J Econ Interact Coord (2013) 8:277–293
DOI 10.1007/s11403-012-0096-3

REGULAR ARTICLE

On the problem of calibrating an agent based model for financial markets

Annalisa Fabretti

Received: 19 October 2011 / Accepted: 4 June 2012 / Published online: 20 June 2012

Abstract Agent based models are very widely used in different disciplines. In financial markets, they can be used to explain well known features called stylised facts and fit statistical properties of data. For this reason, they can model price movements better than standard models using gaussianity. Calibration and validation are essential issues in agent-based modeling. However, calibrating such models is not yet sufficiently considered in the literature. In this paper, a Nelder–Mead simplex algorithm coupled with threshold accepting algorithm (Gilli and Winker in Comput Stat Data Anal 42:299–312, 2003) and a genetic algorithm have been implemented to calibrate the model presented by Farmer and Joshi (J Econ Behav Org 49:149–171, 2002) and the outcomes have been compared and discussed. The data used are closing prices of S&P500 Composite index and a particular attention has been devoted to the choice of the objective function.

Keywords Agent based model · Artificial markets · Calibration · Genetic algorithm

JEL Classification C13 · C52 · G10

1 Introduction

In recent years, an increasing number of researches have focused on Agent Based Models (ABMs). Traditional economic models adopt Gaussian distribution considering extreme events as outliers or, alternatively, building a statistical process that is able to reproduce market features without providing any understanding of how these

A. Fabretti (✉)
Department of Economics and Finance, University of Rome Tor Vergata,
Via Columbia 2, Rome, Italy
e-mail: annalisa.fabretti@uniroma2.it

features are observed; while ABMs drop classical ideas of the representative agent and market efficiency to focus on describing agents' behaviours and interactions.

ABM are widely used in social science and economics (Gallegati et al. 2003; Raberto et al. 2011; Delli Gatti et al. 2011). The agent-based approach considers a population of intelligent adaptive agents or firms and lets them interact in order to maximise their financial performances. For financial markets, they place attention on interactions, learning dynamics, herding behaviour (Lux and Marchesi 1999) or different approaches to trade (Day and Huang 1990). Some use simplified technical trading rules (Farmer and Joshi 2002), and others focus on the price formation mechanism, i.e. the order matching mechanism, to argue what matters in reproducing stylised facts (Pellizzari and Dal Forno 2007). For financial markets, a comprehensive view of the literature of ABM can be found in Hommes (2006) and LeBaron (2000).

ABMs are able to reproduce some of the statistical properties of returns seen in real stock markets (Cont 2001). These features include distribution of returns that is more peaked than the Gaussian distribution, periods of persistent high volatility, periods of persistent high trading volume and correlation between volatility and trading volume. Moreover, ABMs can help to identify the causes which such features can be attributed to.

Since ABMs are useful in reproducing market features, validation and calibration can play an important role. Indeed, as stated by Leombruni and Richiardi (2005), Leombruni et al. (2006) and Marks (2007), a better validation of ABM would reduce scepticism about their results and their usefulness. According to Tesfatsion's web site (http://www2.econ.iastate.edu/tesfatsi/empvalid.htm), there are three alternative ways of validating computational models: a descriptive output validation; a predictive output validation; and an input validation. But how can these be done in a scientifically meaningful manner?

Generally speaking, ABMs can be validated if the statistical properties of simulated data match those of real data and the two series can be said to belong to the same distribution, but it is difficult to justify choosing one specification of model parameters instead of another. Therefore, calibration is useful either for predictive purposes or to provide criteria for the best choice of the parameter set reproducing real data. This can be used as a discriminant in choosing between two models. Indeed, ABMs can have different features and not any of them can fit the stylised facts, qualitatively and quantitatively well (Veryzhenko et al. 2010).

On the one hand, a simulation model can be set up on theoretical considerations or ad-hoc assumptions. On the other hand, the model can be set up on empirical data. The adopted approach must often motivated by the aim of the model, i.e. predictive purposes, experimentation or better understanding of dynamics and causes. So far there are in the literature only a limited number of contributions dealing with a systematic way of calibration. An interesting and wide discussion of the literature can be found in Windrum et al. (2007) and in Fagiolo et al. (2007). Most of these methods are descriptive or heuristic, generally based on similarity of generated sample data (Palin 2002). In fact, a qualitative calibration can be sufficient when validating a model for better understanding the general nature and/or origin of some phenomena. In contrast, a quantitative calibration is needed when a selected model must be applied for predictive purposes.

Some interesting quantitative examples are in Alfarano et al. (2005), in Bianchi et al. (2007) and Mike and Farmer (2008). In Alfarano et al. (2005), an asymmetric version of Kirman's model is proposed; for such a model a closed form solution of the distributions of returns can be derived. Hence, the model parameters can be estimated by applying the maximum likelihood method. In Bianchi et al. (2007, 2008), some parameters of CATS model (see Gallegati et al. 2003) are calibrated via indirect inference (Gourieoux and Monfort 1997) and a minimisation method proposed by Gilli and Winker (2003). Analogous approaches are proposed in Delli Gatti et al. (2011) and Cirillo and Gallegati (2012), too. In Mike and Farmer (2008), each component of a model, used to simulate the price formation under a continuous double auction, is validated against real data; in fact using data from the order book, they can fit the distribution of orders signs , prices and cancellation frequency. However, it is not always possible to have a closed form for the returns distribution as in Alfarano et al. (2005) or a direct simulation of some observables like in Mike and Farmer (2008), and hence a more general method for calibration is needed. In Bianchi et al. (2007), once the auxiliary model is chosen, they deal with the minimisation of the distance between distribution using Gilli and Winker (2003), which is in itself an interesting attempt to provide a general solution to ABM calibration problems. Indeed, Gilli and Winker (2003) proposed a Nelder–Mead simplex algorithm coupled with a local search heuristic called Threshold Accepting. The Kirman's model has been calibrated using an objective function combining estimation errors on kurtosis and ARCH(1)-effect. In Fabretti (2011), such a method has been replicated for a model with fundamentalists and chartists agents (Farmer and Joshi 2002) and some drawbacks have been highlighted. In particular, in Gilli and Winker (2003) and Fabretti (2011), the objective function is naive, the kurtosis and ARCH(1) effect are insufficient to represent statistical properties of data and their weights have been arbitrarily set. In order to capture the main features of data, the objective function must be a blend of higher moments and other statistics. According to Winker et al. (2007), the moments and the statistics must be robust, reflect statistical properties of data and exhibit the potential to discriminate between alternative models or parameter values.

Other heuristic methods, such as Genetic Algorithms (GAs) (Holland 1975; Goldberg 1989), appear to be promising (Calvez and Hutzler 2005; Rogers and Von Tessin 2004). GAs take inspiration from the evolutionary process and select the best set of parameter with respect to a fitness function, which must reflect the goodness of the set parameters. GAs differ from traditional search or optimisation algorithms because they are able to evaluate at the same time a population of points (instead of a single point) and they do not require any analytical tractability of the objective function or additional information (Goldberg 1989).

The present paper offers just one frame of the compelling and extensive issue of parameters estimation for ABMs. The discussion does not claim to be exhaustive. However, a wider view on experiments and comparison of the methodology proposed by Gilli and Winker (2003) and of Genetic Algorithms is provided in the case of a model with a large set of parameters, no closed form solution for distribution and only one observable; at this scope the model proposed by Farmer and Joshi (2002) has been selected. The model is able to reproduce some market dynamics even if more complex and realistic mechanism of price formation could be considered. An indirect inference

approach as in Bianchi et al. (2007) has also been taken in consideration. This approach would allow to be avoided problems related to the choices of the moments and the statistics in the objective function; however, the choice of the auxiliary model is seriously tricky and the minimisation of the distance between auxiliary model parameters and FJ model parameters would re-address to Gilli and Winker (2003) procedure. For such reasons, the direct application of Gilli and Winker (2003) method has been preferred and indirect inference has been left for future research. As we explain later in detail, the two approaches proposed in parameters estimating for FJ model remain between the hypothetical approach (some parameters are set according to some theoretical assumptions) and the empirical approach (parameters are evaluated by using empirical data). Using Brenner and Werker (2007) taxonomy, such approaches can be classified closed to a microsimulations one.

The complexity and the stochasticity of the problem does not allow to be produced an exact solution in any case with certainty. However, a stochastic approximation of a global optimum can be more useful than a deterministic local minimum. In the studied case, it appears reasonable to accept that a unique global minimum does not exist, while a minimum region exists. In fact, both the two methods reach a minimum region, where a slight perturbation of the set of parameters generates a negligible difference in the objective function and in the observed properties of data.

The simulated data show less fat tails and less volatility clustering. However, the failure to reproduce the leptokurtosis of real data is more likely to be attributable to the model. Moreover, the model is quite complex to be properly tested and it could be the case that we have an over parameterisation. For this reason, the test for accuracy and complexity (see Marks 2007), evaluation of the model, could be useful to shed light on this point, which is nevertheless beyond the scope of this paper.

The paper is organised as follows: Sect. 2 presents Farmer Joshi Model; Sect. 3 explains the objective function choice, the methodologies and data; Sect. 4 explains the results; and finally Sect. 5 concludes.

2 The Farmer Joshi model

In the following the agent based model chosen to be calibrated is described. The model (Farmer and Joshi 2002) considers two types of agents characterised by widespread strategies in the market: fundamentalist and chartist. In order to aggregate demands the model also includes a risk neutral market maker.

A single time step t in the trading process can be decomposed into two parts:

(1) the traders observe the most recent prices $P_t, P_{t-1}, \ldots P_{t-d}$ and the information I_t and submit an order ω_t^i;
(2) the market maker fills all the orders at the new price P_{t+1}.

Let N be the number of agents, the market maker bases the price formation only on the net order:

$$\omega_t = \sum_{i=1}^{N} \omega_t^i.$$

The market maker fills the price for the net order using the so-called market impact function:

$$P_{t+1} = P_t \cdot e^{\frac{\omega_t}{\lambda}} \tag{1}$$

where λ is called the liquidity parameter.

Letting $p_t = \log P_t$, and adding a noise term ξ_{t+1}, Eq. (1) becomes

$$p_{t+1} = p_t + \frac{1}{\lambda} \sum_{i=1}^{N} \omega_t^i + \xi_{t+1}. \tag{2}$$

The quantity ω_t^i is a function of $p_t, p_{t-1} \ldots, I_t$ and is

$$\omega_t^i(p_t, p_{t-1} \ldots, I_t) = x_t^i - x_{t-1}^i, \tag{3}$$

where $x_t^i = x_t^i(p_t, p_{t-1} \ldots, I_t)$ is the position at time t of the i-th trader. Each agent i set x_t^i according to his strategy, chartist or fundamentalist.

Trend followers invest based on the belief that price changes have inertia. A trend strategy takes a positive (long) position if prices have recently been going up, and a negative (short) position if they have recently been going down, hence

$$x_{t+1}^i = c(p_t - p_{t-d}), \tag{4}$$

where c is a positive constant and d is the time lag.

Value investors make a subjective assessment of the value in relation to price. They believe that their perceived value may not be fully reflected in the current price, and that future prices will move towards their perceived value. They attempt to make profits by taking positive (long) positions when they think the market is undervalued and negative (short) positions when they think the market is overvalued.

Let be v_t the logarithm of the fundamental value, it follows a random walk:

$$v_{t+1} = v_t + \eta_{t+1}, \tag{5}$$

where η_t is a noise process IID with mean μ_η and standard deviation σ_η. The value investor takes the position

$$x_{t+1}^i = c(v_t - p_t), \tag{6}$$

where $c > 0$ is a constant proportional to the trading capital.

From the point of view of a practitioner, a concern with the simple position-based value strategies and the simple trend follower strategies is excessive transaction costs. Trades are made whenever the mispricing changes. To ameliorate this problem and reduce trading frequency strategies, a threshold can be used for entering a position and another threshold for exiting. Such strategies are based on the belief that the price will revert to the value. By only entering a position when the mispricing is large,

and only exiting when it is small, the goal is to trade only when the expected price movement is large enough to beat transaction costs. Assume that a short position $-w$ is entered when the mispricing exceeds a threshold T and exited when it goes below a threshold τ. Similarly, a long position w is entered when the mispricing drops below a threshold $-T$ and exited when it exceeds $-\tau$. Thus, this strategy depends on its own position as well as the mispricing. In general, different traders will choose different entry and exit thresholds. Let trader i have entry threshold T^i and exit threshold τ^i extracted from a uniform distribution of entry thresholds ranging from T_{min} to T_{max} and uniform density of exit thresholds ranging from τ_{min} to τ_{max}. Finally, parameter c, in (4) and (6), is chosen so that $c = a(T - \tau)$, where a is a positive constant called scale parameter for capital assignment.

3 Data and methodology

In the following the choice of the objective function and the optimisation methods are discussed in detail.

3.1 Objective function

The objective function is a combination of estimation errors on some chosen quantities (either moments or statistics). Therefore, the problem to be solved is a minimization problem

$$\min_{\theta \in \Theta} f(\theta), \tag{7}$$

where θ is the vector parameters, Θ is the space of feasible parameters.

Two important issues arise: the choice of moments and statistics, the determination of weights. According to Winker et al. (2007), the moments and the statistics must assure the potential to reflect statistical properties of data. For example, in financial data fat tails, volatility clustering and persistence are well known stylised facts. The statistics are also required to be robust, however few of these measures are. Thus, the estimation can be performed using moving block bootstrap and the obtained variance can be used to weight each statistic in such a way that their contribution is equally balanced in the determination of the function value.

Formally, denoting by $\mathbf{m}^e = (m_1^e, \ldots, m_k^e)'$ and $(\mathbf{m}^s|\theta) = (m_1^s, \ldots, m_k^s)'$ the vector of empirical moments and statistics of real and simulated data , respectively; the estimation error is defined as

$$\hat{G}(\theta) = \frac{1}{R} \sum_{i=1}^{R} [(\mathbf{m^s}|\theta) - \mathbf{m}^e].$$

Denoting by $\mathbf{W}$ the $k \times k$ matrix of weights of moments and statistics, the objective function is hence defined as a quadratic form

$$f(\theta) = \hat{G}(\theta)' \mathbf{W} \hat{G}(\theta). \tag{8}$$

Following Winker et al. (2007), the matrix **W** is given by the inverse of the covariance matrix obtained from the bootstrap distribution of moments using a window length of b.

The statistics and moments involved in the objective function are the following:

- *mean* and *standard deviation*
- *kurtosis*
- *Kolmogorov-Smirnov statistics*
- *Hurst exponent*

The mean, the standard deviation, the kurtosis and the K-S statistics are chosen to represent the overall shape of the data distribution. The Hurst exponent has been chosen to represent the scaling properties of data. It has been largely investigated in the literature because of its appealing feature to synthesise the scaling characteristic as a unique index. Researches on the Hurst exponent have inspired many models for financial phenomena and many authors have proposed different methods for estimating it (Peng et al. 1994). Hurst's (1965) original method is called Rescaled Range Analysis. However, this method is sensitive to outliers and the deriving statistic is not robust. For this reason, the present study has adopted the generalised Hurst Exponent $H(q)$ with $q = 1$ (see Di Matteo et al. 2003; Di Matteo 2007).

3.2 Calibration methods

Standard methods are subject to fail because such objective function does not always behave well to guarantee a global optimum solution. Indeed, an objective function built taking into account the statistical properties of data is not smooth. In different fields of science and engineering, heuristics optimisation have been applied successfully. Such heuristics include simulated annealing, threshold accepting heuristic, neural networks and genetic algorithms (Fonseca and Fleming 1993; Calvez and Hutzler 2005). These kinds of methods require a strong computational power and thus are really time-consuming. Due to the complexity of the problems and the stochastic elements of the algorithm and the model, they cannot produce an exact solution in every case with certainty. However, a stochastic global optimum is probably more useful than a deterministic local minimum provided by a classical method or even no solution at all.

In the following two methods are described: the methodology proposed in Gilli and Winker (2003), which consists of a classical algorithm, such as the Nelder–Mead simplex coupled with an heuristic called threshold accepting, and a genetic algorithm.

It should be noted that for both methodologies, any evaluation of the objective function is performed on truncated means of the moments for R simulations of data. The evaluation of the objective function requires a computational cost proportional to the number of days simulated NT and the number of replications R. A high value of R is preferable because it reduces the objective function variance.

3.2.1 Nelder–Mead with threshold accepting method

The method proposed by Gilli and Winker (2003) is a combination of the Nelder–Mead (NM) simplex algorithm and a heuristic called threshold accepting (TA). While

the NM provides an efficient way to identify the search direction, the TA helps to avoid local minima. Gilli and Winker (2003) proposed different ways of coupling the two mentioned methods. In the present study all the versions have been tested, but description and results are reported only for the one reckoned to be most successful. At each step, an NM search or a shift of the simplex by TA occurs with equal probability.

Given $(n+1)$ $\theta(i)$'s vertices of the simplex, where n is the number of parameters to be set, the θ's are sorted increasingly with respect to the correspondent objective function values as the $\theta(1)$ is the set of parameters which has generated the lowest value of f, while $\theta(n+1)$ did the worst. The vertex $\theta(n+1)$ is replaced by one among the expansion, the reflection, the in or the out contraction points; if none of them improves the value of f, the simplex shrinks (Nelder and Mead 1965).

Given a $\bar{\theta}$ the threshold algorithm explores n_R rounds (with n_S steps for each round) of the neighbourhood of $\bar{\theta}$; in any round $j = 1, \ldots, n_R$ the procedure perturbs $\bar{\theta}$ randomly n_S times. If the resulting value of the objective function f_{new} is lower than $f(\bar{\theta}) + \rho_j$ the point $\bar{\theta}$ is substituted by the new point and the procedure carries on. The threshold ρ_j is decreased successively for each round and reaches the value of zero in the last round (Winker 2001).

As in Gilli and Winker (2003) the procedure implemented here first considers low values of R and large value of ρ, in order to find the descending direction. Later on the value of R is increased to reduce the objective function variance and ρ is reduced to have more precise shifting.

For the experiments $R = 5, 10, 15, 20, 25, 30, 50$, $n_R = 10$, $n_S = 10$, ρ goes from $\frac{1}{R}$ to $\frac{1}{R} \times 10^{-1}$.

The procedure takes long computational times. In fact, while NM replicates the objective function for each point in the simplex, TA must replicate the computation $n_R \cdot n_S$ times.

3.2.2 Genetic algorithm

Genetic algorithms (GAs) is inspired by evolutionary theory. According to evolution, the individuals which best fit the environment survive and breed. GAs are adaptive heuristic search algorithms able to evaluate in a parallel fashion many hyperplanes of the search space and to direct the search into the region of better performance. GAs were first introduced by Holland (1975); then, they received a large interest as optimisation tools and this led to many variations, see Goldberg (1989).

The basic implementation of a GA can be decomposed into two stages. It starts with a current population, which consists of a number of competing candidate parameters set solutions; then selection is applied to create an intermediate population which, by means of recombination and mutation, generates the next population. GAs can differ in the selection, recombination and mutation processes (see Whitley 1994). The scope of selection is to prefer better individuals for passing genes to the next population. For each individual the fitness function is computed; in nature, the fitness function represents the individual's adaptability to the environment, while here it is the ability to reproduce the observed statistical properties of data, i.e. the objective function in (8). The individuals can be ranked with respect to their fitness function. Indeed, the lower the function the higher the capacity of the set of parameters to reproduce the

data. Here, the two best individuals are voted in being parents. [For other mechanism of selection see Sivaraj and Ravichandran (2011)]. Once we have selected the parents a new intermediate population is generated by genes crossover or recombination. Here, we apply a multiple crossover where each parameter can be inherited by one of the two parents with equal probability. After recombination, the mutation operator is applied with probability p_{mut}; a gene subject to mutation undergoes a slight random percentage change.

In general, in the recombination process, the population can shrink or expand; in the GA implemented here the population dimension is fixed to NI individuals to control the execution time. In fact the algorithm complexity is in the order of the population dimension NI. Unfortunately, this choice can drive the algorithm to converge prematurely without a satisfactory solution; that is, all the individuals have the same genes. Within the scope of preventing the loss of diversity, the mutation becomes necessary; in particular, to compensate the small and fixed dimension of the population, a high value of mutation probability p_{mut} is needed. Moreover, an n-point crossover helps to overcome the limited information capacity of small populations and the tendency to homogeneity (see Eshelman 1991).

3.3 Data

In this paper the data considered are closing prices of S&P500 Composite index from 14/10/2005 to 12/10/2007. Simulated data are produced by using Eq. (2), hence moments and statistics are estimated for the log prices.

In order to study the data's characteristics a basic statistical analysis has been done on the logarithmic returns (see Fig. 1). Returns have fat tails (on the top right panel), and the autocovariance of returns (on the bottom left) is almost negligible, while the autocorrelation of the absolute returns is significant (on the bottom right). As it is well known, financial data show heavy tails, absence of autocorrelation in logreturns and volatility clustering.

The statistics have been estimated for the logarithmic prices using the moving block bootstrap with block size $b = 100$ and resampling 10,000 times (see Table 1). The weights matrix W has been estimated to be:

$$W = \begin{pmatrix} 3{,}308 \times 10^6 & -1{,}807 \times 10^6 & 2{,}766 \times 10^3 & -4{,}346 \times 10^3 & 8{,}320 \times 10^4 \\ -1{,}807 \times 10^6 & 1{,}006 \times 10^6 & -1{,}260 \times 10^3 & -1{,}623 \times 10^3 & -4{,}393 \times 10^4 \\ 2{,}766 \times 10^3 & -1{,}260 \times 10^3 & 18{,}924 & -99{,}851 & 11{,}187 \\ -4{,}346 \times 10^3 & -1{,}623 \times 10^3 & -99{,}851 & 8{,}163 \times 10^3 & 3{,}011 \times 10^3 \\ 8{,}320 \times 10^4 & -4{,}393 \times 10^4 & 11{,}187 & 3{,}011 \times 10^3 & 4{,}324 \times 10^3 \end{pmatrix}.$$

In Fig. 2 two examples of the objective function are plotted against the couple (μ_η, σ_η) (on the left) and the couple (λ, σ_ξ) (on the right) varying in the range $(0.001, 0.01) \times (0.001, 0.02)$ and $(0.6, 1.3) \times (0.001, 0.02)$, respectively. The number of replications R is set to 30. The objective function appears quite rough and shows a large area with low values where a unique minimum cannot be distinguished.

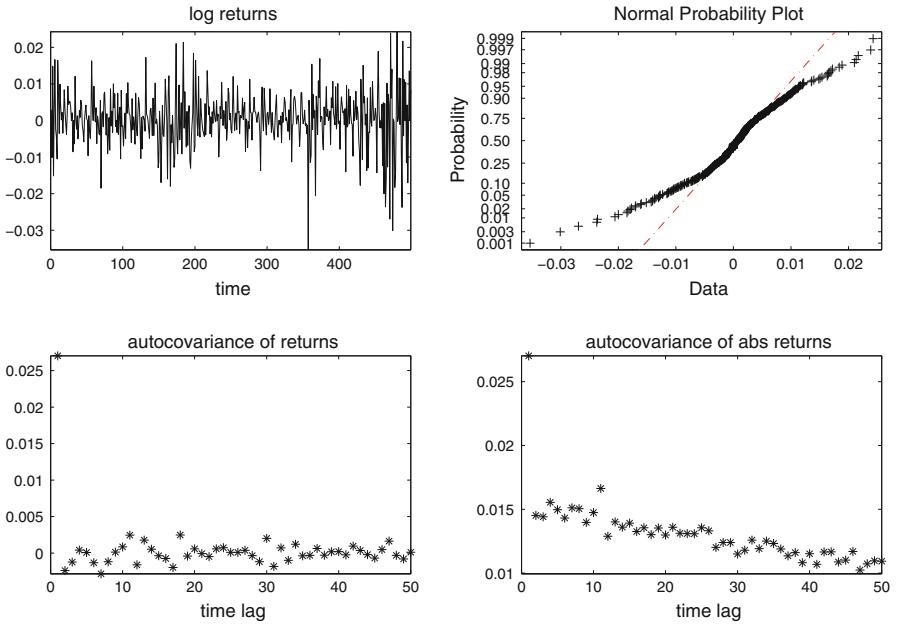

Fig. 1 Data S&P 500 Composite index from 14/10/2005 to 12/10/2007. Logarithmic returns are in the top *left-hand panel*; in the *top right-hand panel*, empirical distribution compared with the normal one; in the *bottom left-hand panel*, autocovariance of returns; in the *bottom right-hand panel* autocovariance of absolute returns. Data show volatility clustering and a distribution with fatter tails than the normal one

Table 1 On the right column moments and statistics estimation of S&P500 composite index from 14/10/2005 to 12/10/2007 are reported

	θ^*_{NM+TA}	θ^*_{GA}	S&P
Mean	[6.9553, 7.3508]	[6.9742, 7.3132]	7.2128
Standard deviation	[0.0203, 0.0960]	[0.0203, 0.0901]	0.0709
Kurtosis	[1.4761, 3.963]	[1.4970, 3.8355]	1.7737
H(1)	[0.3493, 0.5150]	[0.4129, 0.5583]	0.4129
K–S	[0.1936, 0.9980]	[0.2034, 0.9920]	0.5023
Rate of acceptance K–S test (%)	26.1	41.2	
Rate of acceptance C–VM test (%)	38.8	30.1	

In the central columns there are the correspondent confidence intervals computed via Monte Carlo methods using the model with parameters θ^*_{NM+TA} and θ^*_{GA}. At the bottom there are the rates of acceptance of two goodness of fit statistical tests: the Kolmogorov Smirnov (K–S) and the Cramer Von Mises (C–VM) tests. The null hypothesis states that the two samples (the simulated data and the actual data) derive from a common distribution

4 Results and discussion

In the following the details of the implementation along with its results are reported and discussed. As in Farmer and Joshi (2002), $T_{\min}$ and τ_{max} are both set to 0 and

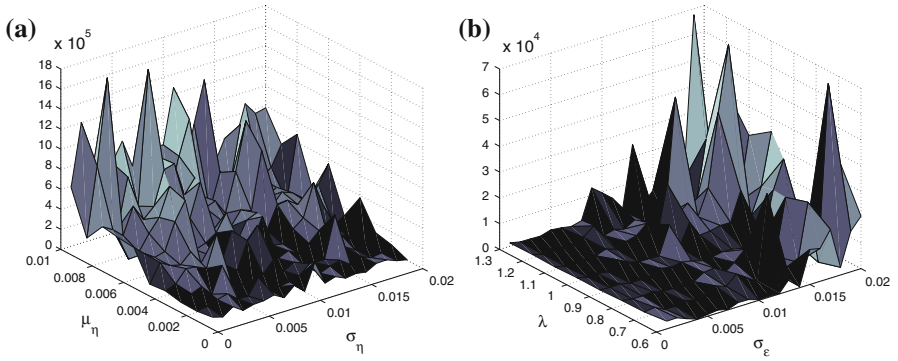

Fig. 2 The objective function is plotted for μ_η and σ_η (*left hand side*), λ and σ_ξ (*right hand side*). Others value are $N_{traders} = 200$, $\lambda = 1$, $a = 0.0001$, $d_{max} = 35, d_{min} = 5$, $\mu_\eta = 0.002$, $\sigma_\eta = 0.002$, $T_{max} = 4$, $T_{min} = 0$, $\tau_{max} = 0$, $\tau_{min} = -0.2$, $v_{max} = 0.5$, $v_{min} = -0.5$, $\sigma_\xi = 0.040$

$\tau_{min} < 0$. In addition, the common parameters for trend followers and value investors (thresholds and capital assignment parameter) are set to be the same to match the empirical fact that the correlation of log-returns is close to zero. In fact the trend followers induce positive short-term autocorrelation and value investors induce negative. Therefore, the set of model parameters to be estimated is

$$\theta = \begin{cases} N & number\ of\ traders \\ \lambda & liquidity\ parameter \\ a & Scale\ parameter\ for\ capital\ assignment \\ d_{max}, d_{min} & Max\ and\ Min\ Time\ delay\ for\ trend\ followers \\ \mu_\eta, \sigma_\eta & mean\ and\ std\ of\ noise\ process\ in\ v_t \\ \sigma_\xi & noise\ variance \\ T_{max} & Max\ threshold\ for\ entering\ positions \\ \tau_{min} & Min\ threshold\ for\ exiting\ positions \\ v_{max}, v_{min} & Max\ and\ min\ offset\ for\ log\ of\ perceived\ value \end{cases} \tag{9}$$

The N traders are split equally into fundamentalists and chartists, (assuming N even), such that $\frac{N}{2}$ traders are chartists and $\frac{N}{2}$ are fundamentalists. The NM + TA algorithm gives problems for integers, hence the number of traders N and the time lag distribution range (d_{min}, d_{max}) have not been included in NM+TA optimisation. By contrast, the GA does not involve such a drawback and the calibration is performed over all the parameters in θ.

Before starting the experiment, it was expected that the procedures were able to identify the descending direction and stop at the minimum. Hence, the experiments were first run with stopping conditions consisting of a small error and a large maximum number of iterations (necessary to stop the procedure if the minimum error is not reached). However, this strategy was suboptimal in finding the minimum and instead was replaced with a sequence of many runs, each of them with a small number of maximum iterations and a lower error than the previous run. Each run was initialised with the outcome of the previous one, while the first one was initialised randomly. The procedures were replicated many times until the solution found was unable to improve

Table 2 The optimal parameters sets θ^*_{NM+TA} and θ^*_{GA} have been obtained using the Nelder–Mead method coupled with threshold accepting and the genetic algorithm, respectively

θ	θ^*_{NM+TA}	θ^*_{GA}
N	208	210
λ	0.4308	0.5145
a	0.0025	1.15e−04
d_{max}, d_{min}	[34, 5]	[42, 41]
μ_η, σ_η	[0.0010, 8.605e−06]	[4.77e−05, 2.35e−04]
σ_ξ	0.0058	0.0051
T_{max}	0.0114	0.0057
τ_{min}	−0.0051	−0.0190
v_{max}, v_{min}	[0.0419, −0.1998]	[0.0424, −0.0120]
min	141.8772	55.6969

Such minimal points are not unique, rather they represent a region of minima where slight changes in the variable do not correspond to tangible differences in the properties of the simulated data. The Nelder–Mead method does not allow to work with integers, thus in θ^*_{NM+TA} the quantity N and d_{max}, d_{min} have been chosen manually after many trials. In comparison, Genetic algorithm allow to optimise over the whole parameters set, therefore the minimum reached by GA is strongly lower than the one obtained by NM + TA

the one reached in the previous run. This strategy allowed the discovery of a solution that is not unique but exists in a region. In fact, different runs reached different points differing in values, but with negligible difference in the objective function and indistinguishable differences in the simulated data. For this reason, the solutions in Table 2 can be considered a proxy for this region, with more detailed exploration left to future research. It is not surprising that the minimum reached by the GA is significantly lower than the minimum reached by NM + TA. Indeed, GA can tune more parameters and hence looks data-mining for the most suitable combination between them.

To test the goodness of fit, the statistical properties of simulated data were investigated and compared with those of real data. In addition, two statistical tests were performed: the Kolmogorov Smirnov and the Cramer Von Mises tests (see Darling 1957). In Table 1 the confidence interval at 95 % level for the moments are reported; the moments were estimated by simulating data using the optimal parameters θ^*_{NM+TA} and θ^*_{GA}; in Table 1, the rates of acceptance at 95 % level of confidence of the Kolmogorov-Smirnov and the Cramer-Von mises tests are reported, too. The null hypothesis consists of stating that the two samples (a simulated one and the actual data) are drawn from a common distribution. The tests were performed on the logarithmic returns. Although the rates of acceptance are lower than expected, it is worth observing that the same tests applied on simulated samples obtained by a parameters set chosen randomly gave 0 % of acceptances.

In Fig. 3 examples of simulated data obtained with optimal parameters are shown. Samples appear quite similar to real data. However, simulated data shows less volatility clustering and less fat tails than real data (see Figs. 4, 5). In fact, as shown clearly in the top right-hand panel of Figs. 4 and 5, when simulated data quantiles are compared with those of a normal distribution not many differences emerge, while differences are evident for actual data (see Fig. 1 at the top right hand side).

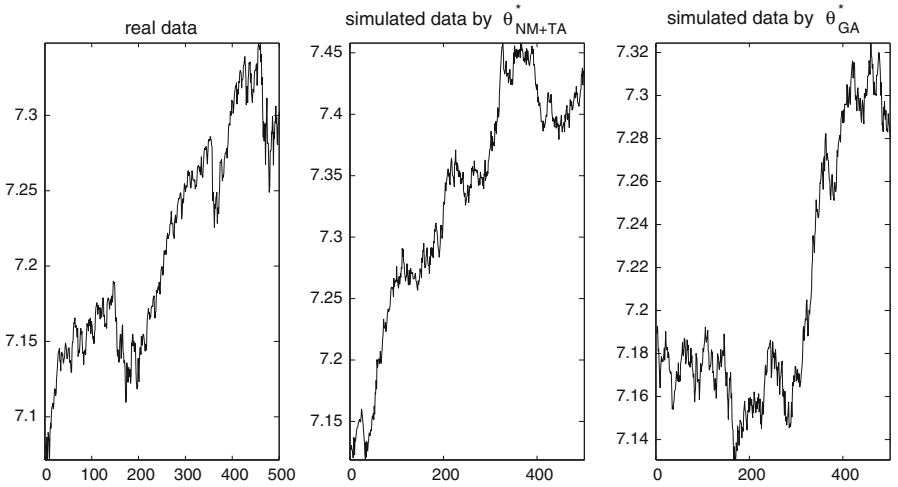

Fig. 3 The estimation has been done on the logarithmic prices of S&P500 from 14/10/2005 to 12/10/2007 (on the *left*). Data generated by θ^*_{NM+TA} (in the *middle*) appear more similar to those generated by θ^*_{GA} (on the *right*). However, in this figure only one determination is reported as an example

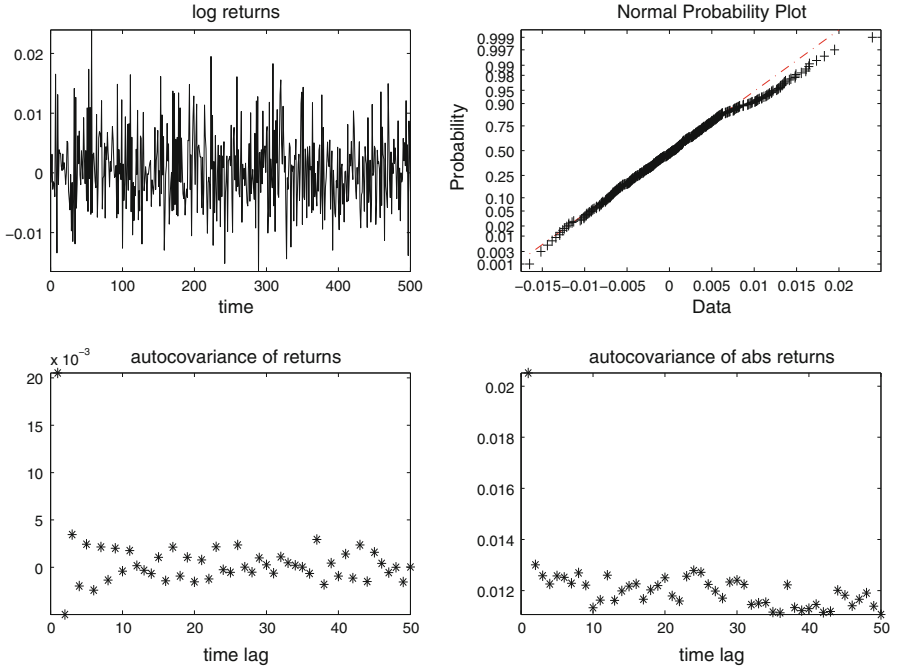

Fig. 4 In *top left-hand panel*, logarithms returns of simulated data by θ^*_{NM+TA}; data shows less fat tails than real data (see in the *top right-hand panel* the qqplot versus a normal distribution) and volatility clustering less marked than real data (in the *bottom left and right-hand panel*)

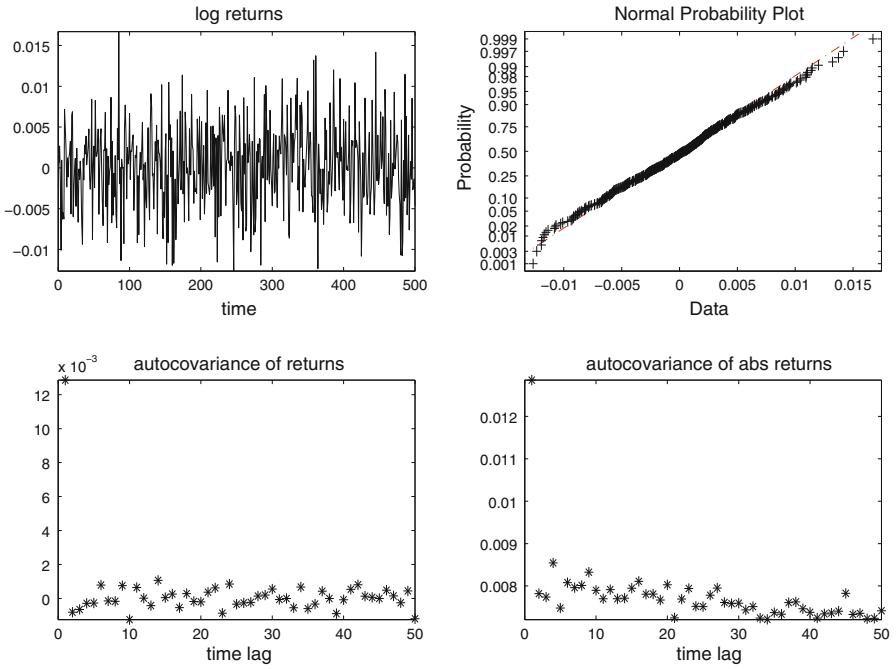

Fig. 5 On *top left-hand panel*, logarithms returns of simulated data by θ^*_{GA}; data shows less fat tails than real data (see on the *top right-hand panel* the qqplot versus a normal distribution) and volatility clustering less marked than real data (on the *bottom left and right-hand panel*)

Finally, in Fig. 6 the boundaries derived as 95 % confidence intervals are plotted for simulated data generated by θ^*_{GA} and by θ^*_{GW}; the actual data stay always inside such boundaries. Differences between boundaries generated by θ^*_{GA} and by θ^*_{GW} are almost intangible.

In conclusion we can observe that the fitting is not completely satisfactory and simulated data cannot replicate faithfully actual data properties. However, the less leptokurtotic nature of simulated data should not be attributed to the optimisation procedures but to the model characteristic or to the fact that chartists and fundamentalists are equally balanced in number and with equal threshold parameters. In fact, fat tails could be generated by chartists activity and here, under the hypothesis of equally distribution, the chartists effect could not emerge.

5 Conclusion and further discussion

Since the estimation problem is a complex and vast topic, the present paper does not claim to be exhaustive. Rather, the aim of this paper is to report the results of experiments on a model comprising of many parameters (Farmer and Joshi 2002), while applying heuristic estimation methods, specifically Gilli and Winker (2003) and Genetic Algorithms. The main results highlight the potential of both the methods to be candidates to fill the gap of parameters estimation in ABMs field. However, many aspects remain unexplored. The main drawback to applying the described

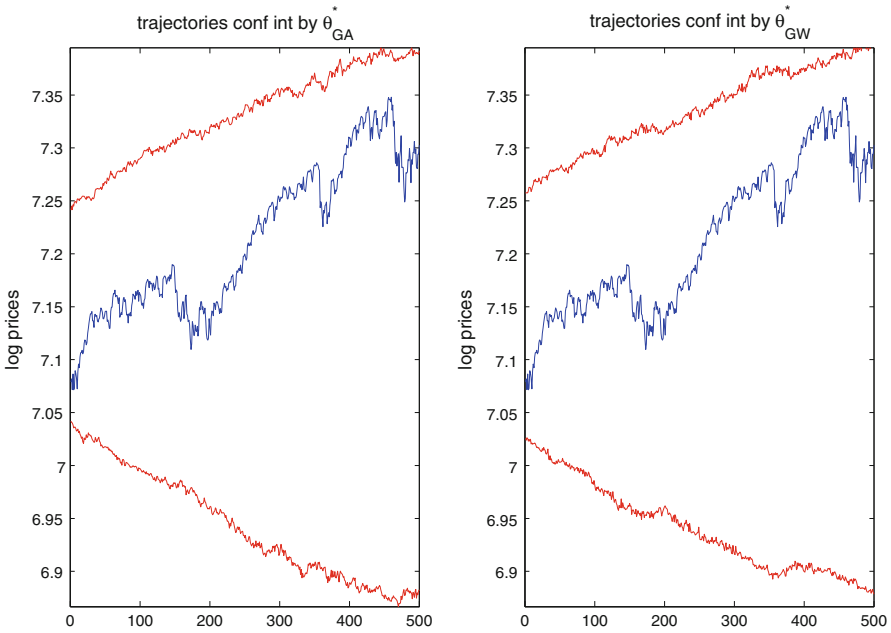

Fig. 6 Actual data are plotted with an upper and a lower boundary derived at a 95 % confidence level by simulated data with parameters set θ^*_{GA} (*left hand side*) and θ^*_{GW} (*right hand side*). The boundaries always contain actual data

procedures consists of long execution times required for computation. Any procedure is so time-consuming that each experiment must be designed carefully before starting. In addition, two questions remain unanswered and require wider explorations: the best choice of the objective function and the time span. Some attempts have been made to test whether an objective function with or without the Hurst exponent, the GARCH parameters and other indexes may perform better, but the results of these experiments cannot be compared in a straightforward way. Some calibration tests have been done on longer (6 years) and shorter (1 year) time periods. Results change significantly either with time span or objective function. For this reason, such choices are very crucial and in absence of some systematic criteria must be principally motivated by the purposes of the calibration (prediction, decision making, validation). Finally, it can be concluded that both methodologies represent good tools to tackle the problem of parameters estimation, but both of them require caution in application. NM with TA sometimes can fail to find the descending direction and the procedure must be restarted, while this never happens with GA. Genetic Algorithms can be preferable when some parameters are integers. Moreover, they are more robust when applied in noisy environment.

References

Alfarano S, Wagner F, Lux T (2005) Estimation of agent-based models: the case of an asymmetric herding model. Comput Econ 26:19–49

Bianchi C, Cirillo P, Gallegati M, Vagliasindi PA (2007) Validating and calibrating agent-based models: a case study. Comput Econ 30:245–264
Bianchi C, Cirillo P, Gallegati M, Vagliasindi PA (2008) Validation in agent-based models: an investigation on the CATS model. J Econ Behav Org 67:947–964
Brenner T, Werker C (2007) A taxonomy of inference in simulation models. Comput Econ 30:227–244
Calvez B, Hutzler G (2005) Automatic tuning of agent-based models using genetic algorithms. Proc Int Work Multi-Agent Based Sim (MABS'05) 3891:39–50. doi:10.1007/11734680
Cirillo P, Gallegati M (2012) The empirical validation of an agent-based model. East Econ J. doi:10.1057/eej.2011.34
Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quant Fin 1:223–236
Darling DA (1957) The Kolmogorov-Smirnvo, CRamer-von Mises tests. Ann Math Stat 28(4):823–838
Day RH, Huang W (1990) Bulls, bears and market sheep. J Econ Behav Org 14:299–329
Delli Gatti D, Desiderio S, Gaffeo E, Cirillo P, Gallegati M (2011) Macroeconomics from the bottom-up. Springer, Berlin
Di Matteo T (2007) Multi-scaling in finance. Quant Fin 7:21–36
Di Matteo T, Aste T, Dacorogna MM (2003) Scaling behaviors in different developed markets. Physica A 324:183–188
Eshelman LJ (1991) The CHC adaptive search algorithm: How to have safe search when engaging in non-traditional genetic recombination. In: Rawlins GJE (ed) Foundations of genetic algorithms Morgan Kauffam, San Mateo, CA
Fabretti A (2011) Calibration of an agent based model for financial markets. Highlights in practical applications of agents and multi-agents systems. Adv Int Soft Comput 89:1–8
Fagiolo G, Moneta A, Windrum P (2007) A critical guide to empirical validation of agent-based models in economics: methodologies, procedures and open problems. Comput Econ 30:195–226
Farmer JD, Joshi S (2002) The price dynamics of common trading strategies. J Econ Behav Org 49:149–171
Fonseca CM, Fleming PJ (1993) Genetic algorithms for multi-objective optimization: formulation, discussion, generalization. In: Proceedings of fifth international conference on genetic algorithm, pp 416–423
Gallegati M, Giulion G, Palestrini A, Delli Gatti D (2003) Financial fragility, patterns of firms' entry and exit and aggregate dynamics. J Econ Behav Org 51:79–97
Gilli M, Winker P (2003) A global optimization heuristic for estimating agent based models. Comp Stat Data Anal 42:299–312
Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison Wesley, Reading MA
Gourieoux C, Monfort A (1997) Simulation based econometric methods. University Press, Oxford
Holland JH (1975) Adaptation in natural and artificial systems. MIT Press, Cambridge
Hommes CH (2006) Heterogeneous agent models in economics and finance. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics, vol 2. Elsevier Science B.V., Amsterdam, pp 10–13
Hurst HE, Black R, Sinaika YM (1965) Longterm storage in reservoirs: an experimental study. Constable, London
LeBaron B (2000) Agent based computational finance: suggested readings and early research. J Econ Dyn Control 24:679–702
Leombruni R, Richiardi M (2005) Why are economists sceptical about agent-based simulations. Physica A 335:103–109
Leombruni R, Richiardi M, Saam NJ, Sonnessa M (2006) A common protocol for agent-based social simulation. J Art Soc Soc Sim 9(1). http://jasss.soc.surrey.ac.uk/9/1/15.html
Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397:498–500
Marks RE (2007) Validating simulation models: a general framework and four applied examples. Comput Econ 30(3):265–290
Mike S, Farmer JD (2008) An empirical behavioral model of liquidity and volatility. J Econ Dyn Control 32:200–234
Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
Palin J (2002) Agent-based stockmarket models: calibration issues and application. PhD thesis, University of Sussex, UK

Pellizzari P, Dal Forno A (2007) A comparison of different trading protocols in an agent based market. J Econ Interact Coord 2:27–43

Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685–1689

Raberto M, Teglio A, Cincotti S (2011) Debt delevaring and business Cycles. An agent-based perspective. Economics Discussion Papers 2011-31, Kiel Institute for the world economy

Rogers A, Von Tessin P (2004) Multi-objective calibration for and agent-based model. In agent-based simulation May 2004, Lisbon, Portugal

Sivaraj R, Ravichandran T (2011) A review of selection methods in genetic algorithm. Int J Eng Sci Tech 3(5):3792–3797

Tesfatsion L, Website on ACE: verification and empirical validation of agent-based computational models. http://www2.econ.iastate.edu/tesfatsi/empvalid.htm

Veryzhenko I, Brandouy O, Mathieu P (2010) Agent's minimal intelligence calibration for realistic market dynamics. In: Li Calzi M, Milone L, Pellizzari P (eds) Progress in artificial economics. Springer, Heidelberg, pp 3–14

Whitley D (1994) A genetic algorithm tutorial. Stat Comput 4(2):65–85

Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: Alternatives and prospects. J Art Soc Soc Sim 10(2). http://jasss.soc.surrey.ac.uk/10/2/8.html

Winker P (2001) Optimization heuristics in econometrics: applications of threshold accepting. Wiley, Chichester

Winker P, Gilli M, Jeleskovic V (2007) An objective function for simulation based inference on exachange rate data. J Econ Interact Coord 2:125–145