Of ants and voters: maximum entropy prediction and agent based models with recruitment

Sylvain Barde

OF ANTS AND VOTERS
MAXIMUM ENTROPY PREDICTION OF AGENT-BASED MODELS WITH RECRUITMENT[1]

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

Maximum entropy predictions are made for the Kirman ant model as well as the Abrams-Strogatz model of language competition, also known as the voter model. In both cases the maximum entropy methodology provides good predictions of the limiting distribution of states, as was already the case for the Schelling model of segregation. An additional contribution, the analysis of the models reveals the key role played by relative entropy and the model in controlling the time horizon of the prediction.

Keywords: Information entropy, Agent-based models, Voter models.

The maximum entropy (MaxEnt) methodology was first introduced as a general method of statistical prediction by (Jaynes 1957a,b), who showed that its use in predicting the dynamic evolution of an unobserved system could be extended beyond its initial use in physics. This insight was incorporated into the bayesian image reconstruction framework of Cornwell & Evans (1985), Narayan & Nityananda (1986) and Skilling & Gull (1991). Assuming that the received data d about a signal is noisy or distorted, the observer is interested in obtaining a reconstruction μ of the original clean signal. If $p(x)$ is the probability measure for x, the

1. The author is grateful for the suggestions received at an OFCE seminar in June 2011, relating to applications of them ethodology to the two models investigated here. Any errors are of course the author's.

best reconstruction satisfies the maximum a posteriori criteria *max* $p(\mu \mid d)$. Bayes' theorem states that this should be proportional to a prior probability on the reconstruction, $p(\mu)$, multiplied by the likelihood $p(d|\mu)$ that the observed data originated from the reconstructed signal.

The MaxEnt methodology assumes an entropic prior of the form $p(\mu) \propto exp(S(\mu \mid m))$, where $S(\mu \mid m)$ is the relative Shannon (1948) entropy between the reconstruction μ and a model m, which is the observer's ex ante guess of the reconstruction, based on the data d. This choice of prior in the image restoration literature is underpinned by the rigorous bayesian formulation of Shore and Johnson (1980), who provide an axiomatic proof that the entropy measure S is the only prior that does not introduce biases into the reconstruction. As a result, the reconstruction can be identified as the one that maximises the following expression, where $\ell(d \mid \mu)$ is the log-likelihood $\log(p(d \mid \mu))$:[2]

$$p(\mu \mid d) \propto exp(\alpha S(\mu \mid m) + \ell(d \mid \mu)) \tag{1}$$

The MaxEnt methodology was initially introduced in economics by Foley (1994) and extended by Toda (2010) as a way of deriving the statistical equilibrium of a market, i.e. the equilibrium distribution of endowments over agents. In a companion paper to the present study, Barde (2012) shows that the problem of allocating goods between rational agents can be modeled as a congestion game that possesses the finite improvement property. This means that any initial condition is linked to a Nash equilibrium by a finite path. Because each step on this path is the result of agents performing welfare-increasing trades, the reversed improvement path (which starts at the Nash equilibrium and ends at the initial condition) can be interpreted as a noise process, where agents make systematic mistakes. This is shown to imply that the problem of predicting the Nash equilibrium from the initial condition is formally equivalent to the problem of retrieving an image that has been corrupted by noise.

2. The multiplicative α term allows for the fact that the entropic prior $p(\mu)$ is defined only up to a multiplicative constant. α therefore effectively serves as a lagrangian parameter for the maximisation.

The image reconstruction interpretation of MaxEnt, which rests on the existence of a finite improvement path linking initial and final states, suggests that the use of relative entropy in the prior increases the flexibility of the methodology compared to Shannon (1948) entropy suggested in Foley (1994). This is because prior knowledge of the initial condition and of the fact that the initial and final states are linked by a finite path reduces the uncertainty of the observer with respect to the final state. This should be reflected in the entropy measure uncertainty by the inclusion of a correction term for this prior knowledge, embodied in m. A specific aspect of this, raised in Barde (2012), is that the model term m should reflect the length of the finite improvement path. If the improvement path is known to be short, the model should be strongly peaked around the initial condition. Conversely, if the path is long, the model should be flatter, reflecting the fact that the initial condition is no longer informative as to the final equilibrium.

In Barde (2012) the use of MaxEnt image reconstruction as a prediction methodology is investigated by applying it to the Schelling (1969, 1971) model of segregation. This empirical application was chosen specifically because for a given set of parametrisations the Schelling model is known to possess the finite improvement property, where every initial condition leads to a Nash equilibrium in a finite number of steps. The Schelling model thus provides an ideal setting for illustrating the image reconstruction interpretation detailed above.

The purpose of this companion paper is to investigate the MaxEnt methodology further, by attempting to predict the outcome of two agent-based models with recruitment, the Kirman (1993) model of ants and the Abrams and Strogatz (2003) model of language competition, a type of voter model. In both of these models there exists different populations, and agents within them can be recruited, *i.e.* convinced to switch group, by social pressure from members of other groups. As a result, the growth of one type of population depends on the size of the other populations. This setting is more complicated to predict than that of the Schelling model, as population sizes are not constant and a final absorbing state may not even exist. Nevertheless, our first central finding is that MaxEnt can predict the evolution of these models. The second

important finding, which stems from the image reconstruction interpretation of MaxEnt presented above, is the confirmation that the width of the underlying model m does indeed play a role in controlling the time horizon of the MaxEnt prediction, which strongly supports the use of relative entropy rather than absolute Shannon (1948) entropy.

The remainder of the paper is organised as follows. Section 1 first presents the Kirman model of ant recruitment and then investigates the effectiveness of the MaxEnt methodology. Section 2 does the same on the voter model, and finally section 3 concludes.

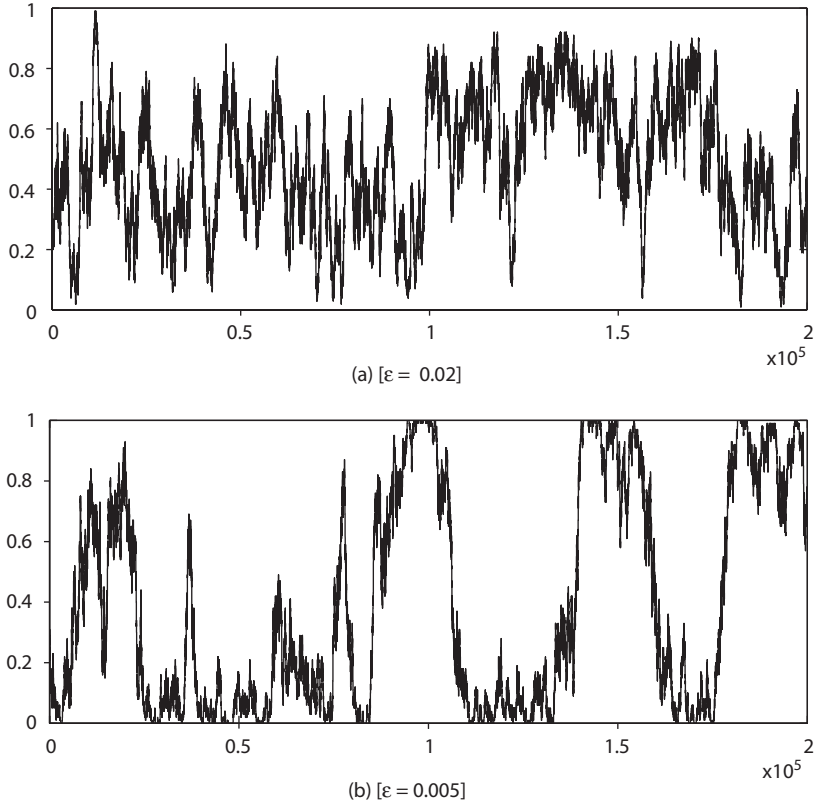
1. The one-dimensional problem: Kirman's model of ants

The Kirman (1993) model of ant recruitment was initially developed to provide a theoretical explanation for a curious empirical puzzle in an experiment involving ants feeding from two different food sources. In the wild, ants that encounter a food source recruit other ants, quickly causing a large amount of ants to feed from that source. The experimental puzzle came from the "cascading" behaviour exhibited by the ants, where most ants used a single source of food for a period of time, and then suddenly switched to the other in a very short period. The central advantage of starting with this model is that because of the simple recruitment process that governs the evolution of the system, the limit distribution of the system is well known, which facilitates the process of verifying the improvement in prediction brought by the use of the MaxEnt methodology.

1.1. Kirman's model of ants

In this model, two sources of food are available to a group of N ants, which are denoted "black" and "white". Describing the state of the system is simple: at any point in time, let $k \in \mathrm{N}$ be the number of ants feeding from the black source, with the remaining $N - k$ ants feeding from the white source. In the following discussion, it will be convenient to refer to $x = k/N$ as the share of the ant population feeding at the black source, with $1 - x = (N - k)/N$ the share feeding from the white source. By extension, we will refer directly to the color of the ant as identifying the food source it uses.

Figure 1. Time evolution of the share of black ants x



As pointed out in Kirman (1993), ants can change color over time, either spontaneously (by making a mistake, for instance), or because they are recruited by an ant of the other color. Because of this, the system will evolve over time. If ε is the probability of an ant spontaneously changing color and $1 - \delta$ the probability of an encounter between two differently coloured ants leading to a successful recruitment, then the dynamic evolution of the system is governed by the following probabilities:

$$\begin{cases} p_{w\to b}(k) = \left(1-\dfrac{k}{N}\right)\left(\varepsilon + (1-\delta)\dfrac{k}{N-1}\right) \\ p_{b\to w}(k) = \dfrac{k}{N}\left(\varepsilon + (1-\delta)\dfrac{N-k}{N-1}\right) \end{cases} \tag{2}$$

At every point in time, the number of black ants k can therefore jump, either $k \to k + 1$ with probability $p_{w\to b}$, or conversely $k \to k - 1$ with probability $p_{b\to w}$. One can see that for large values of the ant population N, the share of black ants $x = k/N$ can be approximated by a continuous interval [0,1], which allows us to rewrite the transition probabilities (2):

$$\begin{cases} p_{w\to b}(x) = \varepsilon(1-x) + (1-\delta)x(1-x) \\ p_{b\to w}(x) = \varepsilon x + (1-\delta)x(1-x) \end{cases} \tag{3}$$

As is the case in Kirman (1993), we assume as a simplification that $\delta = 2\varepsilon$, which allows the simulations to depend on a single parameter. Figure 1 shows the evolution of the state of $N = 100$ ants over time for the two main parametrisations of ε used by Kirman (1993). Of the two, the second case, where $\varepsilon = 0.005$ is the most interesting, as it displays the cascading transitions mentioned previously.

The research question initially addressed by Kirman (1993) was to find the limit distribution $\mu(k)$ of the proportion of time the system spends in a state k. We show below that the MaxEnt methodology can not only replicate this finding, but in fact provide a more general prediction of the time-density of the system for any number of steps τ.

1.2. Prior model and likelihood

We start by specifying a model $m(x, \tau)$ for the relative entropy term in (1). As stated in the introduction, one would intuitively expect this to change depending on the desired time-horizon of the prediction. For low values of τ (short horizons), one would expect the model to be peaked around the initial condition x_0. Conversely, for large values of τ (long horizons) the system will be able to explore large areas of the state space, and the model should be flatter.[3] This movement away from the initial condition x_0 after τ steps is modeled by the diffusion process of 1-dimensional stopped random walk. Given the transition probabilities (3), at any point in time the probability that a jump occurs is $p_j(x) = p_{w\to b}(x) + p_{b\to w}(x)$, while the system remains unchanged with probability $1 - p_j(x)$.

3. As pointed out in Barde (2012), if the system is ergodic, then in the limit $\tau \to \infty$, the model should be a uniform distribution, as all states become accessible.

$$p_j(x) = \varepsilon + 2(1-\delta)\, x\,(1-x) \tag{4}$$

Comparing equations (4) and (3), one can see that in the limit $\varepsilon \to 0$, we have $p_{w\to b}(x) = p_{b\to w}(x) = p_j(x)/2$. This also holds for all values of ε if $x = 0.5$. As a result, the diffusion away from a known initial state x_0 after τ units of time have elapsed can be approximated by a one-dimensional random walk where $p_j(x_0)\,\tau$ jumps occur, each of which takes values $\{+ 1/N, - 1/N\}$ with equal probability. The standard result for such a process is that the probability at time τ of having moved a specified distance is given by a binomial distribution with probability parameter 1/2. Because the prediction $\mu(x, \tau)$ relates to the predicted share of time system will spend in each state, the model is obtained by averaging the binomial density over the expected $p_j(x_0)\,\tau$ jumps for each value of x[4]

$$m(x,\tau) = \frac{\sum_{T=0}^{T=\tau p_j(x_0)} 2^{-T} \binom{T}{\frac{T + N|x - x_0|}{2}}}{\tau p_j(x_0) + 1} \tag{5}$$

Clearly this is an imperfect representation of the diffusion generated by the recruitment process (3), as one can immediately see from table A in appendix 5 that after $\min(k_0, N - k_0)$ jumps, there is a non-zero probability that the random walk process has gone beyond the [0,1] bound for x. This is because the model (5) assumes that probability of a jump $p_j(x_0)$ is constant, and the probability of a positive and negative jump is always equal to 1/2. This is not the case in the actual process (3), as the transition probabilities adjust to guarantee the process remains within the bounds. It will be shown, nevertheless, that this simple random walk diffusion (5) provides a reliable model for the MaxEnt prediction.

The likelihood term for the MaxEnt program (1) can be obtained from the net transition probabilities. The intuition is that the transition probabilities (3) provide a stochastic growth process for each population, which can be integrated to provide an expected time path. However, given the fixed overall number of ants, it must be that these expected time paths of both populations

4. The specification used for calculating the time-average of the binomial density is explained in appendix A.

cancel out for the limiting distribution. As mentioned previously, an attractive aspect of the ants model is that the transition probabilities in fact allow for direct derivation of the limit distribution as $\tau \to \infty$. A particularly elegant derivation is provided in the appendix of Alfarano and Milakovic (2009), which uses a model of herding that is very similar to Kirman (1993).

As shown by Alfarano and Milakovic (2009), deriving the limit distribution directly requires a second order approximation of the transition process, using a drift and a diffusion term obtained through a Taylor expansion of the transition process.[5] The likelihood obtained below uses only the drift term, *i.e.* a first-order approximation of the process. Given the number of black ants k, the share of black ants x, the transition probabilities (3) lead to the following expected state after a jump:

$$(k+1)\,p_{w\to b}(x) + (k-1)\,p_{b\to w}(x) + k\left(1 - p_j(x)\right) = k + \varepsilon(1-2x) \quad (6)$$

Assuming that an interval of time $[t, t+1]$ is short enough that a only single jump is expected to occur, this expected jump directly determines the expected change in the share of black agents x:

$$E\left[x_{t+1}\right] - x_t = \frac{\varepsilon}{N}(1-2x_t) \quad (7)$$

One can see from this expression that assuming a minority of black ants ($x < 1/2$) the expected change in the share of black ants is positive. Conversely, if $x > 1/2$ and a majority of ants are black, one would expect to see the share of black ants k fall. Thus, the expectation is that the transition probabilities will bring the state towards $x = 1/2$ over time.[6] Dividing on both sides by x gives the expected growth rate of the black ant population during over the time interval:

$$\frac{E\left[x_{t+1}\right] - x_t}{x_t} = \frac{\varepsilon}{N}\left(\frac{1}{x_t} - 2\right) \quad (8)$$

5. This is outlined in appendix B.

6. Expression (7) helps to clarify the simplifying assumption that $\delta = 2\varepsilon$. Because in (3) the probability of a black ant recruiting a white ant is equal to the probability of a white ant recruiting a black ant, the expected effects cancel out and δ does not enter the expected change in population (6).

As we are assuming a large ants population N, the share of black ants $x \in [0, 1]$ can be treated as a continuous variable. As a result, the left hand side of this expression can be expressed as the time-derivative of a logarithm, $\partial \ln x/\partial t$, leading to an ordinary differential equation. For τ units of time, the expected time path of the black ant population is approximated following expression, where $\mu(x, \tau)$ is the share of time spent in state x and $\ln x$ is the expected time path in that state:

$$\sum\nolimits_x \mu(x,\tau) \ln x \tag{9}$$

Given the inherent symmetry of the system, it is possible to obtain a similar expression involving $\mu(x, \tau)\ln(1 - x)$ for the expected time path of the white ant population.[7] Furthermore, because the overall number of ants is fixed at N and growth of one population implies an equivalent reduction in the other, the sum of the two time paths should cancel out. This is used for formulate the following likelihood for a candidate prediction μ:[8]

$$\ell\left(\mu\left(x,\tau\right)\right) = \sum\nolimits_x \mu(x,\tau)\left(\ln x + \ln\left(1-x\right)\right) \tag{10}$$

Given the model (5) and likelihood (10), the MaxEnt program for the share of time τ spent in state x in the ants recruitment model is given by:

$$\max_{\mu}\left(\alpha S\left(\mu\left(x,\tau\right)\middle|m\left(x,\tau\right)\right) + \ell\left(\mu\left(x,\tau\right)\right)\right) \tag{11}$$

The first order condition of (11) provides the predicted value of $\mu(x, \tau)$. As explained in Barde (2012), one can see that this is effectively a mixture density between the model (5) and the limit distribution, which in this case is a symmetric beta distribution:

$$\mu\left(x,\tau\right) \propto m\left(x,\tau\right) x^{-\frac{1}{\alpha}} \left(1-x\right)^{-\frac{1}{\alpha}} \tag{12}$$

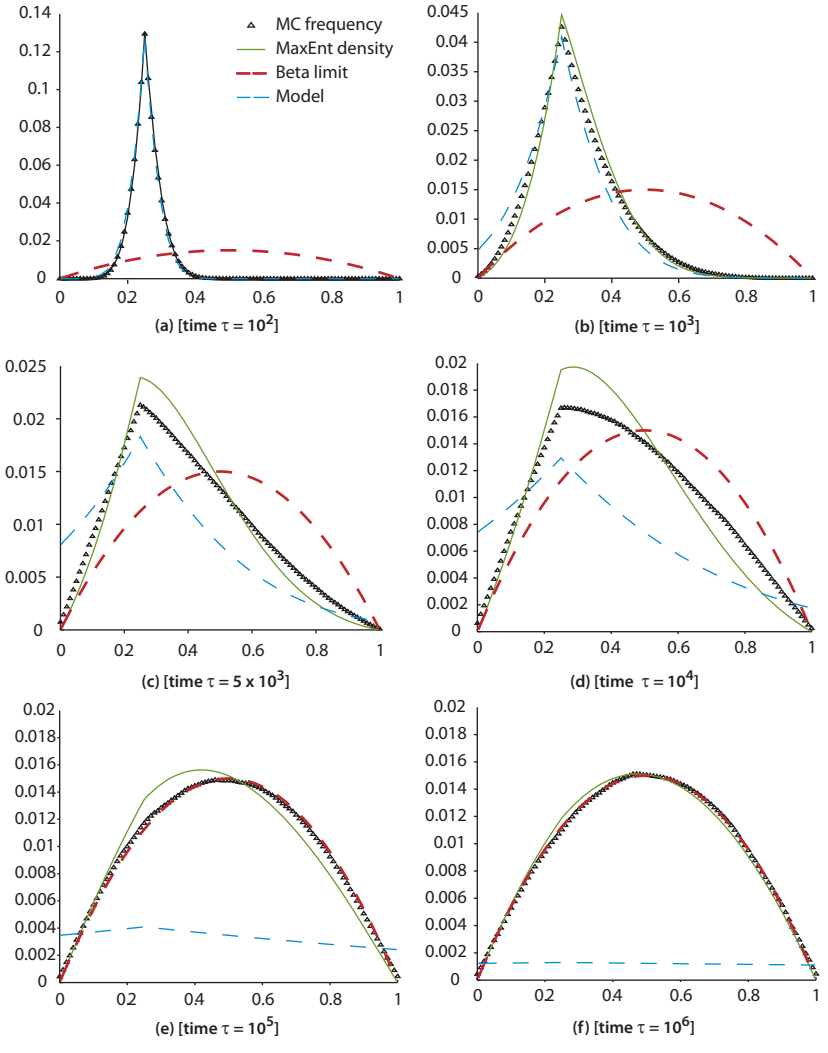
The alpha parameter in (12) is effectively the Lagrange parameter from the maximisation problem (11), and controls the relative weight of the entropy and likelihood terms. In this case,

7. In fact, there is no a priori reason for the state of the system to be measured using the number of black ants k. Intuitively, using instead the number of white ants $N - k$ as the state variable should not change the predictions that an observer can make about system.

8. One can see that this expression exhibits the important properties of a log-likelihood: given a candidate distribution μ and $x \in [0, 1]$ it will be negatively valued, and for the limit beta distribution obtained below, it reaches a maximum value of zero.

one can see that as $\tau \to \infty$ and the model term $m(x, \tau)$ becomes a uniform distribution, the value of α also controls the exponent of the limit beta distribution. As shown in appendix B, the limit distribution is is simply $x^{N\varepsilon-1}(1-x)^{N\varepsilon-1}$, which corresponds with the one identified in Kirman (1993). We therefore set $\alpha = -1/(N\varepsilon - 1)$.

Figure 2. Time evolution of the MC frequencies vs. MaxEnt prediction, $x_0 = 0.25$, $\varepsilon = 0.02$



1.3. Maximum entropy prediction of the ants model

The reliability of the mixture density (12) in predicting the proportion of time τ the system spends in each state x is assessed by comparing the MaxEnt prediction to a Monte-Carlo simulation of the system defined by the transition probabilities (3), using the same values of ε as for figure 1. In both cases, $N = 100$ and the initial condition was set to $x_0 = 0.25$, in order to explicitly examine how the system moves from an asymmetric initial condition to its symmetric limit distribution.[9]

In both figures 2 and 3, the Monte-Carlo frequencies are represented by the sequence of triangular markers, while the solid line represents the MaxEnt prediction (12). As an illustration of how the mixture density is reached, the dashed lines represent the components of this prediction, with the thin dash representing the model (5) and the thicker dash showing the limit beta distribution. Goodness-of-fit statistics are displayed in table 1 for both parametrisations, and report the Spearman rank correlation ρ and mean square error of the prediction (12) relative to the variance of the Monte-Carlo frequency of the for each of the time steps in the figures.

The first case, shown in figure 2, uses the setting that produced the path shown in figure 1a. As expected, in the early stages of the process, for low values of τ, the prediction is dominated by the model term (5), and over time it gradually converges to the limit beta distribution. For intermediate values of τ (particularly $\tau = 5 \times 10^3$ and $\tau = 10^4$), one can see that the empirical frequencies are converging towards the limit distribution faster than suggested by the prediction. Nevertheless, the mean square error of the prediction relative to the variance of the Monte-Carlo frequencies remains low even for these intermediate values of τ. Furthermore, an important aspect is that the prediction successfully captures the asymmetry in the empirical frequencies about x_0 whenever the initial condition is not located at 1/2.

---

9. The number of Monte-Carlo iterations R carried out for a given time-horizon τ is $R = 10^8/\tau$. The implication is that the resulting time-averages in figures 2 and 3 are all calculated over 10^8 time steps, the only difference between sub-figures being that the system is essentially reset to the initial condition x_0 every τ steps. This this is done to ensure that the goodness-of-fit statistics, which are based on the variance of the Monte-Carlo frequencies, are comparable across time horizons τ.

Figure 3. Time evolution of the MC frequencies *vs.* MaxEnt prediction, $x_0 = 0.25$, $\varepsilon = 0.005$

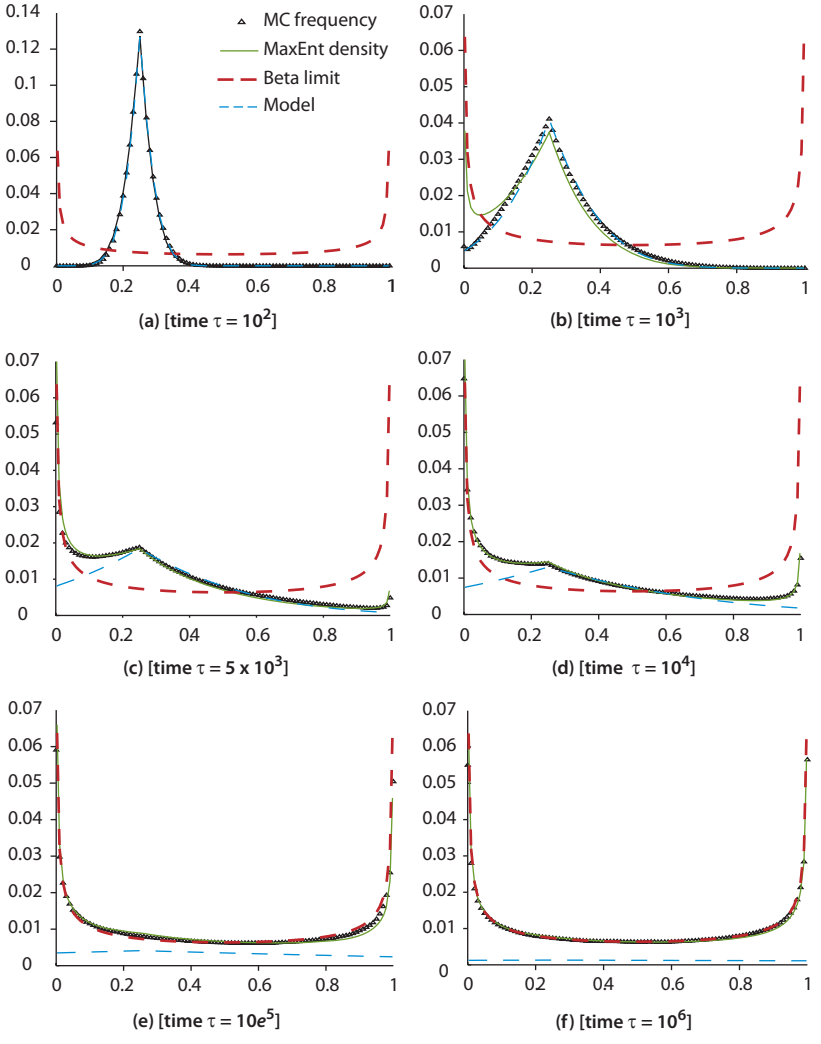


Table 1. Goodness of fit, MaxEnt prediction *vs.* MC frequencies

		$\tau = 10^2$	$\tau = 10^3$	$\tau = 5x10^3$	$\tau = 10^4$	$\tau = 10^5$	$\tau = 10^6$
ε = **0.02**	Spearman ρ	0.9418	0.9985	0.9975	0.9969	0.9537	0.9917
	p-value	$5.74x10^{-49}$	0	0	0	0	0
	MSE/var	0.0011	0.0154	0.0669	0.1183	0.0553	0.0092
ε = **0.005**	Spearman ρ	0.9333	0.9716	0.9923	0.9989	0.8816	0.9523
	p-value	$3.96x10^{-46}$	0	0	0	0	0
	MSE/var	0.0015	0.1378	0.0664	0.0098	0.04	0.0224

The second case, in figure 3, corresponds to the cascading path in figure 1b. As for the previous setting of ε, the goodness-of-fit is shown in the second set of rows in table 1. The qualitative behaviour is similar, with the model component of the prediction dominating in the early stages and a convergence to the limit distribution for large values of τ. Again, the predicted and Monte-Carlo frequencies deviate slightly for intermediate values of τ ($\tau = 10^3$ in this case), but the relative mean square error is low and as for the previous case the methodology correctly predicts the asymmetry involved in shifting from an early distribution that is practically symmetric about x_0 to a limit distribution that is symmetric about 1/2.

Two important observations stem from these results. The first is that the MaxEnt methodology produces a good prediction of the proportion of elapsed time τ the system will spend in each state x, for all values of τ, and not just in the limit $\tau \to \infty$. This provides an improvement with respect to Kirman (1993), as it allows a description of all the phases of the adjustment from an initial condition to the limit distribution. The MaxEnt prediction is even able to capture the asymmetry in the distribution caused by the adjustment from an arbitrary distribution to the symmetric limit. While this adjustment to the limit distribution can also be obtained using a traditional Monte-Carlo approach, the MaxEnt result can be obtained at a greatly reduced computational cost.

The second observation, which results from this ability to predict over a wider range of time horizons, is that when observed over short horizons the ant recruitment process (3) behaves very much like a simple random walk (5). Indeed the early stages of the

adjustment, the model term dominates the mixture distribution completely. It is only over longer time horizons that it converges to the limit distribution identified in Kirman (1993). This suggests that in practical terms it might be very difficult to distinguish recruitment processes from a random walks over short time horizons by looking only at the time density of states.

2. The two-dimensional problem: Voter models

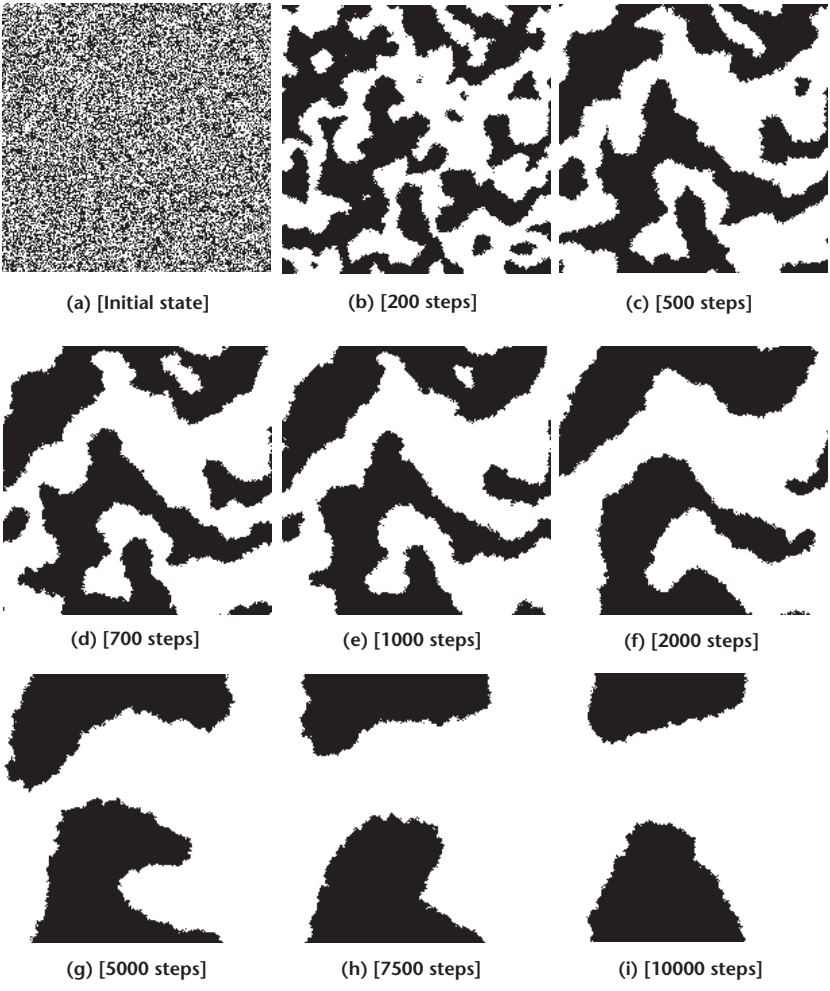
Voter models, also known as consensus models, are similar in spirit to the ants model seen in the previous section. Typically, several populations coexist in the same space and members of each group attempt to convince members of competing groups to switch over, much like the recruitment process described above. One of the main attractions of these models is that it is straightforward to integrate localised spatial effects that are similar to those in the Schelling (1971) model, which provides a further setting for investigating where the predictive power of the MaxEnt methodology.

2.1. The Abrams-Strogatz model of language competition

In the Abrams and Strogatz (2003) model of language competition, two languages, W and B are spoken within a population, and individuals switch from one language to the other according to its attractiveness. Assuming, as was the case for the ants recruitment model, that x is the share of individuals speaking language B and $1 - x$ is the share speaking W, the transition probabilities (13) are determined by three elements. The first is the intrinsic prestige of the languages, controlled by a parameter $s \in [0, 1]$ for B and $1 - s$ for W. The second element is the effect of social pressure, as measured by the shares x and $1 - x$ of individuals speaking the language, and the third is a volatility parameter a which increases or reduces the effect of social pressure through exponentiation of this social pressure term:

$$\begin{cases} p_{w\to b}(x) = sx^a \\ p_{b\to w}(x) = (1-s)(1-x)^a \end{cases} \tag{13}$$

Figure 4. Simulated time evolution of a voter model, $s = 0.5$, $a = 2$



As a result of these transition probabilities, the dynamic evolution of the system is given by:

$$\frac{dx}{dt} = (1-x)p_{w\to b}(x) - xp_{b\to w}(x) = x(1-x)\left(sx^{a-1} - (1-s)(1-x)^{a-1}\right) \quad (14)$$

The equilibrium predictions are qualitatively similar to the ones obtained in Kirman (1993) and shown in figures 2 and 3. The system displays one interior equilibrium $0 < x < 1$ and two corner

solutions at $x = 0$ and $x = 1$.[10] In the low volatility case, where $a > 1$, the corner solutions are stable and the interior solution is unstable, while the opposite is true of the high volatility case $a < 1$.

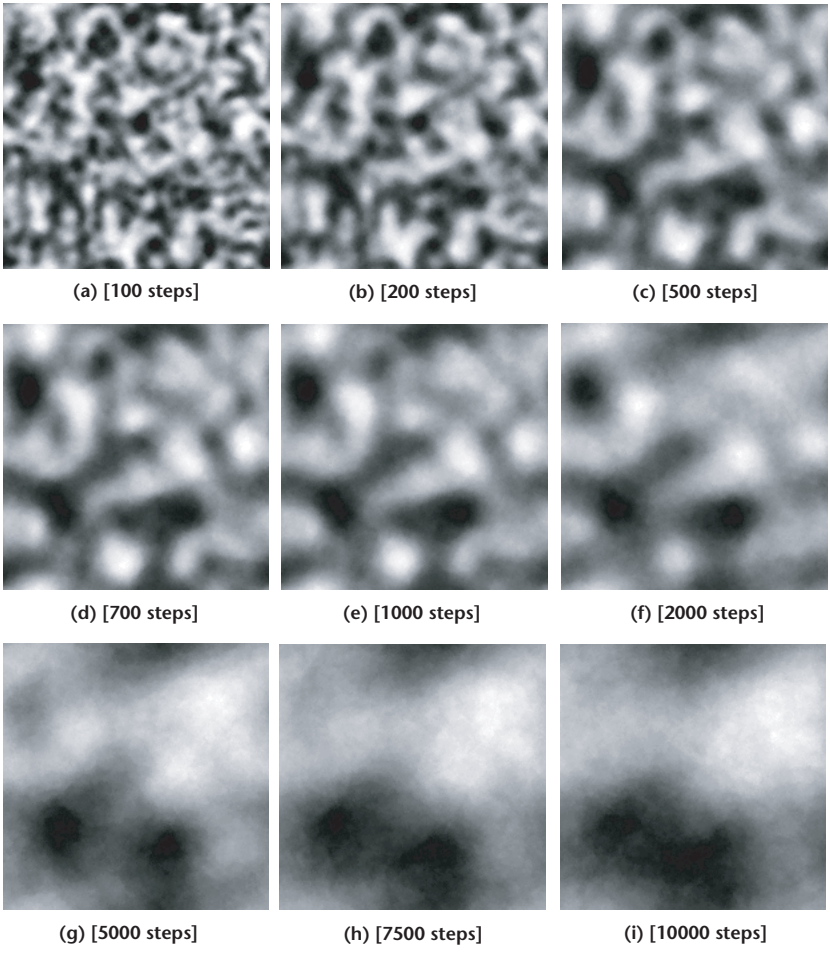
There are several key differences, however, compared to the ant recruitment model in section 1. These are due to the assumption that individuals cannot accidentally switch languages, as was the case in the ants model through the ε term in the transition probabilities (3). As a consequence, once a corner equilibrium $x = 0$ or $x = 1$ is reached in the voter model, the system will remain in that state, which is not the case in the ant model, as is visible in Figure 1b. This clarifies why in the voter model the discussion centers on the stability or instability of the three defined equilibria, while the ant model focuses instead on the share of total time spent in each state. The absence of a ε term allowing accidental individual switching also explains the difference in the parameter that controls the amount of social interaction and thus the stability of equilibrium or shape of the distribution. In both models, the probability of an interaction between agents of different colour is given by $x(1 - x)$. In the ants model, this is translated into a switching probability through the additive ε term, while in the voter model this is done by exponentiating the interaction probability with $a - 1$.

In an important analysis of this model of language competition, Stauffer *et al.* (2007) show that simulations carried out with a finite number of individuals produces different results compared to the continuous equations shown above.[11] Furthermore, they show that local interaction matters for understanding the dynamics of the system and the time until a stable equilibrium is reached. Local interaction is defined as a situation where the social pressure on any given agent to switch from language B to W comes from the share of the agent's direct neighbours that already speak W rather than the share of the overall population that speaks W, as is the case in (14).

10. The interior equilibrium is located at 1/2 in the case of two equivalent languages, with prestige $s = 1 - s = 1/2$. This is the parameter value that was used in the simulation reported below.

11. In Stauffer *et al.* (2007) the continuous version is referred to as a "mean-field approximation".

Figure 5. Time evolution of voter model state frequencies, 200 MC iterations



Stauffer *et al.* (2007) focus on the case where the languages have the same prestige ($s = 1 - s = 1/2$) and the volatility is low ($a > 1$), therefore this is the parameter setting that will be used below to investigate the effectiveness of the MaxEnt predictions. In the following simulation there are $N = 40000$ agents arranged in a 200 x 200 lattice. The neighbourhood which determines the local social pressure to switch language is a 3 x 3 square centered on the agent of interest. As was the case in the analysis of the Schelling model in Barde (2012), the space occupied by the agents is assumed to be toroidal, which means that localised neighbou-

rhood effects can be calculated directly by applying a N x N circulant matrix A to the state vector, a N x 1 vector recording the language spoken by each agent. This implies that x, the share of agents speaking B and $1 - x$, the share speaking W in the transition probability (13) is a local variable that is determined by the 8 closest neighbours of an agent rather than a global variable, averaged over the overall population.

Figure 4 presents the simulated state of such a system at several points in time, starting from a random initial condition where half the agents speak B and half speak and W. One can see that starting from a dispersed state in the initial condition, relatively few steps suffice for distinct clusters to appear.[12] As the number of steps is increased, the smaller clusters tend to disappear and the interfaces between the large clusters of different colours tend to smooth out, a process which Stauffer *et al.* (2007) refer to as an increase in the surface tension of the system. Another characteristic outlined by their analysis is the existence of long-lived meta-stable equilibria. This is visible in the last few panels of figure 4, where the two clusters remain similar over a large number of steps. In the limit, however, the system always ends up in one of two absorbing states, $x = 0$ or $x = 1$.

2.2. Maximum entropy prediction of the voter model

The MaxEnt methodology used to predict the evolution of the state of the voter model is broadly similar to the one used for the analysis of the Schelling (1971) model of segregation carried out in Barde (2012). First of all, a Monte-Carlo analysis was run in order to obtain a point of comparison for the MaxEnt prediction. 400 random initial conditions were drawn and for each of these 200 separate simulations were run, replicating the process shown in in figure 4. As an illustration, figure 5 shows the result of running 200 such simulations on the initial condition provided in figure 4a. Each sub-figure shows the share of runs in which an agent is in state B after the specified number of steps.

The setting of the voter model implies that agents do not have ex-ante preferences for a language, and only determine their state

12. As is the case in Stauffer *et al.* (2007) one step in time corresponds to one update opportunity for all N agents to update their state.

relative to the language spoken by neighbouring agents. As a result, assuming that languages *B* and *W* have equal status and initial populations have equal size, the initial state of any given agent *i* does not provide any information that can be used to directly provide a model for the relative entropy term (1). The best guess an observer might make is that $m_i^W = m_i^B = 0.5$, which is the uninformative uniform distribution. This situation is effectively the same as in the Schelling (1971) model, in which agents do not have any intrinsic preference for a particular location, and the attractiveness of a location to an agent is only determined by the state of the agents around that location. As was the case in Barde (2012), this is dealt with by using the following double-space entropy, which measures the information entropy of a message revealing the state of two randomly picked agents *i* and *j*:

$$S\left(\mu_i, \mu_j \middle| m_{i,j}\right) = \frac{1}{N}\left(-2\sum_i \sum_L \mu_i^L \ln \mu_i^L + \frac{1}{N}\sum_{i,j}\sum_L \mu_i^L \mu_j^L \ln m_{i,j}^L\right) \quad (15)$$

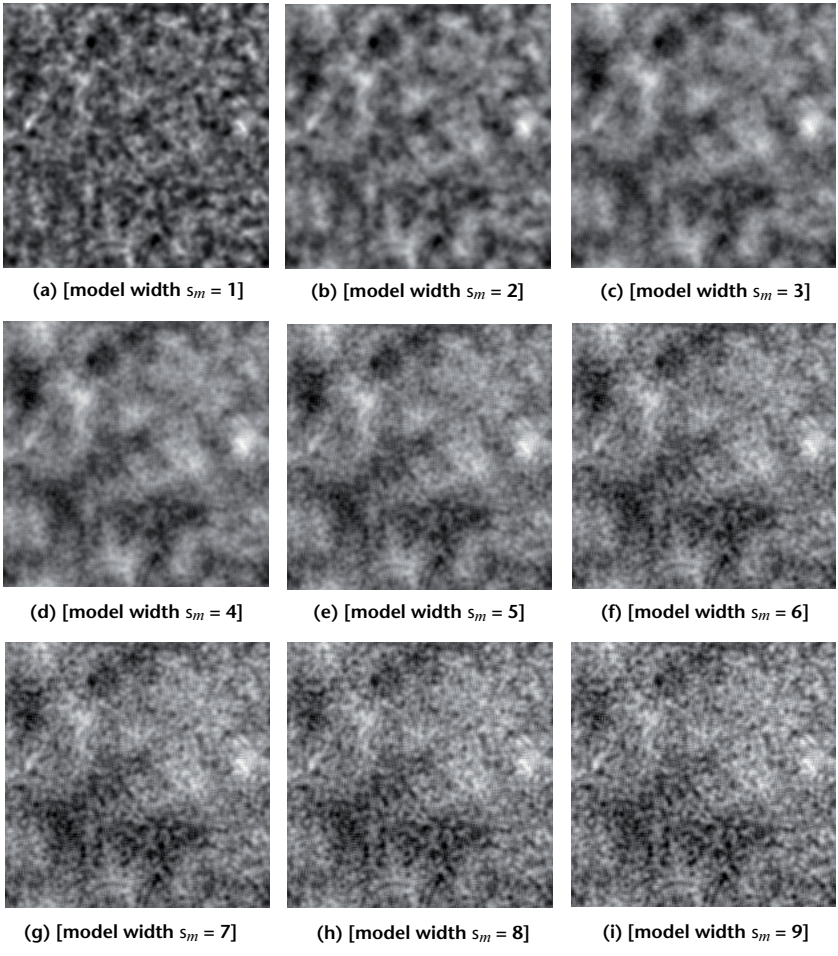
The use of a double-space entropy (15) allows the model to encode correlations across agents: if two agents *i* and *j* are located close to each other, the probability that they both speak a given language *L* is higher than if they are far from each other.[13] The model $m^L_{i,j}$, which models the probability that *i* and *j* both speak language *L* is assumed to be a normal distribution over the distance between agents, as was the case in the analysis of the Schelling model in Barde (2012). Because the width of this normal distribution is determined by its standard deviation σ_m, this parameter directly controls the distance over which agent decisions are likely to be correlated.

The likelihood component of (1) that is used to generate the prediction is assumed to be Gaussian, following again the methodology used in Barde (2012). This effectively measures the similarity be the data available in the initial condition and the MaxEnt prediction for each of the two languages *L*:

$$\ell\left(d^L \mid \mu^L\right) = -\sum_i \frac{\left(\left(A\mu^L\right)_i - d_i^L\right)^2}{\left(\sigma\right)^2} = \frac{-\chi^2\left(\mu^L\right)}{2} \quad (16)$$

13. The reader is referred to the appendix of Barde (2012) for a derivation of double space entropy.

Figure 6. MaxEnt state density prediction of voter model



(a) [model width s_m = 1] (b) [model width s_m = 2] (c) [model width s_m = 3]

(d) [model width s_m = 4] (e) [model width s_m = 5] (f) [model width s_m = 6]

(g) [model width s_m = 7] (h) [model width s_m = 8] (i) [model width s_m = 9]

Here A is the N x N symmetric adjacency matrix, with entries 0 or 1 in the i^{th} row indicating which J agents are neighbours to i. The data d_i^L used in the comparison is given by the initial social pressure $d^L = A{\mu_0}^L$, where ${\mu_0}^L \in \{0, 1\}$ is the vector indicating the language spoken by each agent in the initial condition, taken from figure 4a. It is important to point out that given the choice of parameterisation $a = 2$, the local interaction term $A\mu^L$ in the likelihood is in fact a linear approximation of the social pressure term in the transition probabilities (13).[14] It will be shown that this simplification nevertheless produces good predictions of the Monte-Carlo frequencies.

Given the relative entropy (15) and likelihood (16), the MaxEnt program is given below. Its solution, displayed in figure 6 for increasing values of σ_m (and therefore increasing model widths) is obtained numerically using the image reconstruction algorithm of Skilling and Gull (1991), modified in Barde (2012) to predict the outcome of the Schelling model.

$$\max_{\mu}\left(\alpha S\left(\mu_i,\mu_j\middle|m_{i,j}\right)+\ell\left(d^L\mid\mu^L\right)\right) \tag{17}$$

Figure 6 illustrates the MaxEnt predictions obtained with the initial condition provided in figure 4a for various values of σ_m. These are visually comparable with the Monte-Carlo frequencies in figure 5, also generated from the same initial condition. In addition, the Spearman rank correlation and mean-square-error relative to the variance of the Monte-Carlo frequencies were calculated for each of the 400 sets of predictions and frequencies. The resulting means and standard deviations are presented in table 2. The bold entries in each column indicate identify the value of σ_m (row) that best fits the Monte-Carlo frequencies for the relevant number of steps. The diagonal pattern made up by these bold entries indicates that the Monte-Carlo frequencies of the system at successively higher time steps are, up to a point, better predicted by successively wider models. As was the case for the MaxEnt prediction of the ants recruitment model in section 1.3, this supports the suggestion that the width of the model in the relative entropy term, determined in this case by σ_m, controls the time-horizon of the MaxEnt prediction.

A further observation that can be made from table 2, however, is that the predictive power of the MaxEnt methodology falls as the width of the model is increased. Indeed, the bold entries in each column show a gradual reduction in the correlation coefficient ρ and an increase in the size of the mean-square error as the number of steps is increased, coupled with a widening of the standard deviations around the means of the two statistics.

---

14. This is intended as a simplification: using $d^L = (A\mu_0{}^L)^2$ and $(A\mu^L)^2$ in (16) produces a Hessian matrix for the MaxEnt methodology where all N x N entries are non-zero, requiring an intractable amount of storage and computation time. Using instead the linear approximation produces a Hessian matrix that is basically A x A, and therefore can be stored and manipulated efficiently using sparse matrix algorithms.

Table 2. Goodness of fit, predicted vs. MC voter model state densities (standard deviations in parenthesis)

		Steps								
	σm	100	200	500	700	1000	2000	5000	7500	10000
1	ρ	0.859	0.740	0.562	0.498	0.434	0.325	0.215	0.179	0.157
		(0.006)	(0.011)	(0.017)	(0.019)	(0.021)	(0.024)	(0.026)	(0.026)	(0.027)
	MSE	0.291	0.518	0.865	0.992	1.120	1.340	1.561	1.635	1.678
		(0.011)	(0.020)	(0.033)	(0.037)	(0.042)	(0.049)	(0.052)	(0.052)	(0.054)
2	ρ	**0.950**	**0.921**	0.810	0.752	0.685	0.548	0.383	0.323	0.287
		(0.004)	(0.007)	(0.016)	(0.020)	(0.025)	(0.034)	(0.042)	(0.045)	(0.047)
	MSE	**0.114**	**0.167**	0.376	0.487	0.618	0.888	1.217	1.337	1.411
		(0.007)	(0.012)	(0.028)	(0.036)	(0.046)	(0.065)	(0.082)	(0.088)	(0.094)
3	ρ	0.887	0.918	**0.898**	0.869	0.826	0.713	0.538	0.465	0.417
		(0.009)	(0.008)	**(0.014)**	(0.018)	(0.024)	(0.036)	(0.052)	(0.059)	(0.064)
	MSE	0.228	0.170	**0.205**	0.259	0.340	0.557	0.901	1.047	1.142
		(0.016)	(0.013)	(0.023)	(0.031)	(0.041)	(0.066)	(0.101)	(0.115)	(0.127)
4	ρ	0.813	0.861	0.890	**0.884**	**0.865**	0.791	0.638	0.564	0.513
		(0.015)	(0.014)	(0.016)	(0.019)	(0.023)	(0.035)	(0.056)	(0.066)	(0.074)
	MSE	0.366	0.274	0.217	**0.228**	**0.263**	0.403	0.698	0.845	0.945
		(0.026)	(0.022)	(0.026)	(0.031)	(0.040)	(0.064)	(0.106)	(0.128)	(0.145)
5	ρ	0.753	0.801	0.850	0.857	0.855	**0.815**	0.694	0.626	0.577
		(0.019)	(0.018)	(0.020)	(0.022)	(0.026)	(0.037)	(0.058)	(0.070)	(0.080)
	MSE	0.480	0.387	0.290	0.276	0.280	**0.355**	0.586	0.719	0.815
		(0.033)	(0.030)	(0.033)	(0.038)	(0.045)	**(0.066)**	(0.109)	(0.134)	(0.153)
6	ρ	0.708	0.747	0.801	0.814	0.821	0.806	0.717	0.659	0.615
		(0.022)	(0.022)	(0.025)	(0.027)	(0.031)	(0.041)	(0.061)	(0.073)	(0.083)
	MSE	0.566	0.489	0.383	0.356	0.341	0.370	0.539	0.653	0.740
		(0.038)	(0.036)	(0.042)	(0.047)	(0.055)	(0.075)	(0.114)	(0.139)	(0.159)
7	ρ	0.675	0.701	0.750	0.766	0.779	0.779	**0.718**	**0.669**	**0.630**
		(0.023)	(0.023)	(0.029)	(0.032)	(0.037)	(0.047)	**(0.065)**	**(0.076)**	**(0.086)**
	MSE	0.630	0.577	0.479	0.447	0.422	0.420	**0.538**	**0.632**	**0.708**
		(0.040)	(0.039)	(0.050)	(0.057)	(0.066)	(0.087)	**(0.123)**	**(0.145)**	**(0.164)**
8	ρ	0.652	0.663	0.702	0.718	0.732	0.742	0.700	0.660	0.627
		(0.022)	(0.023)	(0.033)	(0.037)	(0.042)	(0.054)	(0.071)	(0.081)	(0.090)
	MSE	0.675	0.651	0.571	0.540	0.512	0.491	0.571	0.649	0.714
		(0.040)	(0.041)	(0.059)	(0.068)	(0.078)	(0.100)	(0.134)	(0.154)	(0.172)
9	ρ	0.635	0.632	0.658	0.672	0.686	0.700	0.673	0.641	0.613
		(0.021)	(0.023)	(0.036)	(0.042)	(0.048)	(0.060)	(0.076)	(0.084)	(0.092)
	MSE	0.709	0.712	0.657	0.629	0.601	0.572	0.624	0.687	0.743
		(0.037)	(0.042)	(0.066)	(0.077)	(0.089)	(0.113)	(0.144)	(0.162)	(0.177)

The visual comparison of figures 5 and 6 also supports this: As the model with is increased the prediction gradually becomes more "grainy", to the point where it becomes difficult to distinguish an image. The intuitive conclusion that can be drawn from this is that there is a limit to the time-horizon over which reliable a prediction can be made.

3. Discussion and Conclusion

The analysis of the Kirman model of ant recruitment and the locally-interacting Abrams-Strogatz model of language competition both show that the maximum entropy methodology can be used to predict the state distributions of agent-based models with recruitment, where agents can switch groups based on a measure of social pressure. This provides support for the use of the MaxEnt image reconstruction methodology as a prediction methodology in economics. A first aspect is that MaxEnt can reliably predict the state space of these agent-based models, even in the case where there is no defined final state, as in the Kirman (1993) model of ant recruitment, or in the case where the transition of a given agent is probabilistic rather than a best response, as was the case with the initial MaxEnt analysis of the Shelling model carried out in Barde (2012). A second important aspect is the confirmation of the suggestion made in Barde (2012) that the width of the model term controls the time-horizon of the prediction, by serving as a proxy for the length of improvement path between initial and final state.

In methodological terms, the maximum entropy methodology used here and in Barde (2012) therefore provides a generalisation of the existing applications of MaxEnt in economics, mentioned in the introduction. These typically rely on Shannon (1948) entropy in their analysis, the justification being that this measures the absolute uncertainty of an observer as to the state of the system. As was suggested in Barde (2012) and demonstrated here, this implicity corresponds to using relative entropy with respect to a uniform model m. Given the link between model width and time horizon, this implies that Shannon MaxEnt predicts over a large time horizon only. The predictions obtained here, using relative entropy, carry over a much larger range of time. This is potentially

relevant, as dynamic systems may behave differently over different time horizons.

This last point has potential implications given the suggestion made in Kirman (1993) as to the relevance of recruitment models in economics. Indeed, Kirman suggests that recruitment is pervasive in many markets, in particular financial markets, where individual agents make decisions based not only on objective information, but also based on imitation of surrounding agents. This point is reinforced by the use Alfarano and Milakovic (2009) make of a very similar model to analyse herding behaviour in agent-based finance. Importantly, the results in section 1.3 reveal that even when recruitment is present, such that herding occurs over long horizons of time, it may be nevertheless very difficult to detect this process over short time horizons, as the system will be difficult to distinguish from a standard random walk.

References

Abrams Daniel M. and Steven H. Strogatz. Modelling the dynamics of language death. *Nature*, 424:900, 2003.

Alexis A. Toda, 2010. Existence of a statistical equilibrium for an economy with endogenous offer sets. *Economic Theory*, 45:379–415.

Alfarano Simone and Mishael Milakovic, 2009. Network structure and n-dependence in agent-based herding models. *Journal of Economic Dynamics and Control*, 33:78–92.

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

Cornwell T.J. and K.F. Evans. A simple maximum entropy deconvolution algorithm, 1985. *Astronomy and Astrophysics*, 143:77–83.

Foley Duncan K. A statistical equilibrium theory of markets, 1994. *Journal of Economic Theory*, 62:321–345.

Jaynes Edwin T. Information theory and statistical mechanics i, 1957a. The Physical Review, 106:620–630.

Jaynes Edwin T. Information theory and statistical mechanics ii, 1957b. *The Physical Review*, 108:171–190.

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

Narayan Ramesh and Rajaram Nityananda. Maximum entropy image restoration in astronomy, 1986. *Annual review of astronomy and astrophysics*, 24:127–170.

Schelling Thomas C. Dynamic models of segregation, 1971. *Journal of Mathematical Sociology*, 1:143–186.

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

Shannon Claude E. A mathematical theory of communication, 1948. *The Bell System Technical Journal*, 27:379–423.

Shore John E. and Rodney Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, 1980. *IEEE Transactions on Information Theory*, 26:26–37.

Skilling John and Stephen F. Gull. Bayesian maximum-entropy image reconstruction, 1991. *Spatial Statistics and Imaging*, 20:341–367.

Stauffer Dietrich, Xavier Castelló, Victor M. Eguíluz, and Maxi San Miguel, 2007. Microscopic Abrams-Strogatz model of language competition. *Physica A*, 374:835–842.

Appendix

A. Time density of states in a stopped random walk

The diffusion model used in section 1.2 is a stopped one-dimensional random walk, with $T = \tau p_j(x_0)$ expected jumps of equally probably size $\pm 1/N$. The probability of having moved distance $\pm k/N$ for the first six steps is shown in table A. One can see that these are simply the relevant binomial coefficient divided by two to the power of the number of steps. As a result, the general probability of having moved distance k/N after T steps is given by:

$$p\left(d_T = \frac{k}{N}\right) = \begin{pmatrix} T \\ \frac{T+k}{2} \end{pmatrix} 2^{-T} \tag{A-1}$$

Table A. Diffusion from initial condition in a random walk model

	Distance traveled from χ_0												
Steps	-6/N	-5/N	-4/N	-3/N	-2/N	-1/N	0	1/N	2/N	3/N	4/N	5/N	6/N
0							1						
1						1/2	0	1/2					
2					1/4	0	2/4	0	1/4				
3				1/8	0	3/8	0	3/8	0	1/8			
4			1/16	0	4/16	0	6/16	0	4/16	0	1/16		
5		1/32	0	5/32	0	10/32	0	10/32	0	5/32	0	1/32	
6	1/64	0	6/64	0	15/64	0	20/64	0	15/64	0	6/64	0	1/64

The proportion of the T steps spent at a given distance k/N, needed for the model (5) is then simply the average over the relevant column in table A. The major difference from the standard "Pascal triangle" visible in this table is that given the transition probabilities, even distances can only be reached with an even number of steps, and conversely, odd distances require an odd number of steps. As a result, in order to simplify the calculation of the average over the number of steps, a recurrence rule is developed that links every other entry in a column. This uses the two central recurrence rules for binomial coefficients:

$$\begin{pmatrix} n+1 \\ k \end{pmatrix} = \begin{pmatrix} n \\ k \end{pmatrix} = \begin{pmatrix} n \\ k-1 \end{pmatrix} \tag{A-2}$$

$$\binom{n}{k+1} = \binom{n}{k}\frac{n-k}{k+1} \tag{A-3}$$

Combining the two and rearranging the indexes provides the following recurrence rule:

$$\binom{n+2}{k+1} = \binom{n}{k}\left(2 + \frac{n-k}{k+1} + \frac{k}{n-k+1}\right) \tag{A-4}$$

This rule can be used to link directly adjacent non-zero entries in a given column of table A. By using the formula recursively, one can eliminate the binomial coefficient from the right hand side, and express the binomial coefficient for any T and k as a product of terms generated in the same column for lower values of k, where $\boldsymbol{k} = N|x - x_0|$ represents the absolute number of steps the system has traveled away from the initial condition x_0. This allows us to specify (A-1) as follows:

$$2^{-T}\binom{T}{\frac{T+k}{2}} = \begin{cases} 2^{-k}\prod_{i=0}^{i=\left\lfloor\frac{T-k}{2}\right\rfloor-1}\frac{1}{4}\left(2 + \frac{i}{k+i+1} + \frac{k+i}{i+1}\right) & \text{if} \quad \left\lfloor\frac{T-k}{2}\right\rfloor \geq 1 \\ 2^{-k} & \text{if} \quad 1 > \left\lfloor\frac{T-k}{2}\right\rfloor \geq 0 \\ 0 & \text{if} \quad T-k < 0 \end{cases} \tag{A-5}$$

The first right hand-side element forms the core of the expression. The third element simply states that the number of time steps T is also a strict upper bound on the distance than can be traveled in that time, while the second element states that the probability of being on this upper bound is given by a negative power of 2.[15]

In practical terms one starts by computing for each value of $\boldsymbol{k}$ a vector containing the argument in brackets for all values of $i \in \{0,1,2,\ldots,\lfloor(\tau p_j(x_0) - k/2)\rfloor - 1\}$, using the argument of the product term in (A-5). The cumulative product of this vector provides all the non-zero probabilities (A-1) in the k^{th} column in table A. The sum of these vector entries, divided by $\tau p_j(x_0) + 1$, then provides the required model:

15. These two expressions can be seen directly in table 3: The top sides of the triangle are simply formed by increasing powers of 1/2. Above these, the distribution is not defined.

$$m(x,\tau) = \frac{\sum_{T=0}^{T=\tau p_j(x_0)} 2^{-T} \begin{pmatrix} T \\ \frac{T+N|x-x_0|}{2} \end{pmatrix}}{\tau p_j(x_0)+1} \tag{A-6}$$

B. Fokker-Planck derivation of the limit distribution of the ant model

Alfarano and Milakovic (2009) show, using a Taylor expansion of the step operator formed by the Markov transition matrix of transition probabilities (3) that the following Fokker-Planck equation describes the evolution of the distribution over states:

$$\frac{\partial \mu(x)}{\partial \tau} = -\frac{\partial}{\partial x} A(x)\mu(x) + \frac{1}{2}\frac{\partial^2}{\partial x^2} B(x)\mu(x) \tag{A-7}$$

The drift term $A(x)$ of the equation corresponds to the expected jump size (7):

$$A(x) = N(p_{w\to b} - p_{b\to w}) = N\varepsilon(1-2x) \tag{A-8}$$

The diffusion term $B(x)$ corresponds to the probability of a jump $p_j(x)$ given by (4) and used to model the diffusion process away from the initial condition (5).

$$B(x) = p_{w\to b} + p_{b\to w} = \varepsilon + 2(1-\delta)x(1-x) \tag{A-9}$$

The Fokker-Planck equation (A-7) is a second order differential equation with variable coefficients, and as shown by Alfarano and Milakovic (2009), the general solution is of the following form, where c is a constant of integration that can serve to normalise the probability distribution:

$$\mu(x) = \frac{c}{B(x)} exp\left(2\int_{x'=0}^{x} \frac{A(x')}{B(x')} dx'\right) \tag{A-10}$$

Replacing $\delta = 2\varepsilon$ in (A-9) and assuming, as is done both in (Kirman, 1993; and Alfarano Milakovic, 2009) that $\varepsilon \to 0$ and $N \to \infty$ in such a way that that $N\varepsilon$ remains constant, the diffusion term can be simplified to $B(x) = 2x(1 - x)$. Replacing the drift term $A(x)$ and diffusion term $B(x)$ in the general solution gives:

$$\mu(x) = \frac{c}{2x(1-x)} exp\left(N\varepsilon \int_{x'=0}^{x} \frac{(1-2x')}{x'(1-x')} dx' \right) \tag{A-11}$$

The integral term is equal to ln $x(1 - x)$, which leads to the limit distribution identified in Kirman (1993):

$$\mu(x) \propto x^{N\varepsilon-1}(1-x)^{N\varepsilon-1} \tag{A-12}$$