What questions to ask in order to validate an agent-based model?

Problem presented by

Dr Iqbal Adjali

Unilever

Problem statement

There are two types of influences on consumer behaviour: global, through advertising and price setting, and local, via a network of social imitation. Unilever was most interested in knowing how to detect the presence of social imitation networks in consumer behaviour and how to extend and validate agent-based models to include social imitation.

We proposed to apply recent results and managed to detect the presence of a social imitation network in a dataset of soap-buying by Swiss customers, thanks to a very simplified agent-based model.

Study Group contributors

Damien Challet (University of Oxford)
Andreas Krause (University of Bath)
Eddie Wilson (University of Bristol)

Report prepared by

Damien Challet (University of Oxford)
Andreas Krause (University of Bath)

1 Introduction

The automatic collection of customer transaction data, through either online shops or reward cards, is producing very large databases which contain much information about consumer behaviour. What kind of information and how exploitable it is are very relevant questions. Two approaches are being used. Either one concentrates on individual behaviour and tries to apply various theoretical frameworks and results of the literature on discrete choice, or one uses clustering algorithms in order to determine several classes of customers. The very existence of such categories is likely to be the result of social interactions and influences. The literature on discrete choice cannot easily be generalised to networked interactions, which are known to be widely present in various contexts [4]. Another approach is to use toy models of individual behaviour and concentrate on global, aggregate quantities such as market share or demand fluctuations. This raises the question of how to validate such kind of model, hence the request of Unilever. The latter should also be understood with respect to the contribution of ESGI 2004, where a very sophisticated agent-based model of consumer behaviour was proposed (but not much studied) [1].

2 Results

2.1 ABM validation

Inevitably, validating a toy model is restricted to comparing global quantities produced by the model and real-life data. This should not be taken as a negative statement. Indeed, the aggregation of individual actions of interacting agents is sometimes largely independent from the minute details of individual behaviour, for instance because of sub-dominant contributions of additional terms. The global quantities to be compared should be fixed before designing the model.

In the ideal case, one first designs a simple model and solves or understands it mathematically. One of the essential features is to have a small number of parameters, the maximum acceptable number being around three. Once this model is well understood, one can extend it step by step, extending at the same time the mathematical solution.

The lowest level of validation is the qualitative level, where one is satisfied if the model has features reproducing similar real behaviour. For instance, the unidimensional Ising model has a magnetic phase which disappears if the temperature is high enough; physical systems do display such a transition, but not at the same temperature, and not in one dimension (but do in 2 and 3 dimensions). Nevertheless, the Ising model contains an essential ingredient that gives insights into the microscopic cause of magnetism. At the qualitative level, one also wishes that the signs of the first derivatives of global quantities with respect to the parameters are the same as in reality. If this is the case, it means that one has deep insight into the workings of the problem, which allows us to make predictions on how the system would behave if, for example, a central authority changes a parameter (think of the Tobin tax and its influence on market price fluctuations).

The next level of validation (and success) is the quantitative level, when the model is able to reproduce numerically at least some of the quantities of interest. This happens if there is a region in the parameter space compatible with reality. In this case, not only the sign of the first derivative, but also its value and that of the quantity itself are reasonably close to their real values. As above, one can determine the region in the parameter space that is compatible with reality. This level, rare by nature, brings a vast and most useful understanding of the real system. It is achieved for instance by several physical theories, such as gravitation.

The ultimate aim of modelling is time-series prediction. Before aiming at predicting the time series of a real system, one should first ask oneself to what degree it is possible to predict the behaviour of the model itself. Assume that one knows all the details of the model, except its parameters. Is it possible to determine the latter from a time-series? The case of agent-based models with scalar heterogeneity (*i.e.* each agent differing by a single value, such as their wealth, or their learning rate) is relatively straightforward, since the dynamics are usually described by a set of differential equations, which are known. If the exact dynamical solution of the model is known, one method is to use the maximum likelihood principle. In the other case, one runs several instances of the model with various sets of parameters, and tries to find the set that produces the closest data to the time-series. The reconstruction of the agents' heterogeneity is most often only partly possible, but is still of much value even in the case of models with a complex structure of learning, such as the Minority Game. Determining the probable parameters of the simulation and some sizable fraction of the agents' 'personalities' makes it possible to obtain corridors of predictability several timesteps in advance. When these corridors are narrow, one has a real predictive power.

The connection with real systems is made by a leap of faith: one has to assume that a given system *is* described by the model, and then one reverse-engineers it by the method described above. Once again, one obtains corridors of predictability. This method has been applied with success for example to financial markets [2] and ketchup advertisement [5].

3   Are social networks relevant to soap buying?

Unilever is most interested in the modelling of social influences in online shops and in testing any resulting model against a relatively large set of data that they provided, consisting of $10,000$ consumers and $1,200,000$ transactions.

Therefore, before trying to generalize an already complex model, or before designing yet another ABM with a social imitation network, one should first test the existence of a social network of imitation in the data. We chose to test whether the predictions of a simple model that mixes global and social imitation hold in this case.

The model is defined as follows [3]: there are N customers that have a choice between buying (denoted by $S_i = 1$ for customer i) or not buying ($S_i = -1$). The dynamics of

the choice is

$$S_i(t+1) = \text{sgn}[\phi_i + F(t) + \sum_{j=1}^{N} J_{i,j} S_j(t)]. \tag{1}$$

The three ingredients are

(1) **personal buying propensity** ϕ_i, a real number drawn from a probability distribution with density $p(\phi)$ known or unknown. We shall let R denote the cumulative distribution function of this, as in [3];

(2) **public information, or global pressure** $F(t)$. For instance, $-F$ can be the price, and a variation of F is akin to price change. F can also encode the effect of advertisement and the quality of the product;

(3) **social pressure** $J_{i,j}$ is the adjacency matrix of the social network. If i tends to imitate j, $J_{i,j} > 0$.

The great insight provided by this model is the existence of two universality classes that can be disentangled from the data. Assuming that F varies slowly as a function of time ($F(t) \propto t$) ('slowly' means that the stationary state is always reached), it is possible to show that the collective opinion $\mathcal{O} = \sum_i S_i(t)/N$ plays a central role. In particular, if one fits $d\mathcal{O}/dt$ from several shifts in opinion with the functional form

$$\frac{d\mathcal{O}}{dt} \approx h \exp[-(t - \theta)^2/(2w^2)] + c, \qquad (c \ll h) \tag{2}$$

then there are two possible outcomes:

(1) **No social interaction**: $h \propto 1/w$.

(2) **Presence of a social influence network of mean-field type** ($J_{i,j} = J/N$): $h \propto 1/w^{2/3}$. The exponent $\frac{2}{3}$ is constant for a large variety of distributions $p(\phi)$ and links $J_{i,j}$, which include for instance scale-free network with an exponent of link distribution between 2 and 3.

In detail, the model (the Random Field Ising Model) takes the form

$$\mathcal{O}(t) = \sum_i S_i(t)/N, \qquad S_i(t+1) = \text{sgn}\big(\phi_i + F(t) + J\mathcal{O}(t)\big), \tag{3}$$

and so the individual agents sense their neighbours through the collective opinion $\mathcal{O}$, which is also called the mean field. If F varies slowly enough that equilibrium is reached then we can think of $\mathcal{O}$ varying quasistatically with F, and its equilibrium value is determined by

$$\mathcal{O} = 1 - 2R(-F - J\mathcal{O}). \tag{4}$$

When J is small enough this has a unique solution, varying continuously with F, but for $J > J_c$ (a critical value) there can be multiple equilibria, and hysteresis of $\mathcal{O}$ as F

varies. In the region where J is less than J_c but close to it, there is a universal scaling law taking the form

$$\frac{d\mathcal{O}}{dF} = \frac{1}{\epsilon}\mathcal{G}\left(\frac{F - F_c}{\epsilon^{3/2}}\right), \qquad \epsilon = J_c - J, \tag{5}$$

where the critical value of F is $F_c = -\phi_0 - J\mathcal{O}_c$, the critical value of $\mathcal{O}$ is $\mathcal{O}_c = 1 - 2R_0$, ϕ_0 is the mode of the ϕ-distribution, $R_0 = R(\phi_0)$, and $J_c = 1/\big(2p(\phi_0)\big)$. This $\mathcal{G}$ is a universal function, which has been approximated in (2) by a Gaussian.

One difficulty is to justify why F is assumed to vary linearly as a function of the time. The rationale is that in order to produce a marked change, F has to be swept through the value that corresponds to the maximum of change of $\mathcal{O}$, and that its speed does not change much when it passes this precise value. Note that the applicability of this model also assumes that

(1) the shifts of opinion occur in the subcritical region $J < J_c$, but with J close enough to J_c that the scaling law is a good approximation;

(2) the value of $|dF/dt|$ is the same for each shift of opinion. This is necessary because the law (5) holds with F as the independent variable, but when we rescale to the form (2) with t as the variable, h and w are each rescaled by $|F'|$. If we are looking for a scaling law of the form $h \propto 1/w$ or $h \propto 1/w^{2/3}$ and are attempting to fit the exponent from observing several shifts of opinion, we need to assume that $|F'|$ was the same for each of them.

This model provides therefore a test that is able in principle to determine the presence of a social network, answering one of the questions of Unilever.

4   Algorithms

The Unilever data start from the opening of an online supermarket, and so the system is not in a stationary state. Moreover, new products are introduced in batches, and other products are withdrawn from the supermarket, either due to withdrawal of the product by the manufacturer or the supermarket deciding not to offer them any more. Furthermore, new customers are continuously attracted, while others stop using the shop. In order to overcome these difficulties, we chose to restrict our analysis to market shares of brands in a subset of the data. It is necessary to choose a subset which is stable throughout the time period, with only a few new products being introduced and a few being withdrawn. At the same time it was required that the product be in regular and stable demand by customers to avoid any biases due to seasonal fluctuations.

In our case the collective opinion, $\mathcal{O}$, is represented by the market share and $d\mathcal{O}/dt$ is then equivalent to the change in the market share of the different products in the chosen category. We will thus for subsequent months collect data on the market shares for each brand and then calculate the changes in the market shares. The aggregation of market shares in individual months is necessary to obtain a reliable estimate, thus providing a

compromise between precision in the measurement of market shares and a sufficiently large number of data points for the subsequent analysis.

Using the change in the market share as the dependent variable, we then estimate the parameters h and w from the above equation using a simple nonlinear OLS (orthogonal least squares) regression. Using the parameter estimates for each brand we can then investigate the relationship between h and w to establish whether social interactions are of relevance for the evolution of the market share, by estimating the exponent α in the equation $h = cw^{-\alpha}$. In the absence of social interactions we should find $\alpha = 1$, and $\alpha = 2/3$ otherwise.

5   Examples

In order to establish whether a social network affects the choice of brands we investigated the market for soaps as an example. Using the sales data from the (Swiss) online supermarket we were able to obtain the relative market shares of nine brands of soap, which provided the largest number of data points in the sample available to us. Calculating the market share in each month over a 4-year period we then use equation (2) to estimate the parameters h and w from the change of the market share for each of the brands. We removed four brands from the analysis because their sales were not sufficient to provide enough data points and any estimate would be too unreliable. From the remaining five estimations we then obtain that $\alpha = 0.553$, as shown in Figure 1. Given the small sample size and a standard error of 0.163 we can reject the hypothesis that the exponent is 1, while 2/3 cannot be rejected, thus providing evidence for the importance of the social network in the choice of brands.
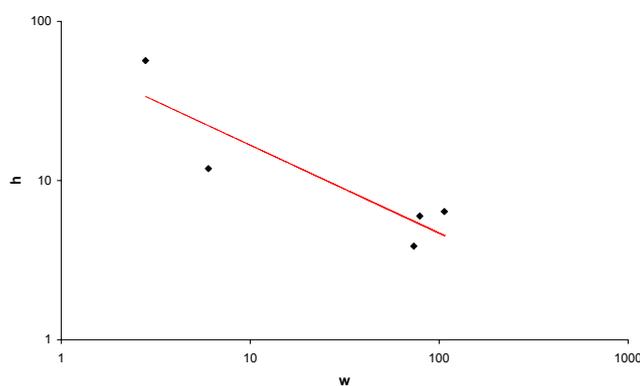

Figure 1: Height h of the first derivative of market share as a function of its width w for several soap brands, and a least-squares fit (on logarithmic axes).
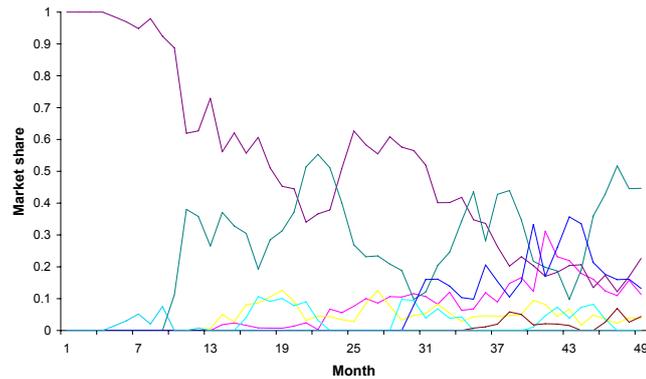
Figure 2: Market share as a function of time for several brands, showing birth-death processes.

These results have to be interpreted with substantial care. Despite having access to all purchases of soaps from the supermarket, a total of about $11,000$ transactions, the need to establish market shares for the nine brands required substantial aggregation of data, thereby reducing the sample size. For a more reliable analysis we would require access to a larger database with more products and a longer time period.

Another complication is that new brands get stocked by the supermarket during the observation time while other brands get discontinued, as can easily be seen in Figure 2. This requires necessarily a shift in demand by the customers which can easily distort our results.

Finally, switching between brands, and thus changing market shares, can also be the result of promotions, as we establish below. Ideally we would need to eliminate this effect because it will distort our results and probably make them less significant.

6   Efficiency of discounting

Using the same dataset we were also able to track the behaviour of individual customers and their choice of soap brand over time. We can identify whether in subsequent purchases a customer has switched brand or stayed with the same brand. With these data we can estimate the probability of a customer switching brands, using a logit model incorporating a number of relevant factors that can affect their choice; these independent variables are other information the dataset provides us with. On analysing the data we found that the main determinant of a switch into a particular brand was the existence of a promotion for that brand. As the dataset contains information on promotions and also on the size of the discount offered, we can directly evaluate the impact of promotions on

the probability of brand-switching. The table below provides the parameter estimates for a logit regression analysing the brand-switching behaviour of consumers, where the variables are a discount indicator I_D (1 if there is a discount, 0 if not) and D, the size of this discount as a fraction of the full price.

The logit model is particularly suitable for this estimation as it is a binary choice model, the two choices in our case being whether to switch the brand or not. The equation being estimated using the maximum likelihood method is

$$\text{Prob(Customer switches brand)} = \frac{\exp(c_0 + c_1 I_D + c_2 D)}{1 + \exp(c_0 + c_1 I_D + c_2 D)}. \tag{6}$$

The results in Table 1 show in column 1 the best fit for c_0 and c_1 with c_2 taken as 0, then in column 2 the best fit for c_0 and c_2 with c_1 taken as 0, and in column 3 the best fit for c_0, c_1 and c_2. Statistical significance at the 90%, 95% and 99% levels is indicated by *, **, ***. We can see from the table of parameter estimates that a promotion which offers the product at half price, a common promotion in our database, increases the probability of a switch between brands from about 25% to 42%, a considerable increase and witness to the effect of promotions on consumer behaviour.

	1	2	3
c_0	-1.0892***	-1.0843***	-1.0892***
c_1	0.6683***		0.5283*
c_2		1.4438***	0.44722

Table 1: Logit estimation of the probability of switching brands.

These results, however, have to be taken with caution. Although the sample size of more than 11,000 transactions is a sufficiently large sample size, the observations can be skewed. It might be that a customer buys different brands for different members of his family, thus we might observe a change of brand from the previous purchase, but this might not represent a genuine switch, and so our estimation of the switching probability is likely to be overstated.

We also cannot know from this analysis whether the switch of brands arising from promotions was permanent or whether consumers switching brands are fickle and constantly switch brands in order to get the lowest prices.

7   Conclusions

We have tried to address the question of validating agent-based models by attempting to establish whether the properties of certain classes of models are consistent with actual data. Our decision was to look for properties that are not dependent on the exact model specification, thereby allowing us to establish the existence of certain properties much more generally. Such a strategy avoids more easily the criticism that the results are dependent on the model (and with a similar model might be reversed).

Using our approach we find evidence for the importance of social networks in the purchasing decisions of consumer goods as well as the influence of promotions on the decision of consumers to switch brands.

The next stage is to find a method to determine the social network of interaction. This may prove impossible with our dataset.

References

[1] P. Casas, S. J. Chapman, R. Hunt, G. Kozyreff, A. Lacey, E. Larrieu, R. Leese, T. Roose, D. Schley, and L. Staron. Models of consumer behaviour. Study Group Report, ESGI 2004. Available at http://www.smithinst.ac.uk/Projects/ESGI49/ESGI49-UnileverConsumers/Report.

[2] N. F. Johnson, D. Lamper, P. Jefferies, M. L. Hart, and S. Howison. Application of multi-agent games to the prediction of financial time series. *Physica A*, 299:222–227, 2001. cond-mat/0105303.

[3] Q. Michard and J.-P. Bouchaud. Theory of collective opinion shifts: from smooth trends to abrupt swings. *European Physical Journal B*, 47:151–159, 2005.

[4] M. E. J. Newman. The structure and function of complex networks. *SIREV*, 45:167–256, 2003.

[5] P. A. D. M. Robert D. Groot. Minority game of price promotions in fast moving consumer goods markets. *Physica A*, (350):553–547, 2004.