A Markov Chain approach to ABM calibration

Annalisa Fabretti
Department of Economics and Finance
University of Rome Tor Vergata
Rome, Italy
Email: annalisa.fabretti@uniroma2.it

***Abstract*—Agent based model are nowadays widely used, however the lack of general methods and rules for their calibration still prevent to exploit completely their potentiality. Rarely such a kind of models can be studied analytically, more often they are studied by using simulation. Reference [1] show that many computer simulation models, like ABM, can be represented as Markov Chains. Exploting such an idea we illustrate an example of how to calibrate an ABM when it can be revisited as a Markov chain.**

I. Introduction

Agent Based Models (ABMs) is an increasing popular approach in social sciences and economics, see [2] and [3]. Generally speaking, in an ABM a set of behavioural rules are defined at the individual level and some aggregate results are derived. Often this approach requires the use of large sets of parameters, needed to describe agent's attitudes. In order to validate the model and then use it for various purposes (prediction, policy analysis etc) finding the parameters value, which better fit the real world, is crucial. Parameters can be explicitly macro variables, as some in [4], or they can be directly estimated by data, as in [5]. Sometimes the peculiarity of the model allows a specific calibration, as in [6], where the maximum likelihood method can be applied since a closed form solution of the distribution of returns can be derived. However it is not always possible to have a closed form for some distributions or a direct simulation of some observables. Often we tackle with individual parameters, that cannot be derived directly from the real world, from which we observe only aggregated variables; in general model complexity does not allow an analytical tractability. That's why the majority of cases must be treated via simulation-based estimation techniques.

A first attempt of calibration can be found in [7], which propose a method of simulated moments with a focus on optimisation heuristic. However, when working with simulated moments, problems related to consistency and identifiability can be relevant, see [8].

In [7] the parameters estimators are obtained as minimum of the simulated minimum distance between moments via a combination of a simplex algorithm [9])and a threshold accepting algorithm [10]. In [11] this optimisation technique has been compared with Genetic Algorithms [12] and the latter appear to perform better; see also [13].

We refer to [14] for a more general discussion about calibrating ABM in economics.

Notwithstanding such interesting attempts, the field of parameters estimation in ABM is not yet enough explored and it needs to find general rules.

According to [1] "..many computer models in the social simulation literature can be usefully represented as time-homogeneous Markov chains." Following this idea we study the problem of setting parameters when the model can be seen as a Markov chain. To this aim we propose a classical minimum distance estimation using a genetic algorithm as minimum searching procedure. The Markov chain approach helps in finding consistent estimators to consider in the objective function. In particular an ergodic and aperiodic Markov chain admits maximum likelihood estimators (MLE) both for the transition probability matrix entries and the equilibrium distribution vector. To test this approach we fit the Kirman model [15], for which it is suitable to use the MLE for the transition probability. Then we compare the result with a direct fit performed using the analytical form of the transition probability.

In order to avoid problems due to the model specification we generated pseudo real data by the model itself with a known set of parameters, then we fitted the model on these data. Future researches are dedicated to improve such a methodology testing other models, enlarging the number of parameters and considering other estimators as the sample occupancy distribution.

We present the main concepts on parameter estimation and Markov chains needed to explain this approach in Section II; in Section III we briefly describe a standard Genetic Algorithm, in Section IV we present an example introducing the model and its Markov chain-version; in Section V we present and discuss the results; Section VI concludes.

II. Parameters Estimation and Markov Chain

Usually, when treating with simulation models, the estimation procedure is based on the comparison of the statistical properties of data with the statistical properties of the model. Indeed, rarely the model properties can be studied analytically directly by the model, hence they are inferred by analysing simulated data. By another hand revisiting the model as a markov chain allows to study some model properties.

In the following we discuss briefly some important issues related to the estimation, as consistency and parameters identification, and we introduce some basic concepts about Markov Chains.

A. Parameters Estimation: consistency and identifiability

We explore the idea to use a Generalized Method of Moments (GMM) or a Classical Minimum Distance (CMD) (see [16]), suitable for simulation models, in which the objective function is built using known estimators for Markov chain. In this paragraph we address briefly problems related to consistency, identifiability and convergence.

Definition 1: A consistent estimator $\hat{\theta}$ is one that converges in probability to the true value θ_0, i.e. $\hat{\theta} \to^p \theta_0$.

Definition 2: An estimator $\hat{\theta}$ is an extremum estimator if there is an objective function $\hat{Q}_n(\theta)$, with $\theta \in \Theta$, such that $\hat{\theta}$ maximizes such a function.

The extremum estimator depends on n and must be denoted by $\hat{\theta}_n$. The following theorem gives the conditions for consistency.

Theorem 3: If there exists a function $Q_0(\theta)$ such that

i) $Q_0(\theta)$ is uniquely maximised at θ_0;
ii) Θ is compact;
iii) $Q_0(\theta)$is continuous;
iv) $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$;

an extremum estimator $\hat{\theta}$ is consistent, i.e. $\hat{\theta} \to^p \theta_0$.

Condition i) and ii) are substantive and are called the identification condition and the compactness condition, respectively. Conditions iii) and iv) are regolarity conditions.

Let us consider an objective function of the form

$$\hat{Q}_n(\theta) = -\hat{g}_n(\theta)'\hat{W}_n\hat{g}_n(\theta), \tag{1}$$

where $\hat{W}_n$ is a positive semi-definite matrix and $\hat{g}_n(\theta)$ is a vector of data and parameters.

When $\hat{g}_n(\theta)$ takes the form

$$\hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta),$$

where $g(z_i, \theta)$ are sample moments, we have the generalized method of moments (GMM), while when

$$\hat{g}_n(\theta) = \hat{h}_n - h(\theta),$$

where $\hat{h}_n$ is a set of estimators, $h(\theta)$ is a vector of functions that map between the model and the estimator, we speak of classical minimum distance (CMD).

By the law of large numbers $\hat{g}_n(\theta) \to^p g_0(\theta)$, where $g_0(\theta) = E[g(z, \theta)]$, so that if $\hat{W}_n \to^p W$, where W is a positive semi definite matrix, by continuity of multiplication $\hat{Q}_n(\theta) \to^p Q_0(\theta) = -g_0(\theta)'Wg_0(\theta)$. Similarly for the CMD, if $\hat{h}_n \to^p h_0$ and $\hat{W}_n \to^p W$ then $\hat{Q}_n(\theta) = -[\hat{h}_n - h(\theta)]'\hat{W}_n[\hat{h}_n - h(\theta)] \to^p -[h_0 - h(\theta)]'W[h_0 - h(\theta)]Q_0(\theta)$.

The identification condition i) consists in requiring that the distribution of data at the true parameter is different than that at any other possible parameter value. For the GMM we have the following lemma.

Lemma 4: If W is positive semi definite and, for $g_0(\theta) = E[g(z, \theta)]$, $g_0(\theta_0) = 0$ and $Wg_0(\theta) \neq 0$ for $\theta \neq \theta_0$ then $Q_0(\theta) = -g_0(\theta)'Wg_0(\theta)$ has a unique maximum at θ_0.

The analysis of the CMD is similar. The condition for $Q_0(\theta)$ to have a unique maximum at θ_0 is that $h(\theta_0) = h_0$ and $h(\theta) - h(\theta_0)$ is not in the null space of W if $\theta \neq \theta_0$. If W is non singular this condition reduces to say $h(\theta) \neq h(\theta_0)$.

If the moment functions $g(z, \theta)$ or the estimator map $h(\theta)$ are linear in θ the previous identification conditions become the rank condition (to have at least as many functions or moments as parameters) otherwise the analysis of global identification is difficult.

Compactness of the parameter space Θ, that is condition ii) of theorem 3 is also relevant, but it depends on the parameters nature. Uniform convergence and continuity, i.e. condition iii) and iv) of theorem 3 are easily verified when we use moments or maximum likelihood estimators (MLE).

B. Markov Chain

A Markov chain is a stochastic process with the Markov property. Formally it is a sequence of random variables $X_1, X_2, ..$ such that

$$P(X_{t+1} = x | X_1 = x_1, ..., X_t = x_t) = P(X_{t+1} = x | X_t = x_t).$$

The random variable X_t represents the state of the system at time t. If the probability of X_t to be x does not depend on t, the Markov chain is time homogeneous. Let $\mathcal{S}$ be the space of possible values of X_t; $\mathcal{S}$ is called the state space, which can be finite or infinite. We consider here only finite state space. Thus, let $\mathcal{S}$ have m possible states, we define the transition matrix as the $m \times m$ matrix, whose element p_{ij} is the transition probability from state i to state j, i.e.

$$P = \{p_{ij}\} = \{P(X_{t+1} = j | X_t = i)\} \qquad i, j = 1, ..., m.$$

A stationary distribution π is a row vector whose entries are non negative, sum to 1 and

$$\pi P = \pi.$$

An irreducible Markov Chain has a stationary distribution if and only if all its states are positive recurrent. Let denote by $p_i^{(t)}$ the probability to be in state i at time t, formally

$$p_i^{(t)} = Pr(X_t = i) \qquad i = 1, .., m.$$

If the positive recurrent chain is both irreducible and aperiodic it has a limiting distribution, formally

$$\pi_i = \lim_{t \to \infty} p_i^{(t)},$$

for $i = 1, .., m$, which is called also the equilibrium distribution and equals the stationary one. On the long run the probability to be in state i at time t equals the equilibrium distribution.

Let

$$N_i = \sum_{t=1}^{n} \mathbf{1}_{x=i}(x_t),$$

where $\mathbf{1}_{x=i}$ is the indicator function, be the number of time the system stays in the state i. Let

$$N_{ij} = \sum_{t=1}^{n-1} \mathbf{1}_{x=i}(x_t)\mathbf{1}_{x=j}(x_{t-1}),$$

be the sample number of transitions between i and j.

We consider two estimators which can both be derived as MLE:

- The sample occupancy distribution

$$\hat{P}_i = \frac{N_i}{\sum_i^m N_i} \qquad i = 1, .., m, \tag{2}$$

  as estimator for the equilibrium distribution.
- The estimator for the transition probability

$$\hat{p}_{ij} = \frac{N_{ij}}{\sum_j N_{ij}}. \tag{3}$$

These estimators are candidates to be $\hat{h}_n$.

III. Genetic Algorithm

Genetic algorithm are search procedure based on natural selection and genetics, see [17].

A GA starts with a population of strings (a string is a coding of parameters) and thereafter generates successive populations of strings. Among the advantages of GA compared to traditional methods, we encounter that GA works simultaneously on a large number of points (a population of strings) while traditional methods on a point at time. In this way, they explore widely the search spaces and the probability of finding a false peak is reduced. Moreover GAs works with a coding of the parameters set, no with parameters themselves and use payoff information (objective function), while traditional methods use other auxiliary information, as for example derivatives in gradient techniques. Finally GAs use probabilistic transition rules.

Briefly, the algorithm starts generating a random population of N chromosomes (which are the suitable solution). Then, the algorithm evaluate the objective function, also called the fitness function, for each chromosomes set and select the individuals (each individual owns a chromosome set), which show the best ability to survive. Those individuals are elected as parents of a new population which will be generated by two following steps: crossover and mutation. Moreover propagation can be also considered. The crossover consists in the generation of new chromosome sets starting from parents' chromosome, then each new chromosome set undergoes to a mutation with a certain probability. Propagation consists in replicating a certain number of old individuals into the new population. Usually these individuals are those with best fitness function. The newborn offspring replaces the old population and the procedure is repeated till a best solution is reached.

The combination of crossover, mutation and propagation gives the evolutionary motion and ensure the convergence to a feasible optimum.

IV. An example of a model seen as Markov chain

The model, we consider here, was proposed in 1993 by Kirman. Taking inspiration by ants behaviour in exploiting two equal sources of food, the model simply considers the interaction between individuals; as result it emerges a herding behaviour which is typically observed also in human behaviour.

There are two sources of food, denoted by two colours (black and white), there are N ants, each feeding at one of the two sources. The state of the system X_t is defined by the number of ants feeding at the black source and can vary between 0 and N. At each time t, two ants meet at random and the first adopts the second's colour with probability $(1-\delta)$. There is also a probability ϵ that the first ant will change autonomously the colour. The dynamical evolution is described by a Markov chain where X_t can assume values k, with $k = 0, 1, ..N$. The transition probabilities $P(k, k+1)$ from k to $k+1$ and $P(k, k-1)$ from k to $k-1$ are the following

$$\begin{aligned} p_1 &= P(k, k+1) = \left(1 - \tfrac{k}{N}\right)\left(\epsilon + (1-\delta)\tfrac{k}{N-1}\right) \\ p_2 &= P(k, k-1) = \left(\tfrac{k}{N}\right)\left(\epsilon + (1-\delta)\tfrac{N-k}{N-1}\right) \end{aligned},$$

where $p_1 + p_2 \leq 1$ and the probability to stay unchanged is $1 - p_1 - p_2$.

This Markov chain is regular, ergodic and admits a stationary distribution $\pi(k)$, for $k = 0, .., N$, which can be obtained solving

$$\pi(k) = \sum_{j=0}^{N} \pi(j) P(j, k). \tag{4}$$

Hence, we have

$$\pi(k) = \frac{\frac{P(0,1)}{P(1,0)}\frac{P(1,2)}{P(2,1)} \cdots \frac{P(k-1,k)}{P(k,k-1)}}{1 + \sum_{j=1}^{N} \frac{P(0,1)}{P(1,0)}\frac{P(1,2)}{P(2,1)} \cdots \frac{P(j-1,j)}{P(j,j-1)}}. \tag{5}$$

The calibration of this model consists in estimating ϵ, δ and N, hence we have $\theta = (\epsilon, \delta, N)$. In a first istance we prefer to avoid the estimation of N, because either it is an integer or it can assume infinite values and the parameter space Θ could not be compact. Instead considering only $\theta = (\epsilon, \delta)$, the parametere space Θ is the square $([0,1]^2)$, which is closed and bounded in $\mathbb{R}^2$, hence compact.

To ensure identifiability it is preferable to choose the MLE for the transition probability, see (3), because the corresponding map $h(\theta)$ is linear in θ so that we have global identifiability. Note that the transition probability is not linear in N, another reason to avoid preliminary a calibration with respect to N.

V. Preliminary Experimental Results

Using the model we produced the pseudo real data to be fitted. The true parameters are set to $\epsilon_0 = 0.02$ and $\delta_0 = 0.5$, the number of agents is 100 and is not fitted at this stage.

Data consists of 10000 simulations of the random variable X_t described above; observed values go from 20 to 91. Generating data from the model itself allows to avoid problems related to model specification. The sample transition matrix has been estimated using a subsampling bootstrap methodology, extracting 1000 times a subsample of length of 300. In this way we could also estimate the variance of each single

probability and derive the matrix $\hat{W}_n$, which is the inverse of the diagonal matrix constituted by the variances. We used only 6 transition probabilities, specifically the $P(k,k)$ for $k = 40, ..., 46$. In this way the rank condition is verified and the computational time is reduced.

The genetic algorithm has been run with 40 individuals, with a mutation probability of 90%, a crossover probability of 90%, a propagation of 2 individuals at each generation and "tournament" as selection rule.

A minimum has been reached in $\hat{\epsilon} = 0.0218$ and $\hat{\delta} = 0.5047$.

The algorithm has fitted quite precisely δ and satisfactorily ϵ. We compare this result to a direct fit possible in this case having an explicit formula for the transition probability. Considering the probability to go from k to $k+1$, p_1 in (IV), we can re-write it as a second-degree function:

$$f(k) = -\frac{(1-\delta)}{N(N-1)}k^2 + k\left(1 - \delta - \frac{\epsilon}{N}\right) + \epsilon.$$

Fitting this function using a non linear least square standard algorithm we get $\hat{\epsilon} = 7.48 \cdot 10^{-5}$, $\hat{\delta} = 0.4387$ and $\hat{N} = 98$. In this case we are able to fit also the number of agents, but results are decisively worste.

VI. Conclusion

In order to contribute to the field of ABM calibration we proposed to exploit the idea to revisit an ABM as a Markov chain, see [1]. To this aim we consider the Kirman model which can be easily described as an ergodic and aperiodic Markov chain and we calibrated it using a Classical Minimum Distance minimised by using a Genetic Algorithm. Since the model allows us to derive analytically the transition probabilities, we also fit directly such a function extracting then the parameters. The first method is more general and performs better than the second. Indeed results are preliminary but satisfactory. We believe that this study can contribute to enlarge our knowledge on the field of ABM calibration and for this reason deserves more investigation. Future researches will test this approach with other models, enlarging the number of parameters and considering cases in which the map $h(\theta)$ cannot be derived analytically but it must be obtained via simulation.

References

[1] Izquierdo, L. S., Izquierdo, S. S., Galan, J. M., Santos, J. I.: Techniques to Understand Computer Simulations: Markov Chain Analysis. J. Art. Soc. Social Sim. **12**(16) (2009)

[2] Hommes, C. H.: Heterogeneous Agent Models in Economics and Finance. In: Tesfatsion L., Judd K.L. (eds.) Handbook of Computational Economics vol. 2 pp. 10-13. Elsevier Science B.V.,(2006)

[3] Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P., Gallegati, M.: Macroeconomics from the bottom-up. Springer, Berlin (2011)

[4] Bianchi, C., Cirillo, P., Gallegati, M., Vagliasindi, P.A.:Validating and calibrating agent-based models: a case study. Comput. Econ. **30**, 245264 (2007)

[5] Mike, S., Farmer, J. D.: An empirical behavioral model of liquidity and volatility. J. of Economics Dynamics and Control **32**,200–234 (2008)

[6] Alfarano S., Wagner F., Lux T. :Estimation of Agent-Based Models: The Case of an Asymmetric Herding Model. Computational Economics **26** 19–49 (2005)

[7] Gilli, M., Winker, P.: A global optimization heuristic for estimating agent based models. Comp. Stat. & Data Anal. **42**, 299–312 (2003)

[8] Grazzini, J.,Richiardi, M.G. : Consistent Estimation of Agent-Based Models by Simulated Minimum Distance. LABORatorio R. Revelli Working Papers Series 130, Centre for Employment Studies (2013)

[9] Nelder, J. A., Mead R.: A simplex method for function minimization. Computer Journal **7**, 308313 (1965)

[10] Winker, P.: Optimization heuristics in econometrics: applications of threshold accepting. John Wiley and Sons Ltd.,Chichester, UK, (2001)

[11] Fabretti, A.: On the problem of calibrating an agent based model for financial markets. J of Econ Int and Coord, Springer, **8**(2), 277-293, (2013).

[12] Holland, J.H: Adaptation in Natural and Artificial Systems. MIT Press (1975)

[13] Calvez B., Hutzler G.: Automatic Tuning of Agent-Based Models Using Genetic Algorithms. MABS'05 41-57 (2005)

[14] Windrum, P., Fagiolo, G., Moneta, A. : Empirical Validation of Agent-Based Models: Alternatives and Prospects. J. of Art. Soc. and Social Sim. **10**(2), (2007)

[15] Kirman, A.: Ants, rationality and Recruitment. Quat. J. Econ. **108**(1), 137–156 (1993)

[16] Newey, W.K., McFadden, D. : Large sample estimation and hypothesis testing. In: Engle, R. F., McFadden, D. (eds.) Handbook of Econometrics, pp. 2111-2245. Elsevier (1986)

[17] Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley (1989)