Estimation of Ergodic Agent-Based Models by Simulated Minimum Distance

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

[a]Catholic University of Milan, Department of Economics and Finance, Milano

[b]Institute for New Economic Thinking and Nuffield College, Oxford

[c]Collegio Carlo Alberto and LABORatorio Revelli, Moncalieri

October 13, 2014

Abstract

Two difficulties arise in the estimation of AB models: (i) the criterion function has no simple analytical expression, (ii) the aggregate properties of the model cannot be analytically understood. In this paper we show how to circumvent these difficulties and under which conditions ergodic models can be consistently estimated by simulated minimum distance techniques, both in a long-run equilibrium and during an adjustment phase.

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

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

*Email: `jakob.grazzini@unicatt.it`

†Email: `matteo.richiardi@maths.ox.ac.uk`

1 Introduction

Agent-based (AB) models are sometimes considered as a candidate to replace or at least complement dynamic stochastic general equilibrium (DSGE) as the standard tool for macroeconomic analysis. 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; this resembles the state of the art in DSGE modeling a few years ago, which has now moved forward toward more formal estimation.[1] Estimation is crucial for the empirical validation of a model, for comparing a model with other available models, and for policy analysis. Also, estimation (as opposed to calibration) involves the attainment of clearly specified scientific standards in the way models are confronted with the data, and this sort of "certification" is highly needed to gain confidence and ultimately support from the wider scientific community and the policy circles in the use of a new tool.

The advantage of AB modeling over more conventional techniques ultimately lies in the increased flexibility in model specification, but this flexibility comes at the price of making estimation more difficult. Paradoxically, AB models can provide a more realistic description of how real economies actually work, but this description resists fine tuning to real data. However, fine-tuning is important as small differences in the values of the parameters might be amplified by the nonlinear feedbacks in the system and result in large differences in the effects of the policies under examination. Moreover, AB models are often conceived to investigate what happens "out-of-equilibrium", during adjustment phases: systemic disruptions, regime shifts, rare events are usual objects of analysis. How to characterize the behavior of a model in such circumstances, beyond a merely qualitative description, is one open issue that we directly address.[2] We then turn to some basic 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 answers to these questions depend in part on the structure of the models, and in part on the specific econometric technique employed. We spell out the general conditions under which

[1] Though estimation of DSGE models arouses a number of serious issues –see section 3 below.

[2] Irrespective of the estimation strategy, whether the data contain enough information to allow inference in these settings is another issue about which we have little to say.

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 (see for instance 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". Because AB models are composed of 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 of AB models is the same as in DSGE models, but in the latter some aggregate properties can be analytically understood or derived from first principles and primitive assumptions –in particular, whether the system is ergodic and stationary.

We see three merits in our work. First, it provides a formalization of AB models and 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.[3] 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.

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 estimating 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 give two examples of estimation of ergodic AB models: in the first, estimation is

[3] In this paper we focus on estimation of ergodic models. The extension to the case of non-ergodic models is discussed in a companion paper (Grazzini and Richiardi, 2014).

performed in a long-run equilibrium, while in the second estimation is performed on an adjustment phase. Section 7 concludes.

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 phenomenon (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. Little data is involved with such a line of research.

Recent years have witnessed a trend in AB macro modeling towards more detailed and richer models, targeting at a higher number of stylized macro facts, and claiming a stronger empirical content.[4] 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 empirical relevance is the ultimate critique that has been moved to DSGE models.[5] 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).[6]

In terms of the terminology of Chen et al. (2012), the AB camp has to move from stage I (the

[4]See eg. Delli Gatti et al. (2005, 2011); Dosi et al. (2010, 2013) –see the reviews in Dawid et al. (2014) and Neugart and Richiardi (2014).

[5]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".

[6]Marks (2013) discusses the use of three measures of similarity to compare simulated and real time series.

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.

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.[7] However, when the models have 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 to estimate models of exchange rate. Amilon (2008) employs the efficient method of moments (and maximum likelihood on a simpler model) to 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. This is because of 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).

AB models are however not the only models possibly plagued by this problem. 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.

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

3 Estimation of DSGE models

In DSGE models aggregation is generally not a problem, thanks to a very low level of heterogeneity.[8] A common strategy to solve the models involves the linearization of first order conditions and constraints by means of a Taylor approximation around the steady state.[9] In presence of non-stationarity, the model is rescaled and detrended, in order to make it stationary around a balanced growth path; the trend is often assumed to follow an exogenous stochastic process.

Turning to 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.[10] The literature on the econometrics of DSGE is vast and increasing.[11] 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.

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, requires a number of steps (Fernández-Villaverde, 2010). First, the solutions

[8] In standard models as Christiano et al. (2005) or Smets and Wouters (2003) the problem of aggregation is circumvented by assuming a representative firm for the perfectly competitive final good sector and a monopolistic producer or monopolistic competition in the intermediate goods market; households are assumed to be homogeneous with respect to consumption and asset holdings, though heterogeneity in hours worked and wages is allowed: each household is a monopolistic supplier of a differentiated labor service, and labor supply decisions are aggregated by using some exogenous technology, generally of the Dixit-Stiglitz type. Attempts to introduce non-trivial heterogeneity among households, for instance with respect to productivity, include Chang et al. (2013); Chang and Kim (2007); Krusell and Smith (2006, 1998), but still rely on the assumption of a representative firm. See Kirman (1992) and Forni and Lippi (1997), among others, for a critique of the representative agent assumption, and Fagiolo and Roventini (2012) for a discussion of its role in DSGE models. Some exact aggregation results hold (eg. Magnani and Mercenier, 2009) but they apply to specific settings only. Similarly, heterogeneity in pricing decisions by firms is modeled according to a variant of the Calvo (1983) or Taylor (1980) mechanism (see Dixon and Bihan, 2012), which allows a solution without explicitly tracking the distribution of prices across firms.

[9] 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).

[10] An intermediate approach is to calibrate some parameters, and then estimate the others conditional on the values of the calibrated set. This raises serious identification issues (see subsection 5.2.2 below).

[11] See Canova (2007), Tovar (2008) and Fernández-Villaverde (2010) for an introduction.

for the steady state values of the observables (the *policy functions*) must 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.

More specifically, if the state space representation is linear and the shocks are assumed to be normally distributed, a simple Kalman filter (KF) can be applied to obtain the likelihood function and perform exact inference. Alternatively, if the state space is nonlinear an extended Kalman filter (XKF) can be implemented on a linearized system around the steady state; this provides sub-optimal inference, but still relies on the Gaussian assumption. 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 filters –see below): 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. The first one is *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.[12] 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.[13] In turn, the flatness of the likelihood

[12] 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.

[13] More in general, DSGE models are likely to lead to ill-behaved objective functions, irrespective of the method

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 due to a limited ability to reproduce the data, possibly because of a 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 "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 lunch bag 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-identified "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 policy functions have no analytical formulation (more on this in the next section). Moreover, AB models are inherently nonlinear[14] and likely non-Gaussian[15]. Given that no equilibrium equations can be written, linearization around the steady state turns out to be problematic.[16] Moreover, the dynamics need not to be smooth, which might impinge on the existence of the Jacobian.

In principle, one could use more sophisticated filters to go around this problem. Since their introduction in the early 1990s (Gordon et al., 1993), particle filters –also known as sequential Monte Carlo methods– have become a popular solution of optimal estimation problems in nonlinear non-Gaussian scenarios. In comparison with standard methods such as the KF or XKF, the principal advantage of particle methods is that they do not rely on any local linearization technique or any

used for estimation –see section 5.2.2 on identification issues.

[14]Even if linearity is assumed at a micro level, heterogeneity disrupts it, as the sum of linear functions is in general *not* linear.

[15]As a referee pointed us, the fact that AB models generally produce non-Gaussian distributions is a positive feature of such models, as the distribution of many macro series are typically non-Gaussian –eg. growth rates (Fagiolo et al., 2008)– and DSGE models are not able to reproduce such regularities even when they are fed with shocks drawn from fat tail distributions (Ascari et al., 2014).

[16]A simplified *meta-model* could be imposed on the steady state behavior of the system. This naturally leads to estimation by II (see below).

crude functional approximation. The price that must be paid for this flexibility is computational: these methods are computationally expensive. The starting point for particle filters is choosing an ensemble of state estimates that captures the initial probability distribution of the state. These sample points are then propagated through the true nonlinear system and the probability density function of the actual state is approximated by the ensemble of the estimates. In many cases however, the number of sample points required is of the same order as the dimension of the system. As a consequence, particle filters in their basic form do not work in high dimensions.

Particle filtering has found only a limited application to DSGE modeling (Amisano and Tristani, 2010; Andreasen, 2011; Fasolo, 2012), while no applications exist to AB models.[17] Alternatively, 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 Simulated minimum distance

The main alternatives to ML are GMM, MSM and II (or extended method of simulated moments, EMSM).[18] MSM and II 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

[17] Though this is an open area for future research.

[18] 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.

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 the parameters). Often, a vector autoregression (VAR) is used as meta-model. In this case there exists a straightforward interpretation in terms of moments, 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.[19]

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 misspecfication 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 (and II) as a natural estimation technique for AB models. Before turning to analyzing the conditions for consistency in SMD estimators, 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

[19] Indirect inference 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 model, and the parameters are changed until the distance between the coefficients of the meta-model estimated on simulated data and those estimated in the literature is minimized.

simulation model without having to go through a process of solution" (Bergmann, 1990).[20]

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[21] may be individual-specific 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.[22] Quite interestingly, this is the same *transition equation* that is used for describing 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 $\boldsymbol{Y}$:

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

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

[21] 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.*

[22] 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.

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.[23] As the behavioral rules f_i and the transformation function $\boldsymbol{G}$ need not to be linear, these random terms cannot be netted out by taking expectations. Therefore, the only way to analyze the mapping of $(\boldsymbol{X}_0, \boldsymbol{\theta})$ into $\boldsymbol{Y}_t$ is by means of Monte Carlo analysis, by simulating the model for different initial states and values of the parameters, and repeating each simulation experiment many times to obtain a distribution of $\boldsymbol{Y}_t$.

However, because digital computers are deterministic machines, it is possible to further pin down the formalization above. 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 $\boldsymbol{\xi}_t$ and $\boldsymbol{\kappa}_t$. By iteratively substituting $\boldsymbol{X}_{t+1}$ with $\boldsymbol{X}_t$, we get

$$\begin{aligned} \boldsymbol{X}_t &= \boldsymbol{F}(\boldsymbol{F}(\cdots \boldsymbol{F}(\boldsymbol{Z}_0, \boldsymbol{\theta}) \cdots)) \\ &= \boldsymbol{F}^t(\boldsymbol{Z}_0, \boldsymbol{\theta}) \qquad (2'') \\ \boldsymbol{Y}_t &= \boldsymbol{G}\big(\boldsymbol{F}^t(\boldsymbol{Z}_0, \boldsymbol{\theta})\big). \qquad (3'') \end{aligned}$$

The law of motion (**??**$''$) uniquely relates the value of $\boldsymbol{Y}$ at any time t to the initial conditions of the system and to the values of the parameters, and is known as the *input-output transformation* (IOT) function.

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

[23] 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 combined matrix of stochastic terms. This is because the variables in excess are a deterministic function of the other observables.

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.

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 $\boldsymbol{\theta}$.[24]

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 $[\underline{T}, \underline{T} + \tau], \tau \to \infty$.

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

By definition, for any given value of the initial conditions (including the random seeds) 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.

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, if the expected value and the variance exist and are finite, the simulated mean –which we label $m_t(Y_a^*, \boldsymbol{\theta})$– converges, both over time and over the replications r, to the theoretical limiting expected value of the underlying DGP conditional on the parameters used for the simulation, $\mu(Y_a^*, \boldsymbol{\theta})$:[25]

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

[24] 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.

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

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 simulated mean of Y, computed either over time or over replications, is a consistent estimator of the theoretical limiting expected value, 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 $m_r(Y^*_j)$ 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 $t \in [\underline{T}_j, \bar{T}_j]$.

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.[26]

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 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 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

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

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 frequency, length and depth of recessions, or an attractor in an appropriately defined state space. 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 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.

5 Estimation of AB models

In the following 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.[27] Under correct specification, the model defines the behavior of the system over the whole set of parameters $\boldsymbol{\theta}$ while the observations are an instance of the system with the *true* set of parameters, $\boldsymbol{\theta}_0$. Given the model and a set of observations, any estimation procedure is based on the comparison of the properties of the model and the properties of the data. In this respect, one important difference between AB models and DSGE models lies in the richer outcomes that the former are able to generate (and that can then be confronted with the real data), that is, in a wider set of statistics $\boldsymbol{Y}_t$. In particular, these statistics can include cross-sectional features of a variable of interest, as the second or higher moments or a parametric characterization of its distribution (eg. the Pareto exponent). On the other hand, 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.

Analytical knowledge of the properties of the model generally implies that ergodicity and stationarity (of the model) can be assessed at a theoretical level.[28] 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, in an AB model knowledge about the statistical properties of the equilibria can only be inferred from the analysis of the simulation data. Consequently, 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

[27] 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.

[28] Use of expected utility theory (EUT) often buys ergodicity, as EUT uses ensemble averages to evaluate uncertain events –see Grazzini and Richiardi (2014).

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 (Stern, 1997, 2000). The second problem requires a deeper analysis of the simulated data.

As we have anticipated in section 3, possibly the most promising techniques for estimating AB models are MSM and II. We now briefly recall the conditions for consistency in minimum distance estimators, and discuss how these conditions apply to simulated minimum distance estimation of AB models.

5.1 Consistency in minimum distance and simulated 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:

$$\hat{Q}_n(\boldsymbol{\theta}) = -\big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big)'\hat{W}_n\big(\hat{\boldsymbol{h}}_n - \boldsymbol{h}(\boldsymbol{\theta})\big) \tag{6}$$

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

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \hat{Q}_n(\boldsymbol{\theta}). \tag{7}$$

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 $\boldsymbol{h}_0 = \boldsymbol{h}(\boldsymbol{\theta}_0)$ as the *true* value of the estimators. For this value, assuming W is a semi-

positive matrix, the objective function then becomes:

$$Q_0(\boldsymbol{\theta}) = -\big(\boldsymbol{h}_0 - \boldsymbol{h}(\boldsymbol{\theta})\big)' W \big(\boldsymbol{h}_0 - \boldsymbol{h}(\boldsymbol{\theta})\big). \tag{8}$$

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, $\boldsymbol{\theta}_0$.[29] 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 (7), 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 $\boldsymbol{\theta}_0$.[30] The conditions above should hold (and, whenever possible, checked) in the real data.

As we have already discussed, 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. (8) 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

[29] As such, consistency can be evaluated with respect to different dimensions: *consistency in size* means that the estimates converge to their true value as the observed population grows bigger; *consistency in time* means that the estimates converge to their true value as the length of the observation period increases; *consistency in replications* means that the estimates converge to their true value as more occurrences of the same stochastic process are observed. With reference to an abstract cube of observations, consistency in size (cross-sectional) refers to the height, consistency in time (longitudinal) refers to the length, consistency in replications refers to the width of the cube.

[30] 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 –see Newey and McFadden (1994, p. 2122).

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 $\tilde{\boldsymbol{h}}_s = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)$. The objective function then becomes:

$$\hat{Q}_n(\boldsymbol{\theta}) = -\big(\boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n) - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)\big)' \hat{W}_n \big(\boldsymbol{h}(\boldsymbol{\theta}_0, \boldsymbol{\varepsilon}_n) - \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_s)\big) \tag{9}$$

The simulation method introduces additional variability into the objective function since the theoretical properties are estimated using a finite number of simulation data.[31] If 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. (9). We now turn to investigate under which conditions an AB model can be consistently estimated by SMD.

5.2 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

[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 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).

objective function.[33] In the following it is assumed that $\boldsymbol{\Theta}$ is a discrete set, and that $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

5.2.1 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 standard, 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.

Each statistics used in the objective function should therefore be tested for stationarity in the simulated data.[34] Given that the distributions of these statistics are *a priori* unknown, a nonparametric test should be preferred (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).[35] 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. Nonparametric tests for stationarity and ergodicity of the time series produced by an AB model are proposed in Grazzini (2012); they rest on an application of the Wald-Wolfowitz runs test for testing whether two samples come from the same population (Wald and Wolfowitz, 1940). The tests are used to assess whether the statistics of interest are constant in time and across runs: the stationarity test uses samples from a given simulation run, while the ergodicity test uses samples from different runs. When an AB model is represented as a Markov process, the tests described in Domowitz and El-Gamal (1993) and Domowitz and El-Gamal (2001) can also be applied.

If each moment in the objective function is ergodic and stationary than the ergodic theorem

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

[34] 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.

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

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

Note that we are talking here about convergence *in time*. Choosing appropriate *longitudinal* moments of stationary and ergodic series produces estimates which are consistent in time: as the observation period increases, both in the real and in the simulated data, the estimates converge to their true value. On the other hand, choosing cross-sectional moments at a specific point in time does not guarantee consistency in population *size*. This is because the cross-sectional moment can display autocorrelation, which is not wiped out by increasing population size. As an example, think of the unemployment rate as a function on an underlying parameter θ. Even in an absorbing equilibrium, where the unemployment rate is stationary, it is likely that the unemployment rate u_t is correlated with u_{t-1}.[36] If we increase the sample size of the simulated population our estimate of the theoretical unemployment rate $u(\theta)$ will become more and more precise, but the value of θ matching the observed unemployment rate will converge to something different from the unobserved parameter of the true DGP. This is a crucial point which distinguishes estimation from many calibration exercises, where cross-sectional statistics are used to characterize a model.[37]

5.2.2 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. Canova and Sala (2009) cite four common identification problems in DSGE models: observational equivalence (which occurs whenever

[36] Recall that stationarity only requires that the autocorrelation structure between u_t and u_{t-h} depends on h and not on t.

[37] For example, the distribution of firm size computed at a specific point in time.

the objective function does not display a unique minimum), under-identification (when the objective function is independent of certain structural parameters), partial identification (when some parameters enter the objective function only as a linear combination of other parameters), weak identification (when the objective function has a unique minimum, but its curvature is very flat at the minimum). As they argue, "[t]hese problems turn out to be inherent to DSGE models –they are typically produced because the mapping from the structural parameters to the coefficients of the solution is ill-behaved– and standard choices of data summaries and of objective functions may make identification deficiencies worse" (p. 432). Identification can be expected to be somewhat less of an issue in AB models, as more flexibility in model specification might result in better specifications but, given the small empirical literature, the relevance of the problem has still to be assessed.

Estimation in the face of severe identification challenges requires a lot of prudence and awareness. For instance, a common practice when dealing with identification issues is to reduce the number of parameters to be estimated (for instance by dropping observationally equivalent parameters from the specification) or calibrate some of the parameters and estimate the others, conditional on the calibrated values. More specifically, one could 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 and fix the other parameters to "reasonable" values (after all, they have little influence on the outcome). This strategy however works only if the calibrated parameters are under-identified; in the more general case it produces a bias in the estimated parameters, unless the calibrated ones are set to their "true" value (Canova and Sala, 2009; Canova, 2009). Canova and Sala (2009) offer four suggestions, in addition to restraining from the mixed calibration-estimation approach described above: i) explore the properties of the objective function before estimation; ii) perform small sample Monte Carlo exercises, iii) include in the objective function as many implications from the model as possible, and avoid the improper use of prior restrictions, iv) use estimation methods which are robust to identification problems, or rethink the model specification and parameterization. These are also valid for AB models. In particular, if the number of parameters is such that it is computationally feasible to compute each moment for each combination of the parameters (conveniently discretized),

it is possible to draw the objective function and visually check whether it is well-behaved. Only in some special cases analytical results can be invoked, as for the following 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 $\boldsymbol{h}(\boldsymbol{\theta}_0) = \boldsymbol{h}_0$. Monotonicity implies that the only point in the parameter space in which all $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{h}_0$ is $\boldsymbol{\theta}_0$. Finally, if the moment conditions are linearly independent functions, then $Q_0(\boldsymbol{\theta})$ is null only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

5.3 Small-sample estimation bias

If the moments are a nonlinear function of the parameters, we get a bias in the estimates. 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

$$h(\theta_0, \zeta_n) = h(\theta_0) + \zeta_n \tag{10}$$

The observed moment is the "true moment" plus an error ζ_n which depends on the sample (and on sample size). Suppose that the underlying process is ergodic and stationary and that $E(\zeta_n) = 0$: the sample moment is an unbiased and consistent estimator of the theoretical moment. The method of simulated moments minimizes the distance between the observed moment and the simulated moment with respect to the parameter:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \ \hat{Q}_n(\theta) = -(h(\theta_0, \zeta_n) - h(\theta))^2 \tag{11}$$

where $\hat{\theta}$ is the estimated parameter and $h(\theta)$ is the theoretical (simulated) moment computed using

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

the model.[39] If the moment function and the parameter are continuous, the maximization condition is:[40]

$$h(\theta_0, \zeta_n) - h(\hat{\theta}) = 0. \tag{13}$$

Using equation (10):

$$h(\theta_0) + \zeta_n - h(\hat{\theta}) = 0 \tag{14}$$

and

$$h(\hat{\theta}) = h(\theta_0) + \zeta_n. \tag{15}$$

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:

$$h(E(\hat{\theta})) \leq E(h(\hat{\theta})) = h(\theta_0) \tag{16}$$

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. 16 implies that $E(\hat{\theta}) \leq \theta_0$, a downward bias. On the contrary if $h'(\theta) < 0$, eq. 16 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).

[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:

$$\frac{d\hat{Q}_n(\theta)}{d\theta} = 2\big(h(\theta_0, \zeta_n) - h(\theta)\big)\frac{dh(\theta)}{d\theta} = 0 \tag{12}$$

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 13.

5.4 Summary: estimation of ergodic models

As we have seen, estimating an ergodic AB model by means of simulated minimum distance is conceptually straightforward, the crucial step being the choice of appropriate *longitudinal* statistics (longitudinal "moments") to match simulated and real data. When the data are distributional characteristics of a variable defined at the micro level (eg. income, or firm size), as some cross-sectional moment (eg. the variance) or the parameters of a given distributional form (eg. the Pareto exponent), this also involves looking at their longitudinal moments (eg. the average over time of the variance of income, or the average over time of the Pareto exponent of the estimated firms size distribution).

In order for these longitudinal moments to be meaningful quantities, the time series considered (both real and simulated) must be stationary: (weak) stationarity implies, in loose terms, that the theoretical mean and variance are constant in time. Stationarity can be a permanent property of the time series, that is, once it becomes stationary, the time series remains stationary. This we call an absorbing equilibrium of the model, with respect to the time series considered. Or, stationarity can be a temporary property, which breaks down at some point. This we call a transient equilibrium of the model. What is important is that estimation is never performed outside those equilibria, or mixing data coming from different equilibria.

Because the theoretical moments cannot be analytically derived, we proceed by simulating them: stationarity (and ergodicity) ensure that the empirical means computed on the artificial data converge to the theoretical ones. However, for finite (simulated) horizons estimating the theoretical moments in the artificial data introduces an additional source of uncertainty. By producing more simulation data (longer time series in an absorbing equilibrium, more replications in a transient equilibrium)[41], this sampling error can be reduced to an arbitrarily small number that tends to zero, in which case the performance of the MSM estimator equals that of the corresponding analytical moment estimator: the only uncertainty that is left in the inference exercise is that relating to the real data, for which the observation period is fixed.

The observed moments are then constructed by averaging over the available real data. Should

[41] To be more precise, if the model is ergodic (with respect to the specific time series considered), estimating the theoretical moments $\boldsymbol{\mu}(\boldsymbol{\theta})$ in an absorbing equilibrium by averaging over one long simulated time series or over many shorter time series produced by different simulation runs is the same, provided the series has reached the stationary state.

the real sample also grow larger (again, this means longer time series in an absorbing equilibrium and more replications of the process under investigation in a transient equilibrium), that uncertainty would evaporate too, and the estimates would converge to the true value of the coefficients: the estimator is consistent in *time* when estimation is performed in an absorbing equilibrium, and consistent in *replications* when performed in a transient equilibrium.

The value of $\boldsymbol{\theta}$ for which the theoretical moments are closer to the observed ones is our estimate. Confidence intervals can then be constructed by bootstrapping the real data.

6 Examples

In this section we present two examples of estimation of ergodic AB models by SMD: the first model is estimated in an absorbing equilibrium, while the second is estimated in a transient equilibrium. In both cases, because the moments used for estimation are non-linear, we get a small-sample bias (of predictable direction). All estimation strategies are explored by means of Monte Carlo experiments: pseudo-true data are created by running the model with some chosen value of the parameters; then, estimates are obtained by matching the simulated moments to the moments computed on the pseudo-true data. This procedure is repeated many times in order to get a distribution of the estimates.

6.1 Estimation in an absorbing equilibrium

The model used here is an AB stock market model proposed in Cliff and Bruten (1997) to reproduce the experimental results obtained by Smith (1962), showing how a small number of inexperienced traders converge rapidly to a competitive equilibrium under the double auction mechanism (Smith, 1962, p.157). There is an order book where traders can ask or bid. Traders only observe the book and know nothing about the demand and supply schedule; they cannot lend or borrow. The limit price of agent i in period t is

$$p_i(t) = v_i(1 + \mu_i(t)) \tag{17}$$

where v_i is the (constant) value of the traded asset for agent i and $\mu_i(t)$ is a profit margin (positive for sellers and negative for buyers). In each period traders look at the book and define a target price $\tau_i(t)$: traders increase their target price if the last trade occurred at a high price, and lower

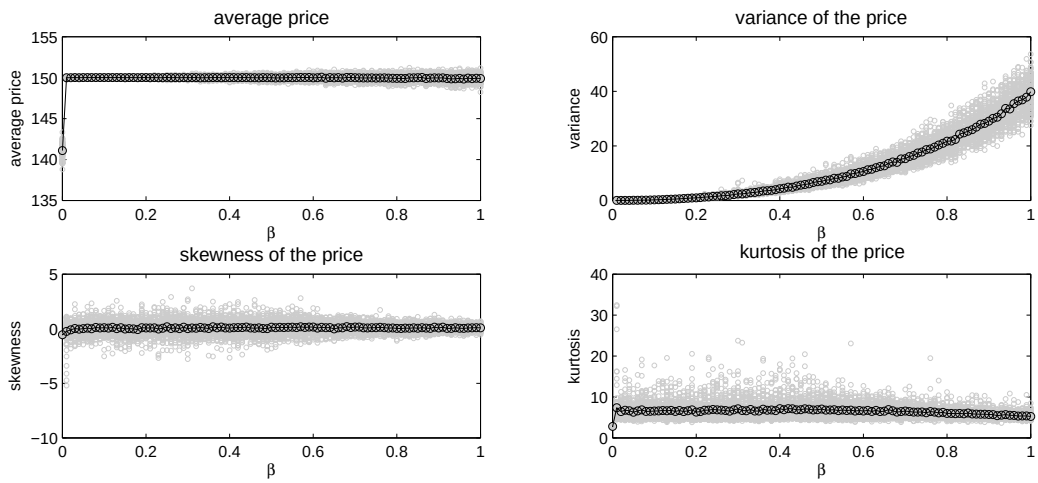

Figure 1: Model behavior in the absorbing equilibrium (stationary price). β is sampled in $[0, 1]$ at intervals of 0.01. 100 runs with different seeds are performed for each value of β (the seeds are kept constant while varying the value of the parameter). Each run lasts for 400 trading days after the absorbing equilibrium has been reached. The gray dots are the moments computed in each simulation; the black dots mark the averages, conditional on β.

it otherwise –see Cliff and Bruten (1997) for details. The actual change in the limit price is:

$$\Delta_i(t) = \beta_i(\tau_i(t) - p_i(t)) \tag{18}$$

where β_i is a behavioral parameter determining how the traders react to a difference between the target price and the current price $p_i(t)$. The profit margin is updated as follows:

$$\mu_i(t+1) = \frac{p_i(t) + \Delta_i(t)}{v_i} - 1 \tag{19}$$

and determines the updated limit price, $p_i(t+1)$. The learning process depends on the interaction between the behaviors of the agents in the market (through the orders) and on β_i, which determines the strength of the interaction. We wish to estimate the learning parameter from the aggregate outcome of the model. Ideally β_i is heterogeneous. To simplify the estimation procedure, we assume all traders share the same parameter $\beta_i = \beta$. We consider 11 buyers and 11 sellers.

Figure 1 summarizes the behavior of the model with respect to β. The theoretical equilibrium is the same irrespective of β and depends on the distribution of the fundamental values each trader attributes to the asset. The behavior around the equilibrium price, on the other hand, depends on β. The only moment that seems to discriminate between different values of the learning parameters

Figure 2: Standard deviation of the price. β is sampled in $[0, 1]$ at intervals of 0.01. 100 runs with different seeds are performed for each value of β (the seeds are kept constant while varying the value of the parameter). Each run lasts for 400 trading days after the absorbing equilibrium has been reached. The gray dots are the moments computed in each simulation; the black dots mark the averages, conditional on β.

is the variance.[42] To asses whether the variance can be used to consistently estimate the model, we need to understand its statistical properties, and in particular test whether it is stationary and ergodic. To this aim, we use the tests described in Grazzini (2012) (see section 5.2.1): the results indeed show that the variance is stationary and ergodic. However, the variance is convex in the parameter: this introduces a downward bias in the estimates. To reduce the bias, we linearize the moment applying a concave transformation. It turns out that the square root of the variance (that is, the standard deviation) is sufficiently linear (figure 2).

The pseudo-true data are generated with $\beta = 0.55$. The theoretical moments are computed by sampling all discrete values of β in [0,1] at intervals of 0.01, averaging over 100 runs each lasting for 400 trading days after the absorbing equilibrium is reached. Optimization is achieved by brute force, that is by computing the value of the objective function at each value of the parameter. Figure 3 shows the shape of the objective function, for one particular Monte Carlo experiment (pseudo-true series) using the variance (left panel) and the standard deviation (right panel).

Figure 4 shows the results of the Monte Carlo experiments, based on 8,000 different pseudo-true series. The average estimated parameter (standard deviation) is 0.5472 (0.00163) using the

[42]The model has a very different behavior with $\beta = 0$, with a much higher variance. As in a real application the no-learning case would be easily detectable, we focus on $\beta > 0$.

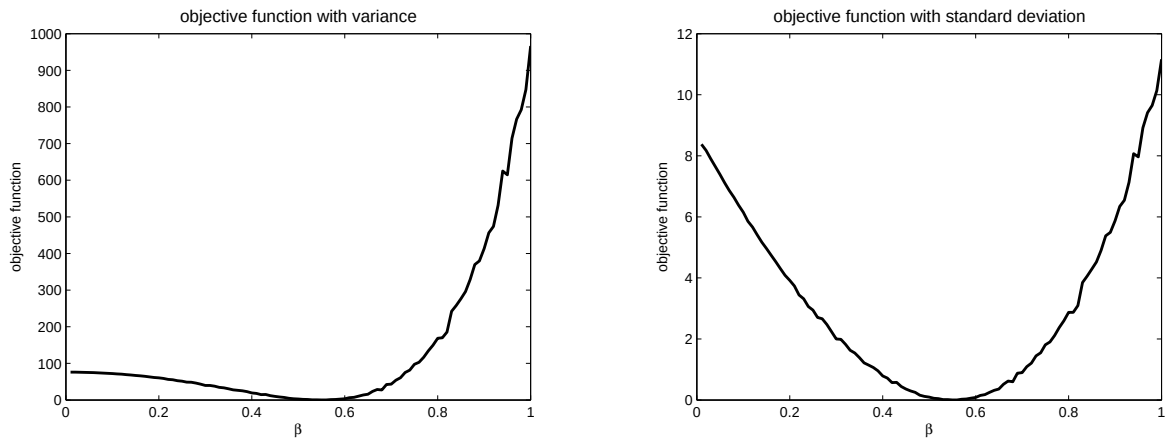

Figure 3: Objective function, constructed over different moments. Left: variance. Right: standard deviation. The theoretical moments are computed, for each value of the parameter, averaging over 100 runs each lasting for 400 trading days after the absorbing equilibrium is reached.

variance of the price as target and 0.5489 (0.00163 again) using the standard deviation of the price as target. As predicted, we get a downward bias, which is reduced when employing the standard deviation rather than the variance in the moment estimator. Also, the bias decreases as the sample size (trading days) increases (results not shown).

6.2 Estimation in a transient equilibrium

As an example of structural estimation of an AB model in a transient equilibrium, we take our earlier work (Grazzini et al., 2012b). In that paper we estimate an AB version of the well known Bass (1969) model, which provides a mathematical explanation of the different stages of product adoption described in Rogers (1962) (innovators, early adopters, early majority, large majority, and laggards), and formalizes the crucial distinction between innovators and imitators. The model is an example of early epidemic models of innovation diffusion (Geroski, 2000). It consists of a differential equation that specifies the rate of adoption $h(t)$ as a function of an external force and an internal (endogenous) one. The external influence is constant over time and represents the effects of advertisement, while the internal influence depends on how many others have already adopted at time t and formalizes word-of-mouth:

$$h(t) = p + qF(t) \tag{20}$$

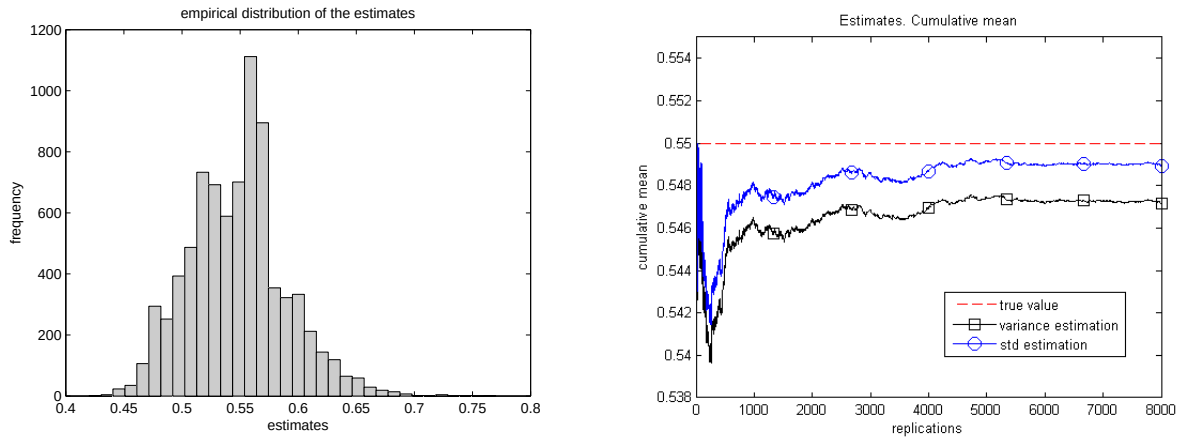

Figure 4: Monte Carlo results, 8,000 experiments (pseudo-true series). The theoretical moments are computed, for each value of the parameter, averaging over 100 runs each lasting 400 trading days after the absorbing equilibrium is reached. Left: empirical distribution of the estimates (the correct value is 0.55). Right: Cumulative mean of the estimates.

where $F(t) = N(t)/m$ is the c.d.f. of adopters, that is the ratio of those who have already adopted $N(t)$ over the number of potential adopters m. p is the parameter for external influence and q is the parameter for internal influence, with $p + q < 1$. Everybody starts as a non-adopter. Our AB version simply considers a discrete version of eq. (20), which applies either when the adoption process is intrinsically discrete, or when the process is indeed continuous but information on cumulative adoption becomes available only at discrete time intervals.

The model is clearly ergodic for $p > 0$ (given enough time, everybody eventually adopts). However, the unique absorbing equilibrium is invariant to the value of the parameters and hence contains no information that can be exploited for estimation. We therefore focus on the adjustment process, looking for regularities which define the transient equilibria of the system.

Our estimation approach involves a combination of analytical and simulation-based techniques, in a three-stage procedure. First, we get an estimate $\hat{p}$ for the external influence parameter, conditional on a specific value of the population parameter m, by matching the observed and simulated adoption rate at $t = 0$. Second, we get an estimate $\hat{q}$ for the internal influence parameter q, conditional on m and $\hat{p}$, by matching the observed and simulated adoption rate of those who have not yet adopted (the population at risk) at $t > 0$.[43] Third, we implement a MSM estimator of

[43]Our estimation strategy for the parameters p and q is equivalent to the standard ML approach to estimating discrete-time survival models, where adoption is modeled as a binary dependent variable conditional on being still at risk (Jenkins, 1997; Stata 13 manual, 2013).

m by iterating the first two steps with different values of the market potential parameter, until we find the set of estimates $(\hat{m}, \hat{p}, \hat{q})$ which minimize the distance between the observed and simulated moment (the mean adoption time for those who have adopted up to a given observation period T). Our three-stage estimator converges to the true values of the parameters as the population of potential adopters m increases.

More specifically, in every period we observe a number of adoptions which is a random draw from a binomial distribution $B(m - N_t, h_t)$, though the parameters of this distribution –the population at risk $m - N_t$ and the probability of adoption $h_t = p + qN_t/m$– change over time. The natural (ML) estimator for h_t is therefore:

$$\hat{h}_t = \frac{n_t}{m - N_{t-1}} \tag{21}$$

which is unbiased, consistent and asymptotically normal for large populations m. Given that $N_0 = 0$, data on the first period of introduction of the new product can be used to estimate p:

$$\hat{p} = \frac{N_1}{m}. \tag{22}$$

This estimator is also unbiased, consistent and asymptotically normal for large populations m. We can then exploit knowledge of $\hat{p}$ to estimate q. By plugging $\hat{p}$ in eq. 21 we obtain:

$$\hat{q}(t) = \frac{m}{N_t}\left[\frac{n_t}{m - N_{t-1}} - \frac{N_1}{m}\right]. \tag{23}$$

In every period after the first one, an estimate $\hat{q}_t$ can thus be obtained, which is also unbiased, consistent and asymptotically normal, for large populations m. As our final estimator of q we take a weighted average of $\hat{q}_t$, using the inverse of the variance of $\hat{q}_t$ as weights.

Finally, to estimate the population parameter m, we minimize the distance between the real (τ_r) and simulated (τ_s) mean adoption time for those who have adopted in the observation period, where the simulated moment is obtained by simulating the adoption time of m individuals with $h_t = \hat{p}(m) + \hat{q}(m)N_{t-1}/m$:

$$\hat{m} = argmin_m[\tau_r - \tau_s(m)]^2. \tag{24}$$

The procedure is iterated until convergence. The estimates are consistent in replications, that is, by increasing the number of observed adoption processes (assuming the parameters remain constant) the estimates converge to their true value. However, because the moment τ is not linear in m (figure 5), we get an upward small sample bias. This is confirmed by a Monte Carlo experiment with pseudo-true values of $m = 10,000; p = .03; q = .4$, where we get average values (standard deviations) of the estimates, over 1,000 replications of the procedure, of $\hat{m}$=10,049 (325); $\hat{p}$=.02998 (.0083); $\hat{q}$=.39960 (.0004). The size of the bias decreases as the sample size (m) increases (results not shown).

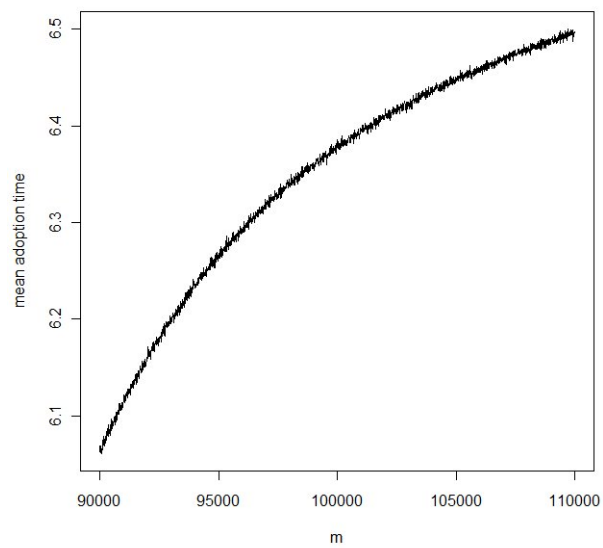

Figure 5: Average adoption time for the adopters $\tau(T, m)$, different values of m. Other parameters: $p = 0.03, q = 0.4, T = 10$. Ten artificial adoption sequences are simulated for each value of m. For each sequence, 10 replications of the estimation procedure are performed, with different pseudo-random numbers. For each set of estimated parameters, $\tau(T, m)$ is computed. The graph reports average values. Source: Grazzini et al. (2012b).

7 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 (in our companion paper (Grazzini and Richiardi, 2014) we extend the analysis to estimation of non-ergodic 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 relevance of identification and misspecification issues in AB models should be the object

of further investigation, similarly to Canova and Sala (2009). Estimating different versions of the same AB model and testing the robustness of the estimates would be just one interesting exercise. Fifth, 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). Finally, 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.

Acknowledgements

Jakob Grazzini received funding from the EU 7th framework collaborative project "Complexity Research Initiative for Systemic InstabilitieS (CRISIS)", grant no. 288501, while Matteo Richiardi acknowledges financial support from Collegio Carlo Alberto and Regione Piemonte within the research projects "Causes, Processes and Consequences of Flexsecurity Reform in the EU: Lesson from Bismarckian Countries" (Collegio Carlo Alberto) and "From Work to Health and Back: The Right to a Healthy Working Life in a Changing Society" (Regione Piemonte and Collegio Carlo Alberto).

The paper benefited from comments by participants at the 2013 and 2014 conferences of the Eastern Economic Association in New York and Boston, the 2013 European Meeting on Applied Evolutionary Economics in Nice, by seminar attendants at INET (2013) and the University Panthéon-Assas Paris II (2014) and by two anonymous referees.

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.

Amisano, G. and Tristani, O. (2010). Euro area inflation persistence in an estimated nonlinear DSGE model. *Journal of Economic Dynamics and Control*, 34(10):1837–1858.

Andreasen, M. M. (2011). Non-linear DSGE models and the optimized cebtral difference particle filter. *Journal of Economic Dynamics and Control*, 35:1671–1695.

Ascari, G., Fagiolo, G., and Roventini, A. (2014). Fat-tail distributions and business-cycle models. *Macroeconomic Dynamics*, FirstView:1–12.

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.

Bass, F. (1969). A new product growth for model consumer durables. *Management Science*, 15:215–227.

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. (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.

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.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Canova, F. (2007). *Methods for applied macroeconomic research*. Princeto.

Canova, F. (2009). How much structure in empirical models? In Mills, T. C. and Patterson, K., editors, *Palgrave Handbook of Econometrics*, volume 2: Applied Econometrics. Palgrave Macmillan.

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

Chang, Y. and Kim, S.-B. (2007). Heterogeneity and aggregation: Implications for labor-market fluctuations. *The American Economic Review*, 97(5):1939–1956.

Chang, Y., Kim, S.-B., and Schorfheide, F. (2013). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of the European Economic Association*, 11(s1):193–220.

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.

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

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

Cliff, D. and Bruten, J. (1997). Minimal-intelligence agents for bargaining behaviors in market based environments. *HP Laboratories Bristol*, (HPL-97-91).

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. (2014). 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.

Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P., and Gallegati, M. (2011). *Macroeconomics from the Bottom-Up*. Springer, Milan.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., and Palestrini, A. (2005). A new approach to business fluctuations: Heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior and Organization*, 56(4):489–512.

Dixon, H. and Bihan, H. L. (2012). Generalised taylor and generalised calvo price and wage settings: Micro-evidence with macro implications. *The Economic Journal*, 122(560):532–554.

Domowitz, I. and El-Gamal, M. A. (1993). A consistent test of stationary-ergodicity. *Econometric Theory*, 9(4):pp. 589–601.

Domowitz, I. and El-Gamal, M. A. (2001). A consistent nonparametric test of ergodicity for time series with applications. *Journal of Econometrics*, 102(2):365 – 398.

Dosi, G., Fagiolo, G., Napoletano, M., and Roventini, A. (2013). Income distribution, credit and

fiscal policies in an agent-based Keynesian model. *Journal of Economic Dynamics and Control*, 37(8):1598–1625.

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

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., Napoletano, M., and Roventini, A. (2008). Are output growth- rate distributions fat-tailed? some evidence from oecd countries. *Journal of Applied Econometrics*, 23(5):639–669.

Fagiolo, G. and Roventini, A. (2012). Macroeconomic policy in DSGE and agent-based models. *Revue de l'OCSE*, (124):67–116.

Fasolo, A. M. (2012). A note on particle filters applied to DSGE models. Working Papers Series 281, Central Bank of Brazil, Research Department.

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

Forni, M. and Lippi, M. (1997). *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford University Press.

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.

Geroski, P. A. (2000). Models of technology diffusion. *Research Policy*, 29(4-5):603–625.

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*.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proceedings F on Radar and Signal Processing*, 140(2):107–113.

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. and Richiardi, M. (2014). Partial identification in non-ergodic models. Technical report, INET, Oxford Martin School, University of Oxford.

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):25–40.

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.

Jenkins, S. P. (1997). Discrete time proportional hazards regression. *Stata Technical Bulletin*, 39:22–32.

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

Kirman, A. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives*, 6(2):117–136.

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.

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

Krusell, P. and Smith, A. A. (2006). Quantitative macroeconomic models with heterogeneous agents. Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress.

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.

Magnani, R. and Mercenier, J. (2009). On linking microsimulation and computable general equilibrium models using exact aggregation of heterogeneous discrete-choice making agents. *Economic Modelling*, 26(3):560–570.

Marks, R. E. (2013). Validation and model selection: Three similarity measures compared. *Complexity Economics*, 2(1):41–61.

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. (2014). 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.

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.

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

Rogers, E. M. (1962). *Diffusion of innovations*. Free Press, New York.

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

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.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175.

Smith, V. L. (1962). An experimental study of competitive market behavior. *The Journal of Political Economy*, 70(2):111–137.

Solow, R. (2010). Building a science of economics for the real world. Hearing before the House of Representatives 111th Congress, House Committee on Science and Technology, Subcommittee on Investigations and Oversight.

Stata 13 manual (2013). Discrete-time survival analysis. Technical report, Stata Corp. Entry: 'discrete'.

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.

Taylor, J. B. (1980). Staggered wage and price setting in macroeconomics. *Journal of Political Economy*, 88(1):1–23.

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

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):pp. 147–162.

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.