

Comments on the paper

"Reconstructing Aggregate Dynamics in Heterogeneous Agents Models" by D. Delli Gatti *et al.*

Sylvain Barde

School of Economics, Keynes College, University of Kent, Canterbury, OFCE

The main aim of the paper is to apply the image processing interpretation of the Maximum Entropy (MaxEnt) method to the Kirman (1993) model and the Abrams and Strogatz (20003) voter model as implemented by Stauffer *et al.* (2007). This follows the initial work in Barde 2012 which showed that the Schelling (1969) model of segregation can be predicted with the methodology. The discussant does point out some of the major issues that are associated with the methodology, many of which I agree with. The most important comment is probably the fact that more exploratory work is needed to establish a taxonomy of valid assumptions for corresponding statistical properties. Having said this, I feel that two important clarifications are needed.

My first comment relates to the claim that the assumptions or simplifications required to obtain the MaxEnt solution are arbitrary. Given some data d (the initial condition in agent-based models), the basic formulation for obtaining the prediction μ the maximum entropy problem is given by:

$$\max_{\mu} [\alpha S(\mu | m) + \ell(d | \mu)]$$

The first part of the expression, $S(\mu | m)$ is the relative entropy of with respect to a model m and $\ell(d | \mu)$ is the likelihood that the initial condition d is a noisy version of the prediction μ . For any given problem, two terms need to be specified: the model term m and log likelihood $\ell(d | \mu)$. While there is an element of 'educated guessing' in specifying these terms, this is not as arbitrary as the discussant claims.

— The model term m is a diffusion term which specifies how far the prediction can stray from initial condition, and this is the term that controls for time in the system. Intuitively, if very little time has elapsed, one should use a very peaked m , as μ will be very close to d . Conversely, long time horizons are represented with a flatter m . It is

also important to note that m can have several dimensions, depending on the nature of the problem: one dimensional for the ants model, two dimensions for the Schelling and voter models.

— The likelihood term ℓ depends on the nature of the path linking the initial condition to the predicted state of the system. The image-reconstruction algorithm treats μ as the true image to be discovered and d as a noisy version of μ . This time-reversed path is conditioned on the fact that if the sequence of actions taking the system from its initial condition to its equilibrium distribution is best-response (a common assumption in economics), then the reverse path is effectively a noise process. The likelihood term is therefore determined by knowledge of the updating process, which determines the implicit noise process in the reversed path.

Both these terms are determined from the updating rules of the system, and are therefore not as arbitrary as it may seem. It is true that if little information is available (for instance if the exact transition probabilities are unknown), they must be approximated. For instance, in the generic version used for the voter model, both a gaussian likelihood $\ell(d | \mu)$, *i.e.* a gaussian noise process, and gaussian correlations over two-dimensional space for the model term m are assumed as an approximation. However this can be refined if more information is available from the updating process. This is the case in the ants model, where the transition probabilities are well known. In this case the model term is the diffusion of a stopped random walk rather than a gaussian diffusion and the likelihood is designed directly from a path integral of the transition probabilities.

Clearly, MaxEnt is no miracle solution: if the researcher has no information about the dynamic updating process of a system, then there is no way that knowledge of the initial condition alone can lead to a decent prediction of future states. In the Kirman ant model, for instance, the initial condition at $t = 0$ is simply a value $x \in [0,1]$ representing the share of ants of a certain colour. If the researcher is ignorant of the recruitment mechanisms, then x alone does not provide much information on the stable distribution of the system at a later time $t = n$. The central argument for using MaxEnt in the context of agent-based models is precisely that the updating rules of the system are known *ex ante*, as they are provided by the researcher.

My second comment is would be that the aim of the methodology is not to replace the traditional Monte-Carlo methods used in agent-based models but instead to provide a complement. The methodology is analytical in so far as the derivation of the maximum entropy

problem is obtained from a rigorous Bayesian approach however, as mentioned by the discussant, in most cases a numerical methodology is required to solve for the solution of the problem. Furthermore, as pointed out by the discussant, the three simple models analysed so far with MaxEnt are a far cry from the complex systems routinely used in the agent-based literature. So given this, what is the usefulness or purpose of the proposed methodology?

An important application in my opinion is to provide a tool for categorising types of agent-based models according to the strength of their convergence to a stable distribution. A key finding of the paper, as well as the companion work on the Schelling model is that while the three models are clearly stochastic, the fact that they are amenable to MaxEnt prediction reveals that they are much more predictable than one might think. In technical terms, this is related to the fact that the image reconstruction MaxEnt algorithm works only if one is able to treat the reversed time-evolution of the system as a noise process, indicating that the time-evolution is in fact a finite improvement path. I agree with the discussant that more work is needed

In the future, rather than providing a direct solution tool for large agent-based model, a potentially important application for MaxEnt is the prediction of those component modules of the larger model that are amenable to MaxEnt. An interesting possibility in this regard is to take advantage of the faster execution speed of the methodology compared to Monte-Carlo to directly provide agents in the model with expectations, by using MaxEnt on the current state to obtain predicted future values for key state variables. Similarly, it could be used to speed-up large agent-based models by using the faster MaxEnt method on those components that are known to be amenable to the methodology.

References

Abrams D. M. and S. H. Strogatz, 2003. "Modelling the dynamics of language death." *Nature*, 424:900.

Barde S., 2012. "Back to the future: economic rationality and maximum entropy prediction." *University of Kent School of Economics Discussion Paper* 12(02).

Kirman A., 1993. "Ants, rationality and recruitment." *Quarterly Journal of Economics*, 108:137–156.

Schelling T. C., 1969. "Models of segregation." *American Economic Review*, 59:488–493.

Stauffer D., X. Castelló, V. M. Eguíluz, and M. San Miguel, 2007. "Microscopic Abrams-Strogatz model of language competition." *Physica A*, 374:835–842.

Reply to Comments

Domenico Delli Gatti

Institute of Economic Theory and Quantitative Methods, Università Cattolica del Sacro Cuore, Milan

Corrado Di Guilmi

School of Finance and Economics, University of Technology, Sydney

Mauro Gallegati

DiSES, Università Politecnica delle Marche

Simone Landini

IRES Piemonte, Turin

In our view, the aggregation problem does not boil down to simple "averaging" in such a way as to resurrect the Representative Agent.¹ In order to elaborate on this, we should start from the following notion: In a (macro) system there can be elementary (micro) and composite (meso) constituents. Micro constituents are units which agglomerate into within-homogeneous but between-heterogeneous sub-systems (meso constituents). Accordingly, a (macro) system can be seen as made of (meso) sub-systems composed by (micro) elementary units, that is a statistical ensemble which represents all the significant configurations the system can assume.

At any level of observation, a quantity is a functional, whose realised values are measurement outcomes. For instance, micro-functions implemented in an ABM are micro-stochastic processes constituting a statistical ensemble. In a single run of simulations, an ABM will generate a sample of numbers which is the realisation of the collection of their outcomes at each point in time: in a sense, an ABM is a space-time random field.

The numeric outcome of each micro-functional can be thought of as the outcome of an experiment, hence it is the measurement of a certain quantity on an observation unit.

A transferable quantity is a variable whose aggregate value is given by the summation of the constituents' values. Only transferable quantities admit an exact/algebraic aggregation. Non transferable quantities are system specific: being realised by the superimposition of

1. The RA is not the average agent, technically it is more properly an estimator for the system as a collective body characterised either by transferable and not transferable quantities.

underlying micro-level quantities they are emergent information. For instance it is not possible to algebraically aggregate individual prices (they are non transferable quantities): their mean is not the market price but the average price in the market. The market price pertains to the market as a collective entity, a system by itself. The inflation rate does not make sense at the individual level but it depends in some way on individual behaviours.

It is possible to associate a stochastic process to each kind of quantity in order to have aggregation in terms of expected values. The expected value of a given observable variable is a functional and does not coincide with the average. The expected value is the estimator of the first moment of a stochastic process, the average is a particular realization of that estimator given a set of experimental outcomes.

From the algebraic point of view aggregation is not a problem if quantities are transferable. A collection of numbers characterising the same property of the system's constituents can always be added up to generate the aggregate value. This makes the aggregation problem somehow misleading. Indeed, if a collection of realised numbers $\{y_{i,t}\}$ from a transferable quantity is available, then $Y_t = \sum_i y_{i,t}$ solves the problem. But what if $y_{i,t} = f(x_{i,t})$? Is it still true that $Y_t = f(X_t)$? Moreover, what if we know Y_t and X_t but cannot observe the micro-data? Given a set of micro-data from a transferable quantity it will always be possible to determine an exact system level number by means of algebraic aggregation. This is not possible in the other two cases. Therefore, the problem is inferential as concerning the macro-functional. In terms of micro-foundation things are even more complicated. The correct question in this case is: given the macro (X_t, Y_t) what is the micro $(\{y_{i,t}\}, \{x_{i,t}\})$ which is consistent with it? The answer is: the most probable one. Therefore, the expected value functional is needed before an average estimate.

As stated above, the system's constituents can be thought to agglomerate into sub-systems which are within-homogeneous and between-heterogeneous with respect to some criterion. This aspect leads to a mean-field approach to aggregation.

Mean-field can be seen as a method to determine aggregate functionals at sub-system level taking into account the phenomenology of micro-functionals and of their realisations. If mean-field were made explicit in terms of expected values, averages would of course be given at sub-system level, but in no way this implies that these values are representative of a collective agent in the same way as the representative agent does. The representative agent is a simplifying assumption

which allows to manage heterogeneity and interaction for practical purposes when dealing with problems of aggregation in a micro-founded context. It is (or behaves) as a collective body on a smaller scale: in its extreme version, the representative agent is associated with the system as a whole, annihilating every kind of heterogeneity for system's constituents Hartley (1997). The representative agent can be thought of as the estimator of a (sub-)system, but it still remains a description of a collective body on a reduced scale level: a per-capita value is not a property of the individual, it is still a system property. With per-capita values we are used to compare (sub-)systems not individuals. The fact that one can think of the average as a numerical *aggregator* of micro values can therefore be misleading because it might be thought that a mean-field approximation of the system is equivalent to the representative agent.

These statements can be made specific by considering the mean-field approach in the master equation framework for the dynamics of a probability density for a given observable on a certain state space. The density $P(N_j(t), t)$ is the probability distribution of micro-constituents over a state space of sub-systems, which are shaping the configuration of the macro system. Therefore, in the master equation framework, the density $P(\cdot)$ is to be conceived as a control-functional. On the other hand, the mean-field observable $N_j(t)$ plays the role of a state-functional.

In mean-field terms, one can specify a model for $N_j(t)$ which links its realisations to some other quantities at macro level to take care of the environment feedbacks, as if they were some force-fields acting on system constituents and inducing their agglomeration into sub-system as an externality effect. It is also possible to specify these effects in terms of *effective interactions* among sub-systems, which is what Aoki (Aoki, 1996; Aoki, 2002; Aoki and Yoshikawa, 2002) calls mean-field interaction by means of transition rates. Moreover, there can be also some emerging characteristic which drives the most probable path trajectory of the state-functional, which is what Aoki calls the macroeconomic equation, best known as macroscopic equation and which can be associated to the notion of pilot-quantity, at least according to the pilot-wave theory in the Bohminian interpretation of quantum mechanics (see Bohm, 1952a; Bohm, 1952b).

Among the methodologies to solve master equations (see Kubo *et al.*, 1973; Gardiner, 1985; Risken, 1989; van Kampen, 1992; Aoki, 1996; Aoki, 2002; Aoki and Yoshikawa, 2002) when the state-functional is known to be distributed as unimodal and peaked about its

expected value, $N_j(t)$ can be expressed by means of the van Kampen ansatz: $N_j(t) = N\phi(t) + \sqrt{N}\varepsilon(t)$. In this representation, the macroscopic equation drives the expected value of the share of agents occupying the j -th state and is itself a function of transition rates, $\dot{\phi}(t) = \phi(\beta(t), \delta(t))$, each of which—in Aoki's interpretation—includes the effect of the environment on $N_j(t)$ by means of the so called externality functions $\psi_j(t)$ depending on system quantities. Therefore, being a fully functional development of the system, and allowing for heterogeneity and interaction, the mean-field/master equation approach cannot be confused with a representative agent, unless the representative agent were specified as an estimator for the system as a collection of collective bodies (sub-systems) each of which takes a place on the state space and obeying an exclusion-like principle, which is not a very reliable assumption. Differently said, two sub-systems in the same state are almost the same sub-system and they can be lumped into a larger body because their elementary constituents belong to the same micro-state.

Finally there is one more technical aspect which needs to be dealt with: a master equation, in general, does not admit a closed form solution but requires an approximation method to be solved. Basically there are three methods of approximation, each of which has been described by Aoki (Aoki, 1996): Kubo method (Kubo *et al.*, 1973), Kramers-Moyal expansion (see Gardiner, 1985; Risken, 1989) and van Kampen system size expansion (see van Kampen, 1992). A fourth method is also available, it is the one developed in our paper and it can be called Aoki method: in essence it is a variant of van Kampen's, even though more natural and easy to deal with. All these methods share a common feature: they are grounded on approximation techniques. In the van Kampen/Aoki perspective, by using the ansatz $N_j(t) = N\phi(t) + \sqrt{N}\varepsilon(t)$ into an explicit definition of transition rates, the master equation for $P(N_j(t), t)$ is transformed into a master equation with respect to the spreading fluctuations term $\varepsilon(t)$, that is concerning the density $Q(\varepsilon(t), t)$. This new master equation is perfectly equivalent to the original one and its approximation is as follows: transition rates are Taylor approximated about the drift $\phi(t)$, the density $Q(\varepsilon(t), t)$ is Taylor approximated about the spread $\varepsilon(t)$. Due to the transformation $P(N_j(t), t) \equiv Q(\varepsilon(t), t)$ and a time rescaling, a system size parameter N enters the new master equation and its approximation. Hence, by applying the polynomial identity principle, two differential equations can be asymptotically isolated: the first one for the dynamics of the most probable drifting path trajectory, $\phi(t)$,

and the second one for the dynamics of the probability density of spreading fluctuations $\partial_t Q(\varepsilon(t), t)$. The first one is the macroscopic equation, and it depends on transition rates, even though it reads as an ordinary differential equation. The second one asymptotically converges to a Fokker-Planck equation as the system size increases. The macroscopic equation can be solved separately from the Fokker-Planck, its solution can therefore be used to solve the latter. Very often, the Fokker-Planck can be analytically solved with standard methods but, if the transition rates are too complicated, the solution can also be found *systematically*, in van Kampen's terminology. Indeed, by setting the stationarity condition $\partial_t Q(\varepsilon(t), t) = 0$, the Fokker-Planck equation boils down to a continuity equation obeying Liouville theorem, and it reads as an Hamilton-Jacobi equation. Since it asymptotically concerns a second order approximation, the stationary distribution is found to belong to the family of exponential distributions of Gaussian type. Therefore, what one really needs is a set of coupled equations (called the mean-field system) for the first and the second moments to get the dynamic functionals for the expected value and the variance driving the density $Q(\varepsilon(t), t)$ through time.

Two remarks are in order at this point. First, the differential equations for the expected value and variance functionals of the spreading fluctuations distribution about the drift depend on the transition rates and the macroscopic equation (here it comes its pilot role), and this shows that fluctuations about the drifting path trajectory have an endogenous specification in terms of mean-field or *effective interaction*. Secondly, what has been found to be Gaussian is not the solution of the master equation itself, but the distribution of fluctuations: van Kampen/Aoki methods do not provide a properly said Gaussian approximation to the model.

This method is not less valid than the Kramers-Moyal or Kubo methods just because of approximation. Indeed it does not properly allow for Gaussian approximation of the master equation, while Kubo method guesses a-priori an exponential probability kernel of Gaussian type and, if the Pawula's theorem (see Risken, 1989) conditions for the second order approximation are not fulfilled, Kramers-Moyal method is by definition approximate without any asymptotic behaviour. Moreover, as it can be done either with Kubo and Kramers-Moyal methods, the van Kampen/Aoki method can deal with higher order moments, which usually characterise asymmetric distributions. The weakness of van Kampen/Aoki methods is that they provide a local approximation about the drift for the state-functional $N_j(t)$, while

Kramers-Moyal and Kubo methods provide a global approximation for the probability density $P(N_j(t), t)$ control-functional. In our model this is almost irrelevant because the markovian nature of the model allows quite naturally for a second order approximation, and because the state space is trivial being made of two states only. Therefore, the state functional $N_1(t)$ is necessarily unimodal, and this allows for the shown ansatz. In general, with more complex state spaces, non-unimodal distributions and asymmetries, Kramers-Moyal method gives better results provided some reasonable order of approximation.

In our opinion, Aoki's interpretation of the Master Equation Approach (MEA) combined with Mean-Field Approximation (MFA) leads to three main theoretical results with promising applications to socioeconomic disciplines, mainly developed in macroeconomics:

1. *stochastic aggregation* of complex systems made of interacting and heterogeneous constituents;
2. *inferential identification* of drift and spread stochastic functionals as dynamic components of time series at the system's level;
3. *endogenous modelling* of interaction and spreading fluctuations about cyclical drift as the macroscopic emergent phenomenon due to the superimposition of microscopic behaviours.

Of course the MEA-MFA does not solve all the methodological and technical problems of macroeconomic modelling but it makes some steps beyond theoretical and practical problems the standard model is facing in micro-foundation of macro-models and aggregation of micro-behaviours. One of the most intriguing suggestions that have emerged from the issues dealt with in the paper and its discussion is to take in account local and global interaction by means of a nested structure consisting of groups made of sub-groups which can be partitioned into even smaller agglomerations over a finite *hierarchical structure of concentric levels*. In principle it might be possible to specify several master equations one nested into the others from the higher to the lower level of description. Each equation should be used to distinguish different interactive environments, from the very global to the most local one. The deeper one goes through this structure the more the interactions become less global, or more local, and the nested combination should take care of field-effects exerted by the level above or outside on the level below or inside. It is our opinion that this structure could be promising for two related purposes: it can describe transitions between different areas of a state space by considering dynamic transitions through partitions within each areas, and it can be

a starting point to develop a phase-transition and self-organised-criticality analysis for complex socioeconomic systems.

References

- Aoki M., 1996. *New Approaches to Macroeconomic Modelling*, Cambridge University Press.
- Aoki M., 2002. *Modeling Aggregate Behaviour and Fluctuations in Economics*, Cambridge University Press.
- Aoki M., and H. Yoshikawa, 2002. *Reconstructing Macroeconomics*, Cambridge University Press.
- Bohm D., 1952a. "A suggested Interpretation of the Quantum Theory in Terms of Hidden Variables I." *Physical Review*, 85: 166-179.
- Bohm D., 1952b. "A suggested Interpretation of the Quantum Theory in Terms of Hidden Variables II." *Physical Review*, 85: 180-193.
- Gardiner C. W., 1985. *Handbook of Stochastic Methods*, Springer-Verlag.
- Hartley E.J., 1997. *The Representative Agent in Macroeconomics*, Routledge.
- van Kampen N. G., 1992. *Stochastic Processes in Physics and Chemistry*, North-Holland.
- Kubo R., K. Matsuo and K. Kitahara, 1973. "Fluctuations and relaxation in macrovariables." *Journal of Statistical Physics* 9: 51-93.
- Risken H., 1989. *Fokker-Planck Equation. Method of Solutions and Applications*, Springer-Verlag.

