables.

[21] For a discussion of the Markov chain representation of AB models, see Izquierdo et al. (2009). Sometimes it is possible to collapse $\boldsymbol{X}$ into a smaller number of states. The conditions for a projection of a MC still to be a MC are known as *lumpability* (or strong lumpability), and necessary and sufficient conditions for lumpability are known (Kemeny and Snell, 1976; Banish et al., 2012) (they have to do with simmetries in the micro state space). In general it may happen that, for a given MC, some projections are Markov and others are not.

4.1 Equilibrium

One important difference between DSGE and AB models lies in the definition of equilibrium. In DSGE models, equilibrium is defined as a consistency condition in the behavioral equations: agents (whether representative or not) must act consistently with their expectations, and the actions of all the agents must be mutually consistent. This is the methodological prescription of *rational expectations*, and logically operates at an individual level before the state space representation of eqs. (2) and (3). The system is therefore always in equilibrium, even during a phase of adjustment to a shock. AB models, on the other hand, are characterized by *adaptive expectations*, according to which consistency might or might not arise, depending on the evolutionary forces that shape the system. An equilibrium can therefore be defined only at the aggregate level and only in statistical terms, after the macro outcomes have been observed.

Let $\boldsymbol{Z}_0 = \{\boldsymbol{X}_0, \boldsymbol{s}\}$ be the set of initial conditions of the simulation, where $\boldsymbol{X}_0$ is the initial state of the system and $\boldsymbol{s}$ stands for the seed(s) used by the random number generator(s) in the simulation, which determine(s) the evolution of $\boldsymbol{\varepsilon}$.

Definition If, in a given time window $[\underline{T}, \bar{T}]$, an aggregate outcome of the AB model Y_t is stationary, the model is said to have a *statistical equilibrium* $Y^*(\boldsymbol{Z}_0, \boldsymbol{\theta})$, with respect to Y and for given initial conditions $\boldsymbol{Z}_0$ and values of the parameters $\boldsymbol{\theta}$.[22]

We distinguish between two types of statistical equilibrium: absorbing and transient.

Definition A statistical equilibrium is said to be *absorbing* if $Y^*(\boldsymbol{Z}_0, \boldsymbol{\theta})$ is stationary in $[\underline{T}, \bar{T} + \tau], \tau \to \infty$.

Definition A statistical equilibrium is said to be *transient* if $Y^*(\boldsymbol{Z}_0, \boldsymbol{\theta})$ is no longer stationary in $[\underline{T}, \bar{T} + \tau], \tau > 0$.

For any given value of the initial conditions and the parameters, there can be at most one absorbing statistical equilibrium –call it Y^*_a. However, there might be many transient statistical equilibria Y^*_j: for instance, a model can oscillate between two (or more) transient statistical equilibria.

[22] By stationarity, here and in the rest of the paper, we mean weak stationarity. A stochastic process $\{y_t\}$ is (weakly) stationary if $E(y_t) = \mu$, that is, its expected value is independent of t, and if $\mathrm{Cov}(y_t, y_{t-j})$ exists, is finite and depends only of j for any i.

If Y_t is also ergodic, irrespective of the initial conditions the model will always end up finding the same statistical equilibria. An implication is that if Y_t is ergodic and the model has an absorbing statistical equilibrium, this equilibrium is unique and depends only on the parameters: $Y_a^*(\boldsymbol{Z}_0, \boldsymbol{\theta}) = Y_a^*(\boldsymbol{\theta})$. Therefore, the simulated mean, variance and autocovariance –which we generally label $m_t(Y^*, \boldsymbol{\theta})$– converge, both over time and over the replications r, to the theoretical limiting moments of the underlying DGP conditional on the parameters used for the simulation, $\mu(Y^*, \boldsymbol{\theta})$ (if they exist):[23]

$$\lim_{t\to\infty} m_t(Y^*, \boldsymbol{\theta}) = \lim_{r\to\infty} m_r(Y^*, \boldsymbol{\theta}) = \mu(Y^*, \boldsymbol{\theta}) \tag{4}$$

The properties of the time series generated by the model are constant both in time and across replications: therefore, they can be inferred from the sample moments. In other words, the main simulated moments of Y, computed either over time or over replications, are consistent estimators of the theoretical limiting moments, for $t > \bar{T}$.

Moreover, the transient equilibria, if any, are also invariant across different instantiations of the model. Across different replications, $\{Y_{j,r}^*\}\ \forall j$ is a sequence of independent and identically distributed random variables. By the law of large numbers, as the number of replications increases any simulated moment $m_r(Y_i^*)$ converges to the theoretical limiting moment of the underlying DGP conditional on the parameters used for the simulation, $\mu(Y_j^*, \boldsymbol{\theta})$ (if it exists):

$$\lim_{r\to\infty} m_r(Y_j^*) = \mu(Y_j^*, \boldsymbol{\theta}),\ \forall j. \tag{5}$$

and the simulated moments of Y_j^* are consistent (over the replications) estimators of the theoretical limiting moments, for $t \in [\underline{T}_i, \bar{T}_i]$.

On the contrary, if Y_t is stationary but not ergodic, different (absorbing and/or transient) equilibria are obtained, for the same values of the parameters, depending on the initial conditions. This can be regarded as a case of multiple statistical equilibria. [24]

In some cases it is possible to relax the ergodicity condition replacing it with *subergodicity*, that is, ergodicity in subsets of the parameters space. A subergodic model is conditionally (to a known

[23] If Y_t is strictly stationary, any simulated moment –not only the mean, variance and autocovariance– converges to its theoretical counterpart, if it exists.

[24] Note the two different flavors of multiple equilibria, over time (different transient equilibria) and over replications (e.g. different absorbing equilibria, depending on the initial conditions, if the model is not ergodic).

set of initial conditions) ergodic. If the model has an equilibrium but the equilibrium depends on a given set of initial conditions, estimation can be performed by treating the initial conditions as additional parameters to be estimated, or by directly introducing the observed data as initial conditions.

Note how flexible and powerful this descriptive apparatus is. For instance, a model can show an absorbing statistical equilibrium for, say, GDP. This means that after an initial adjustment period up to $\underline{T}$, the GDP series becomes stationary, with constant mean and variance. If the system receives a transitory shock, it moves away from the statistical equilibrium. However, once the shock has passed, if the model is ergodic it comes again to the previous steady state, after an adjustment phase. If we rerun the model and shock it 100 times, it would always come down to the same equilibrium: its properties are then suited for estimation. Moreover, it might happen that during the adjustment process some other transformation of the state of the system, for instance the speed of adjustment to the equilibrium level of GDP, becomes stationary. This new regularity breaks down when GDP reaches its steady state: it is therefore a transient statistical equilibrium. Again, if the model is ergodic the properties of the transient equilibrium are invariant across simulation runs and can be used for estimation.

It is possible that a model displays no absorbing equilibrium for a given variable of interest. To continue with our example, think of the evolution of GDP, with business cycle fluctuations of different amplitude and duration, and intermittent crises. This is an interesting case for many AB modelers, who essentially see the world as a disequilibrium process. Even in such situations however, it might be possible to find statistical regularities with respect to some other variable, for instance the distribution of the lenght and depth of recessions. If this is the case, we are back to equilibrium analysis, and if the model is ergodic we might be able to estimate it. Moreover, these regularities might be only transient, and vanish as the simulated time goes by (think for example of the effects of fiscal policies on GDP growth, which are very different depending on whether the economy is close to full employment or not). Again, if they are stable enough across different replications of the model, we might use them for estimation. If, on the other hand, the model exhibits no regularities whatsoever, no matter how the data are transformed, then the model is not suited for estimation, and one might argue that it is also of limited explanatory (not to say predictive) help: "everything can happen" is hardly not a good theory. So, when AB researchers

speak of disequilibrium or out-of-equilibrium analysis, what they have in mind, we believe, is really transient statistical equilibrium analysis of some sort, and this is in principle amenable to estimation.

4.2 A labor market example

The following example, in the spirit of Bergmann (1990), might help clarifying the issue of multiple equilibria and the role of ergodicity (we purportedly keep it very simple so that its dynamics can be understood analytically). Suppose we have a model of the labor market where individuals experience random transitions between employment (E) and unemployment (U): in any period $t = 1, 2, \cdots$ the probability of being fired is p_{EU}, while the probability of being hired is p_{UE}. The initial employment status is random, with a probability of being unemployed equal to u_0. The model defines, for every individual, a time-homogeneous Markov chain with transition matrix

$$P = \begin{array}{c} \\ E \\ U \end{array} \begin{array}{c} \begin{array}{cc} E & U \end{array} \\ \begin{pmatrix} p_{EE} & p_{EU} \\ p_{UE} & p_{UU} \end{pmatrix} \end{array} \tag{6}$$

with $p_{EE} + p_{EU} = p_{UE} + p_{UU} = 1$.

This is at the individual level. The Markov Chain representation of the system at the aggregate level entails expanding the number of states to cover all possible configurations of the N individuals in the population: from the case when no one is unemployed, to the case when everyone is. However, we do not need to do that. Since all individuals are alike and there is no interaction between them, the fraction of time a (representative) individual stays unemployed is equal to the overall average unemployment rate, u. The limiting distribution of states is given by

$$u = p_{EU}(1 - u) + p_{UU}u \tag{7}$$

Hence, the system has a statistical equilibrium where the unemployment rate follows a normal distribution with mean $u^* = \frac{p_{EU}}{1+p_{EU}-p_{UU}}$. Different initial conditions –the random distribution of employment states– do not affect the stationary state: the model is ergodic, and the statistical

equilibrium is unique.[25]

Now, suppose that the model has a regime switch: if the unemployment rate reaches $\bar{u} > u^*$, the economy collapses and the individual transition matrix becomes

$$Q = \begin{array}{c} \\ E \\ U \end{array} \begin{array}{c} \begin{array}{cc} E & U \end{array} \\ \begin{pmatrix} q_E & q_U \\ q_E & q_U \end{pmatrix} \end{array} \tag{8}$$

with $q_E + q_U = 1$, and $q_U > \max(p_{EU}, p_{UU}, \bar{u})$. This introduces a limited degree of interaction in the model, as what happens to any individual agent at time t depends on what happened to the other agents before t. The matrix Q, *per se*, supports an equilibrium $u^{**} = q_U > u^*$. If $u^*, \bar{u}$ and u^{**} are sufficiently far apart the two equilibria u^* and u^{**} coexist as transient equilibria: a series of lucky (or unlucky) draws can actually move the system from one equilibrium to the other. Over the long run, the time spent in the two statistical equilibria is inversely proportional to how close they are to the tipping threshold $\bar{u}$.[26]

Note that the model is still ergodic (the two equilibria are always selected): therefore, if the model was correctly specified and the two regimes were also observed in the real data, we could use our knowledge of the statistical properties of the two equilibria to estimate the four free parameters p_{EU}, p_{UU}, q_U, and $\bar{u}$.[27]

However, for a broad range of values of the parameters, we are not very likely to observe regime switches. Suppose for instance that $p_{EU} = .05, p_{UU} = .5, q_U = .6, \bar{u} = .4$: the two equilibrium values are $u^* = 9.1\%$ and $u^{**} = 60\%$. Even with a small population size of $N = 100$, the probability of switching from the low unemployment equilibrium to the high unemployment one is 1.7×10^{-31}, leading to an average waiting time of 6×10^{30} periods. This is much bigger than the age of the universe in seconds, estimated in 4.3×10^{17} seconds (13.7 billion years). Conversely, the probability of switching from the high unemployment equilibrium to the low unemployment equilibrium is 1.8×10^{-05}, for an average waiting time of 55,000 periods, a more mundane number but still big

[25] Ergodicity of the time series u_t is not the same thing as ergodicity of the Markov Chain P. We do not elaborate on the issue.

[26] If one or both the equilibria are too close to the tipping threshold, the two equilibria merge and the system converges to a new absorbing equilibrium u^{***}.

[27] The four parameters would not be separately identifiable when u^* and u^{**} are too close to each other.

enough to make a regime change a remote possibility.

So, to any practical reason this parameterization is indistinguishable from a model where a first flip of a coin decides in which regime we are, and then transitions take place according either to P, or Q. This however is a non-ergodic system, and non kosher for estimation: while in the real world we only observe data around one equilibrium, the model is capable to produce both equilibria, depending on the initial conditions (that is, depending on u_0). Any statistics mixing the results from different simulation runs will therefore produce nonsense estimates.

However, the model still behaves as subergodic, conditional on u_0. If we define this initial condition as an additional parameter, then the model becomes ergodic as the equilibrium value depends only on the value of the parameters. This additional parameter can be estimated in the data, or treated as given and replaced for with the unemployment rate in the first period of observation.

5 Estimation of AB models

With appropriate micro data, it is in principle possible to estimate directly the behavior of all the agents in the system, and then feed these estimates in the model in order to produce aggregate statistics that could in turn be compared with the available macro evidence. Systematic differences between the simulated and the observed macro data could then be interpreted as evidence of specification errors (in the way individual agents interact and the micro behavior is aggregated). However, such abundance of micro information is rare, and one has often to content with macro data.

In the following we assume that only observations on the aggregate behavior of the system are available; moreover, we assume that the model is correctly specified. This is a fundamental hypothesis, which implies that the model perfectly describes the real world, that the structural parameters exist and have a finite value and that all the parameters in the real system are represented in the model.[28] Under correct specification, the model defines the behavior of the system over the whole set of parameters θ while the observations are an instance of the system with the *true* set of parameters, θ_0. Given the model and a set of observations, any estimation procedure is

[28]The econometric literature has developed a number of specification tests. The extension of these tests to AB models is an interesting line of research that we do not pursue here.

based on the comparison of the properties of the model and the properties of the data. The main difference between estimation of DSGE and estimation of AB models lies in the knowledge which can be obtained about the properties of the model. While in a DSGE model (some) properties of the steady state can be studied analytically, in an AB model they must be inferred from the artificial data which the model itself generates: the only information about the model comes from the simulation of the model. This can be seen directly in eq. (3): while it is *always* theoretically possible to solve for $\boldsymbol{Y}_t$ by substituting in eqs. (1)-(2), individual heterogeneity and the specificities of the behavioral rules can easily grow the analytical expression of $\boldsymbol{Y}_t$ enormous, hindering any attempt at algebraic manipulation. On the contrary, in DSGE models eq. (3) is kept manageable by limiting the heterogeneity of the agents and by conditioning their behavior on a limited set of variables.

In an AB model knowledge about the statistical properties of the equilibria can only be inferred from the analysis of the simulation data.

Analytical knowledge of the properties of the model generally implies that ergodicity and stationarity (of the model) can be assessed at a theoretical level. At an empirical level, ergodicity of the true DGP cannot be assessed in the real data, as Nature only plays once, and is therefore assumed (or derived from the assumption of correct specification). Only stationarity of the real data needs therefore to be assessed by statistical testing (and this leads, in DSGE estimation, to de-trending of the data). This is true even when the criterion function (for instance, the likelihood) in the estimation procedure has no simple analytical expression, as in DSGE Models.

By converse, it is not possible to know *a priori* if an AB model produces stationary and ergodic series, and these properties must be tested in the simulation data. Simulation works here as a powerful exploratory device, generating most of the knowledge about the behavior of the system.

To summarize, there are two issues in estimation of AB models, and they both descend from the fact that the properties of the model are not analytically known: the objective function used for estimation must be numerically evaluated, and ergodicity and stationarity of the aggregated series used for estimation must be tested. The first problem is common to other modeling strategies (as in DSGE), and calls for the use of simulation-based econometric techniques. The second problem requires a deeper analysis of the simulated data.

As we have seen in section 3, possibly the most promising techniques for estimating AB models

are SMS and II. Both techniques belong to the general class of simulated minimum distance. SMD summarizes the real system and the artificial system by means of a set of functions on the observed and artificial data and minimizes the difference between the "short summaries" of the two systems. SMD is a general estimation method as different functions can be used to characterize the data. When these functions are moments of the data we get the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989). When these functions are the coefficients of the "meta-parameters" estimated on an arbitrary (and therefore by definition misspecified) meta-model we get indirect inference (Gouriéroux et al., 1993). In this case the estimation procedure minimizes the distance between the meta-parameters estimated on the real data and the meta-parameters estimated on the artificial data (which are a function of $\boldsymbol{\theta}$). II has a very interesting application which allows to estimate a model without (directly using) the data. This can be done by choosing a meta-model already estimated in the literature: the same specification and the same estimation methods are applied to the artificial data produced by the AB model, and $\boldsymbol{\theta}$ is changed until the distance between the coefficients of the meta-model estimated on simulated data and those estimated in the literature is minimized.

In the next section we review the conditions for consistency in minimum distance estimators, while in the following one we discuss how these conditions apply to simulated minimum distance estimation of AB models.

6 Consistency in minimum distance estimators

Define a set of estimators $\hat{\boldsymbol{h}}_n$ and a vector of functions $\boldsymbol{h}(\boldsymbol{\theta})$, where the estimators are functions of the data, the functions $\boldsymbol{h}(\boldsymbol{\theta})$ are the mapping between the model and the estimators and $\boldsymbol{\theta}$ are the structural parameters. An estimation of $\boldsymbol{\theta}$ can be constructed by maximizing the following objective function:[29]

$$\hat{Q}_n(\boldsymbol{\theta}) = -\big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big)' \hat{W}_n \big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big) \tag{9}$$

where $\hat{W}$ is a positive semi-definite matrix. That is,

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \hat{Q}_n(\boldsymbol{\theta}). \tag{10}$$

[29] This section draws heavily on Newey and McFadden (1994).

If the model is well-specified, the *true* data generator process can be thought of as an instance of the model in which $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The estimators can then be rewritten as $\hat{\boldsymbol{h}}_n = \boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n)$, in which the equivalence between the *true* data generator process and the model is explicit and in which the observation error is denoted as $\boldsymbol{\varepsilon}_n$.

Define $\boldsymbol{h}_0 = \boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n) = \boldsymbol{h}(\boldsymbol{\theta}_0)$ as the *true* value of the estimators. For this value, assuming W is a semi-positive matrix, the objective function then becomes:

$$Q_0(\theta) = -\big(\boldsymbol{h}_0 - \boldsymbol{h}(\boldsymbol{\theta})\big)' W \big(\boldsymbol{h}_0 - \boldsymbol{h}(\boldsymbol{\theta})\big). \tag{11}$$

An estimator $\hat{\boldsymbol{\theta}}$ is said to be consistent if, as the number of observations increase, it converges in probability to the parameters to be estimated, $\boldsymbol{\theta}_0$. The minimum distance estimator belongs to the class of *extremum estimators*, which also includes maximum likelihood, nonlinear least square and generalized method of moments.

The consistency conditions for extremum estimators, including the minimum distance estimator defined in equation (10), are given in the following theorem (Newey and McFadden, 1994, p.2121):

Theorem If there is a function $Q_0(\boldsymbol{\theta})$ such that (i) $Q_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$; (ii) $\boldsymbol{\Theta}$ is compact; (iii) $Q_0(\boldsymbol{\theta})$ is continuous; (iv) $\hat{Q}_n(\boldsymbol{\theta})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$, then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$.

The theorem states that if $\hat{Q}_n(\boldsymbol{\theta})$ converges to $Q_0(\boldsymbol{\theta})$ for *all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $Q_0(\boldsymbol{\theta})$ is maximized only at the *true* parameters, then the limit of the maximum $\hat{\boldsymbol{\theta}}$ is the maximum of the limit $\boldsymbol{\theta}_0$.

The conditions stated in the theorem are actually stronger than necessary: compactness can be relaxed, if the objective function is well-behaved when $\boldsymbol{\theta}$ is unbounded; continuity can be replaced with semi-continuity; $\hat{\boldsymbol{\theta}}$ needs only to be the superior of the objective function, rather than the maximum (Newey and McFadden, 1994, p.2122).

Assuming that the more technical conditions are satisfied, it is essential for consistency that $\hat{Q}(\boldsymbol{\theta})_n$ converges to $Q_0(\boldsymbol{\theta})$ and that the limit $Q_0(\boldsymbol{\theta})$ have a unique maximum at the true parameter value.

6.1 Convergence

Convergence is a property of the real system and requires that the observations are informative of the data generator process: as the data increase the amount of information available also increases.

The essential condition that allows $\hat{Q}_n(\boldsymbol{\theta})$ to converge to $Q_0(\boldsymbol{\theta})$ is that $\hat{\boldsymbol{h}}_n$ converges to $\boldsymbol{h}_0$; that is, the statistics used to characterize the model have a finite *true* value, $\boldsymbol{h}_0$ and they converge toward this value. The ergodic theorem states the conditions for this to happen: the statistics must be first moments of stationary and ergodic series $\boldsymbol{Y}_t$. Given that $\boldsymbol{Y}_t$ is stationary, its variance $\boldsymbol{\sigma}^2_{\boldsymbol{Y}}$ does not change over time, so that by increasing the number of observations the sample mean $\hat{\boldsymbol{h}}_n = \bar{\boldsymbol{Y}}_t$, computed over n observations, provides more accurate estimates of the theoretical mean $E[\boldsymbol{Y}]$ –the variance of the sample mean goes to zero. In other words: if the $\hat{\boldsymbol{h}}_n$ functions are sample averages over a given stationary and ergodic (possibly non linear) transformation of the data, with finite expected value $E[\hat{\boldsymbol{h}}_n] = \boldsymbol{h}_0$ and finite variance, then $\hat{\boldsymbol{h}}_n \xrightarrow{p} \boldsymbol{h}_0$.[30]

The consistency theorem requires a further condition on the type of convergence: uniform convergence. This property is defined in (Newey and McFadden, 1994, p.2121) as

Definition Uniform convergence in probability: $\hat{Q}_n(\boldsymbol{\theta})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$ means $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \xrightarrow{p} 0$.

Uniform convergence requires that the objective function is converging at the same speed for all $\boldsymbol{\theta}$. The conditions for uniform convergence of sample averages are more strict than those for convergence, and are known as the *uniform law of large numbers*: if the objective function is continuous and finite for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the data are stationary and ergodic and $\boldsymbol{\Theta}$ is compact then there is uniform convergence (Newey and McFadden, 1994, p.2129).

6.2 Identification

For consistency it is essential that $Q_0(\boldsymbol{\theta})$ has a unique maximum at the true parameter value. The conditions for identification in minimum distance are similar to the conditions required in the generalized method of moments. Supposing that convergence is satisfied, $\hat{\boldsymbol{h}}_n \xrightarrow{p} \boldsymbol{h}_0$ and $\hat{W}_n \xrightarrow{p} W$, W positive semi-definite, then $\hat{Q}_n(\boldsymbol{\theta}) = -\big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big)'\hat{W}_n\big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big) \xrightarrow{p} \big(\boldsymbol{h}_0 - \boldsymbol{h}(\boldsymbol{\theta})\big)'W\big(\boldsymbol{h}_0 -$

[30] An additional requirement is that $\hat{W}_n$ converges to W, this last condition being less binding since $\hat{W}$ can be defined as an identity matrix without impinging on the consistency of the estimations.

$\boldsymbol{h}(\boldsymbol{\theta})\big) = Q_0(\boldsymbol{\theta})$. The condition for $Q_0(\boldsymbol{\theta})$ to have a unique maximum at $\boldsymbol{\theta}_0$ is that $\boldsymbol{h}(\boldsymbol{\theta}_0) = \boldsymbol{h}_0$ and $\boldsymbol{h}(\boldsymbol{\theta}) \neq \boldsymbol{h}(\boldsymbol{\theta}_0)$ if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and W is nonsingular (Newey and McFadden, 1994, p.2128). When $\boldsymbol{h}(\boldsymbol{\theta})$ is linear in the parameters identification is assured if the rank condition is satisfied, which means that the number of functions in $\boldsymbol{h}(\boldsymbol{\theta})$ is equal to or greater than the number of parameters to be estimated and that the functions in $\boldsymbol{h}(\boldsymbol{\theta})$ are linearly independent.

When $\boldsymbol{h}(\boldsymbol{\theta})$ is nonlinear general conditions for identifications are difficult to find and "a practical solution to the problem of global identification, that has often been adopted, is to simply assume identification" (Newey and McFadden, 1994, p.2127). In some cases it is possible to prove local identification by linearizing the model around the equilibrium, as common in DSGE models.

In the case of simulation methods or in general in the case of objective functions which cannot be maximized analytically it is necessary to use optimization heuristic. Given the complicate nature of the objective function and the stochasticity of the optimization procedure it is possible that the heuristic converges to a local maximum. To verify that the selected maximum is a global maximum it is possible to perform several maximizations with different initial conditions and different parameters in the optimization heuristics and choose the estimate that maximizes the objective function among the converged values.

7 Minimum Distance in AB Models

As we have discussed above, the relationship between the behavior of an AB model and the structural parameters remains hidden in the recursive structure of the model, and only inductive evidence can be obtained by analyzing the simulated data. This implies that it is not possible to write analytically the objective function as in eq. (11) and that the properties of the statistics used in the objective functions are also not analytically known.

Simulation based econometric methods overcome this problem by simulating the model and using the artificial data to estimate $\boldsymbol{h}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Simulations allow to extend minimum distance estimation by replacing the theoretical properties of the model, $\boldsymbol{h}(\boldsymbol{\theta})$, with their simulated counterpart, $\tilde{\boldsymbol{h}}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)$. $\tilde{\boldsymbol{h}}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)$ is a random variable, where $\boldsymbol{\varepsilon}_s$ denotes the random term in the simulated data. For the sake of clarity suppose that the vectorial function $\boldsymbol{h}(\boldsymbol{\theta})$ is defined by a set of moments. Simulated minimum distance minimizes the weighted distance between the set

of observed moments, $\hat{\boldsymbol{h}}_n = \boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n)$ and the set of simulated moments $\tilde{\boldsymbol{h}}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)$. The objective function then becomes:

$$\hat{Q}_n(\boldsymbol{\theta}) = -\big(\boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n) - \tilde{\boldsymbol{h}}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)\big)' \hat{W}_n \big(\boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n) - \tilde{\boldsymbol{h}}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)\big) \tag{12}$$

The simulation method introduces additional variability into the objective function since the theoretical properties are estimated using a finite number of simulation data.[31] Supposing that it is possible (in terms of computational time) to simulate the theoretical moments for each possible combination of the parameters, the technique gives the value of the objective function for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, so that a $\hat{\boldsymbol{\theta}}$ is chosen such that the objective function is maximized.[32] Often maximization is performed by means of an optimization heuristic which introduces further variability in the outcome (Gilli and Winker, 2008). In the following this source of variance is neglected by supposing that an efficient optimization heuristic is available.

In principle it is always possible to build an objective function like eq. (12). We now turn to investigate under which conditions an AB model can be consistently estimated by simulated minimum distance.

7.1 Consistency conditions and AB models

Consistency conditions for AB models refer to the properties of the statistics used for estimation. The attention therefore turns from the objective function to the constituents of the objective function. Preliminarily, note that in a computer program as an AB model the objective function is discrete by construction, since only a countable number of values of the parameters can be tested. A threshold must therefore be chosen which specifies the level of precision in the estimates, that is the level below which further improving the precision gives no remarkable improvements in the objective function.[33] In the following it is assumed that $\boldsymbol{\Theta}$ is a discrete set, and that $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

[31] If the functions $\boldsymbol{h}$ refer to the (unique) absorbing equilibrium of the system, it is equivalent to simulate R runs of the model, each elapsing S periods in the statistical equilibrium, or simulate only one run and look at the statistical equilibrium for $R \times S$ periods. Conversely, if the functions $\boldsymbol{h}$ refer to transient equilibria, it is necessary to simulate $R \times S$ adjustment processes (in different simulation runs, or in the same simulation run but having the system exogeneously shocked $R \times S$ times) to obtain the same amount of data.

[32] Fixing the random numbers across iterations is required to achieve uniform convergence (McFadden, 1989; Pakes and Pollard, 1989).

[33] The same problem is faced by *any* numerical optimization procedure, even when an analytical expression for the objective function is available (e.g. the likelihood function).

7.1.1 Convergence

The first essential condition for convergence of the objective function is the convergence of its elements to their *true* value. In particular, since the simulated moments (or other summary measures) are used as an estimate of the theoretical moments, it is crucial to know whether the corresponding estimators are consistent. The conditions are the same as in section 6.1, but this time they must be checked in the simulated data: the moments must be first moments of stationary and ergodic series, so that the ergodic theorem and the uniform law of large numbers can be invoked.[34]

Each statistics used in the objective function should be tested for stationarity in the simulated data.[35] Given that the distribution of these statistics is *a priori* unknown, a nonparametric test must be used (and this impinges also when testing for stationarity in the real data). However, nonparametric tests have in general sufficient power only with a large amount of data: this is in principle a problem with real data, much less so with artificial data, which can be produced at will (computing time being the only constraint).[36] Each statistics should also be tested for ergodicity. This test is possible only in the artificial data for the simple reason that usually only one extraction from the real DGP is available, and builds on the possibility of simulating the model with different initial conditions.[37]

If each moment in the objective function is ergodic and stationary than the ergodic theorem applies and the moments converge in probability toward their expected value. This implies two fundamental properties of the objective function. The first is that by increasing the number of simulations the theoretical moments converge in probability to their expected value, which implies that the additional variance introduced by the simulation tends to zero. The second is that the observed moments converge in probability to their expected value and in turn the objective function $\hat{Q}_n(\boldsymbol{\theta})$ converges in probability to $Q_0(\boldsymbol{\theta})$.

Given that the objective function is converging, to have uniform convergence we need to inspect

[34] Note that this condition is excluding a large set of statistics often used to calibrate AB models, as for example the distribution of firm size which is computed at a point in time and is not a sample average.

[35] This can also be regarded as a specification test: the results of the stationarity test in the simulated data and in the real data should be similar if the model is correctly specified. Space constraints do not allow us to elaborate further on this point.

[36] If the real data cannot be tested for stationarity due to small sample size, the assumption of correct specification must again be invoked.

[37] Nonparametric tests for stationarity and ergodicity of the time series produced by an AB model are proposed in Grazzini (2012).

the objective function to assure that it exists and is finite for each $\boldsymbol{\theta}$. This can be done by sensitivity analysis.

7.1.2 Identification

Identification is hard in complex models. This is true for DSGE as for AB models, which are possibly plagued not only by a large number of parameters, but also by complicated feedbacks between agents and between the micro and the macro scale. Hence, a first sensible operation is to reduce the number of parameters to be estimated (this is akeen the *limited information approach* to the estimation of DSGE models, as described in Fagiolo and Roventini (2012)). For instance, it might be possible to drop observationally equivalent parameters from the specification, use parameters estimated elsewhere (for example, the productivity of capital and labor and other technological parameters), use directly real data in the model (the stock market returns, for example), or use sensitivity analysis to select the parameters that actually make a difference in the macro behavior of the model. The estimation procedure should therefore strictly follow the research question that originated the model by selecting the parameters that are crucial for the evolution of the model and that are important for the emergence of the properties under investigation.

Given the chosen moments and the parameters to be estimated it is possible to state a very intuitive (although restrictive) condition for identification.

Theorem If the moment conditions are equal to or greater than the parameters to be estimated, they are linearly independent and monotonically increasing *or* monotonically decreasing in all the parameters, then the objective function has a unique maximum.

Proof Convergence implies that the objective function $Q_0(\boldsymbol{\theta})$ is maximized at $\boldsymbol{\theta}_0$, *i.e.* $Q_0(\boldsymbol{\theta}_0) = 0$ and $\boldsymbol{h}(\boldsymbol{\theta}_0) = 0$. Monotonicity implies that the only point in the parameter space in which all $\boldsymbol{h}(\boldsymbol{\theta})$ are null is $\boldsymbol{\theta}_0$. Finally, if the moment conditions are linearly independent functions, then $Q_0(\boldsymbol{\theta})$ is null only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

If the number of parameters is such that it is computationally feasible to compute each moment for each combination of the parameters (conveniently discretized), then it is possible to draw the objective function and visually check whether there exists a unique maximum. The last option, if

an heuristic optimization is required, is to perform the maximization several times and to select the estimated parameters that correspond to the maximum value of the objective function.

7.2 Small-sample estimation bias

Non linear functions (moments) imply a bias. This is easy to show in the case of perfect identification when there is only one parameter and one moment.[38] As before, $h(\theta)$ is the function mapping the structural parameter θ into the moment chosen for estimation, while θ_0 is the true value of the parameter. The observed moment can be written as

$$h(\theta_0, \zeta_n) = h(\theta_0) + \zeta_n \tag{13}$$

The observed moment is the "true moment" plus an error ζ_n which depends on the sample (and on sample size). Supposing that the underlying process is ergodic and stationary and that the moment estimator is unbiased ($E(\zeta_n) = 0$), by applying the ergodic theorem we know that the theoretical moment is consistently estimated by the sample moment. As the number of observations increases the variance of ζ_n, decreases. The moment estimator, given the assumptions, is therefore unbiased and consistent. The method of simulated moments minimizes the distance between the observed moment and the simulated moment with respect to the parameter:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \ \hat{Q}_n(\theta) = -(h(\theta_0, \zeta_n) - h(\theta))^2 \tag{14}$$

where $\hat{\theta}$ is the estimated parameter and $h(\theta)$ is the theoretical (simulated) moment computed using the model.[39] If the moment function and the parameter are continuous, the maximization condition is:[40]

$$h(\theta_0, \boldsymbol{\varepsilon}_n) - h(\hat{\theta}) = 0. \tag{16}$$

[38] See Grazzini et al. (2012a) for an application.

[39] As we have seen, in a simulation based econometric framework $h(\theta)$ is simulated and becomes $\tilde{h}(\theta, \boldsymbol{\varepsilon}_s)$. For simplicity we omit here to consider this additional source of variability.

[40] Supposing differentiability of the conditional moment, $\hat{\theta}$ is selected by setting the first derivative of the objective function to zero:

$$\frac{d\hat{Q}_n(\theta)}{d\theta} = 2\big(h(\theta_0, \zeta_n) - h(\theta)\big)\frac{dh(\theta)}{d\theta} = 0 \tag{15}$$

Since the conditional moment has to be monotonic in order to identify the parameter, it must be that $\frac{dh(\theta)}{d\theta} \neq 0$ for all θ and all h, thus leading to condition 16.

Using equation (13):

$$h(\theta_0) + \zeta_n - h(\hat{\theta}) = 0 \tag{17}$$

and

$$h(\hat{\theta}) = h(\theta_0) + \zeta_n. \tag{18}$$

The estimation procedure selects the estimated parameters such that the expected value of the theoretical/simulated moment is equal to the true moment. If the moment function is non-linear, this condition does not imply that the expected value of the estimates is equal to the true parameter. For instance, if the moment function is convex:

$$h(E(\hat{\theta})) \leq E(h(\hat{\theta})) = h(\theta_0) \tag{19}$$

which implies that $E(\hat{\theta}) \neq \theta_0$. The direction of the bias depends then on the first derivative of the moment function. If $h'(\theta) > 0$, eq. 19 implies that $E(\hat{\theta}) \leq \theta_0$, a downward bias. On the contrary if $h'(\theta) < 0$, eq. 19 implies that $E(\hat{\theta}) \geq \theta_0$, an upward bias. In the same way it is possible to show that if the moment function is concave, the estimated parameter is upward biased if the moment function is increasing and downward biased if the moment function is decreasing.

The bias can be corrected by knowing the analytical expression of the conditional moment (which is not the case here, since we are dealing with an AB model) or reduced either by applying a monotonic transformation (a concave transformation if the moment function is convex, and viceversa) or by increasing the number of observations (*i.e.* by reducing the variance of the error).

8 Conclusions

In this paper we have identified simulated minimum distance as a natural approach to estimation of AB models. In this approach, the theoretical quantities or statistics used for characterizing the model conditional on the values of the parameters (the moments, for instance), for which no analytical expression is available in AB models, are replaced by their simulated counterparts. This requires that these statistics are appropriately chosen so that their estimates in the simulated data converge to the theoretical values. A condition is that they must be sample averages —a property

that is often not satisfied in calibration.

Moreover, because the theoretical behavior of these statistics remain unknown, it must be empirically analyzed in the simulated data. In practice, this means checking that the statistics used for estimation are stationary and ergodic, both in the real and in the simulated data (ergodicity can be tested only in the simulated data). "Test, test, test" in the simulated data is therefore the prescription when no analytical knowledge can be obtained about the behavior of the model, that is the specific prescription for AB models.

Of course, the analysis and recommendations contained in this paper must be regarded as a mere introduction to the problem of estimating AB models. They are purposely maintained at a plain and intuitive level because so little is available in the literature.

Whether our work is of some value can probably be assessed by the number of questions it raises, more than by the number of answers it gives. In this respect, we can think of many open issues and avenues for research.

First, simulated maximum likelihood should be compared with the simulated method of moments and indirect inference on a test AB model, in order to better evaluate their relative performance, in a similar way to Ruge-Murcia (2007).

Second, the estimation procedures should be tested by means of Monte Carlo simulations with the same number of observations as in the real data: for DSGE models, it is often the case that even if the model was perfectly specified, we would still need many more observations than are actually available to obtain accurate estimates (Beltran and Draper, 2008).

Third, the possibility of applying Bayesian methods should be investigated, not only in conjunction with simulated maximum likelihood but also with the other estimation procedures. As the literature on AB macro models is quite critical to the mainstream DSGE approach but at the same time has so far produced only few applications, with even less attempts at calibration (not to say estimation), one could question the need to take into consideration Bayesian priors that are necessarily based on the existing (non-AB) literature. However, including priors allows for a more general estimation procedure, and leaves open the possibility of using little informative priors.

Fourth, the relative performance of AB and DSGE models should be assessed. This has not been possible, so far, because of the lack of appropriately estimated AB models. It would then be very interesting to test whether AB models offer a better description of the real world than DSGE

models, as claimed by their proponents, and whether this conclusion depends on the specific periods considered (for instance, periods of great moderation versus periods of macroeconomic turbulence).

Fifth, one could think of DSGE and AB models as a mechanism for generating priors, rather than a model of the data. A Bayesian VAR model could then be estimated with priors informed either by a DSGE model, or by an AB model, along the lines of Del Negro and Schorfheide (2004). The question would then be, extending Sims (2006), whether the behaviorally-based priors produced by the DSGE and the AB models are helping beyond what could be achieved with uniform priors, and whether the AB model outperforms the DSGE model.

Finally, the relevance of misspecification in AB models should be the object of further investigation. Estimating different versions of the same AB model and testing the robustness of the estimates would be just one interesting exercise.

References

Alfarano, S., Wagner, T., and Lux, F. (2005). Estimation of agent-based models: The case of an asymmetric herding model. *Computational Economics*, 26:19–49.

Alfarano, S., Wagner, T., and Lux, F. (2006). Estimation of a simple agent-based model of financial markets: An application to australian stock and foreign exchange data. *Physica A*, 370(1):38–42.

Amilon, H. (2008). Estimation of an adaptive stock market model with heterogeneous agents. *Journal of Empirical Finance*, 15(2):342–362.

Anderson, P., Arrow, K., and Pines, D., editors (1988). *The Economy as an Evolving Complex System*. SFI Studies in the Sciences of Complexity. Addison-Wesley Longman, Redwood City, CA.

Arthur, W., Durlauf, S., and Lane, D., editors (1997). *The Economy as an Evolving Complex System II*. Addison-Wesley Longman.

Ballot, G. (2002). Modeling the labor market as an evolving institution: model artemis. *Journal of Economic Behavior and Organization*, 49:51–77.

Ballot, G. and Taymaz, E. (1997). The dynamics of rms in a micro-to-macro model: The role of training, learning and innovation. *Journal of Evolutionary Economics*, 7:435–457.

Ballot, G. and Taymaz, E. (1999). Technological change, learning and macro-economic coordination: An evolutionary model. *Journal of Artificial Societies and Social Simulation*, 2:article 2.

Banish, S., Lima, R., and Araújo, T. (2012). Agent based models and opinion dynamics as markov chains. Technical Report WP 10/2012/DE/UECE, School of Economics and Management, Technical University of Lisbon.

Beltran, D. O. and Draper, D. (2008). Estimating the parameters of a small open economy dsge model: Identifiability and inferential validity. International Finance Discussion Papers 955, Board of Governors of the Federal Reserve System.

Bergmann, B. R. (1974). A microsimulation of the macroeconomy with explicitely represented money flows. *Annals of Economic and Social Measurement*, 3(3):475–489.

Bergmann, B. R. (1990). Micro-to-macro simulation: a primer with a labor market example. *The Journal of Economic Perspectives*, 4:99–116.

Bianchi, C., Cirillo, P., Gallegati, M., and Vagliasindi, P. A. (2007). Validating and calibrating agent-based models: a case study. *Computational Economics*, 30:245–264.

Bianchi, C., Cirillo, P., Gallegati, M., and Vagliasindi, P. A. (2008). Validation in agent-based models: an investigation on the cats model. *Journal of Economic Behavior and Organization*, 67:947–964.

Blume, L. and Durlauf, S., editors (2006). *The Economy as an Evolving Complex System, III. Current perspectives and future directions.* Oxford University Press, Oxford, UK.

Boswick, H., Hommes, C., and S., M. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938–1970.

Caballero, R. J. (2010). Macroeconomics after the crisis: Time to deal with the pretense-of-knowledge syndrome. *Journal of Economic Perspectives*, 24:85–102.

Canova, F. and Sala, L. (2009). Back to square one: Identification issues in dsge models. *Journal of Monetary Economics*, 56:431–449.

Chen, S.-H., Chang, C.-L., and Du, Y.-R. (2012). Agent-based economic models and econometrics. *The Knowledge Engineering Review*, 27(2):187–219.

Cirillo, P. and Gallegati, M. (2012). The empirical validation of an agent-based model. *Eastern Economic Journal*, 38:525–547.

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., and Mehrling, P. (2008). Toward an empirically based macroeconomics. *The American Economic Review, Papers and Proceedings*, 98(2):236–240.

Dawid, H., Gemkow, S., Harting, P., van der Hoog, S., and Neugart, M. (2013). Agent-based macroeconomic modeling and policy analysis: The eurace@unibi model. In Chen, S.-H. and Kaboudan, M., editors, *Handbook on Computational Economics and Finance*. Oxford University Press, Oxford, UK. forthcoming.

De Marchi, S. (2005). *Computational and Mathematical Modeling in the Social Sciences, Cambridge*. Cambridge University Press, Cambridge, UK.

Del Negro, M. and Schorfheide, F. (2004). Priors from equilibrium models for vars. *International Economics Review*, 45:643–673.

Dosi, G. and Nelson, R. (1994). An introduction to evolutionary theories in economics. *Journal of Evolutionary Economics*, 4:153–172.

Eliasson, G. (1977). Competition and market processes in a simulation model of the swedish economy. *The American Economic Review*, 67:277–281.

Epstein, J. (2006). Remarks on the foundations of agent-based generative social science. In Tesfatsion, L. and Judd, K., editors, *Handbook of Computational Economics*, volume Volume 2: Agent-Based Computational Economics of *Handbook in Economics 13*. North-Holland.

Fabretti, A. (2012). On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, pages 1–17. forthcoming.

Fagiolo, G. and Roventini, A. (2012). Macroeconomic policy in dsge and agent-based models. EconomiX Working Papers 2012-17, University of Paris West - Nanterre la Dfense.

Fernández-Villaverde, J. (2010). The econometrics of dsge models. *SERIEs*, 1:3–49.

Franke, R. (2009). Applying the method of simulated moments to estimate a small agent-based asset pricing model. *Journal of Empirical Finance*, 16:804–815.

Gallegati, M. and Richiardi, M. (2009). Agent-based modelling in economics and complexity. In Meyer, R. A., editor, *Encyclopedia of Complexity and System Science*. Springer. forthcoming.

Gilli, M. and Winker, P. (2003). A global optimization heuristic for estimating agent based models. *Computational Statistics and Data Analysis*, 42(2):299–312.

Gilli, M. and Winker, P. (2008). Review of heuristic optimization methods in econometrics. *Comisef Working Paper Series*.

Gouriéroux, C. S. and Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford University Press, New York.

Gouriéroux, C. S., Monfort, A., and Renault, E. M. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S):S 85–118.

Grazzini, J. (2012). Analysis of the emergent properties: Stationarity and ergodicity. *Journal of Artificial Societies and Social Simulation*, 15(2):7.

Grazzini, J., Richiardi, M., and Sella, L. (2012a). Small sample bias in msm estimation of agent-based models. In Andrea Teglio, Simone Alfarano, E. C.-C. M. G.-V., editor, *Managing Market Complexity. The Approach of Artificial Economics.*, Lecture Notes in Economics and Mathematical Systems. Springer.

Grazzini, J., Richiardi, M. G., and Sella, L. (2012b). Indirect estimation of agent-based models. an application to a simple diffusion model. *Complexity Economics*, 1(2 (forthcoming)).

Haken, H. (1983). *Synergetics. Non-equilibrium Phase Transitions and Social Measurement.* Springer-Verlag, Berlin, DE, 3rd edition edition.

Hansen, L. P. and Heckman, J. J. (1996). The empirical foundations of calibration. *The Journal of Economic Perspectives*, 10(1):87–104.

Izquierdo, L., Izquierdo, S., Galn, J., and J.I., S. (2009). Techniques to understand computer simulations: Markov chain analysis. *Journal of Artificial Societies and Social Simulation*, 12(1):art. 6.

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov Chains*. Springer.

Kirman, A. (2010). The economic crisis is a crisis for economic theory. *CESifo Economic Studies*, 56(4):498–535.

Kocherlakota, N. R. (2007). Model fit and model selection. *Federal Reserve Bank of St. Louis Review*, 89(4):349–60.

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

Krugman, P. (2011). The profession and the crisis. *Eastern Economic Journal*, 37:303–312.

LeBaron, B. and Tesfatsion, L. (2008). Modeling macroeconomies as open-ended dynamic systems of interacting agents. *American Economic Review: Papers & Proceedings*, 98(2):246–250.

Leombruni, R. and Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A*, 355(1):103–109.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

Neugart, M. and Richiardi, M. (2013). Agent-based models of the labor market. In Chen, S.-H. and Kaboudan, M., editors, *Handbook on Computational Economics and Finance*. Oxford University Press, Oxford, UK. forthcoming.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *Handbook of econometrics*, volume IV, chapter 36. Elsevier.

Orcutt, G. H. (1957). A new type of socio-economic system. *The Review of Economics and Statistics*, 39:116–123.

Orcutt, G. H. (1961). *Microanalysis of socioeconomic systems: A simulation study*. Harper.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 5:1027–1057.

Poirier, D. J. (1998). Revising beliefs in nonidentified models. *Econometric Theory*, 14(4):483–509.

Prigogine, I. and Stengers, I. (1984). *Order out of Chaos: Man's New Dialogue with Nature*. Bantam Books, New York, NY.

Richiardi, M. (2012). Agent-based computational economics: a short introduction. *The Knowledge Engineering Review*, 27(2):137–149.

Ruge-Murcia, F. J. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31:2599–2636.

Schelling, T. (1969). Models of segregation. *American Economic Review*, 59(2):488–493.

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

Sims, C. A. (2006). Comment on del negro, schorfheide, smets and wouters. http://sims.princeton.edu/yftp/DSSW806/DSseattleComment.pdf.

Solow, R. (2010). Hearing before the subcommittee on investigations and oversight committee on science and technology. House of Representatives 111th Congress.

Stern, S. (1997). Simulation based estimation. *Journal of Economic Literature*, 35(4):2006–2039.

Stern, S. (2000). Simulation-based inference in econometrics: motivation and methods. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-based inference in econometrics: methods and applications*. Cambridge University Press.

Stiglitz, J. E. (2011). Rethinking macroeconomics: What failed, and how to repair it. *Journal of the European Economic Association*, 9:591–645.

Tovar, C. E. (2008). Dsge models and central banks. BIS Working Papers 258, Bank for International Settlements.

Trichet, J.-C. (2010). Reflections on the nature of monetary policy non-standard measures and finance theory. Speech at the Opening address at the ECB Central Banking Conference, Frankfurt, 18 November 2010. Available online at `http://www.ecb.int/press/key/date/2010/html/sp101118.en.html`.

Watts, H. W. (1991). Distinguished fellow: an appreciation of guy orcutt. *The Journal of Economic Perspectives*, 5:171–179.

Weeks, M. (1995). Circumventing the curse of dimensionality in applied work using computer intensive methods. *The Economic Journal*, 105(429):520–530.

Winker, P. and Gilli, M. (2001). Indirect estimation of the parameters of agent based models of financial markets. Technical Report 03/2001, School of Business Administration, International University in Germany, Bruchsal.

Winker, P., Gilli, M., and Jeleskovic, V. (2007). An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination*, 2:125–145.