Journal of Economic Dynamics & Control 91 (2018) 391–408

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Estimation of agent-based models using sequential Monte Carlo methods

Thomas Lux

Department of Economics, University of Kiel, Germany

ARTICLE INFO

Article history:
Received 13 April 2017
Revised 9 November 2017
Accepted 13 January 2018
Available online 31 January 2018

JEL classification:
G12
C15
C58

Keywords:
Agent-based models
Estimation
Markov chain Monte Carlo
Particle filter

ABSTRACT

Estimation of agent-based models is currently an intense area of research. Recent contributions have to a large extent resorted to simulation-based methods mostly using some form of simulated method of moments estimation (SMM). There is, however, an entire branch of statistical methods that should appear promising, but has to our knowledge never been applied so far to estimate agent-based models in economics and finance: Markov chain Monte Carlo methods designed for state space models or models with latent variables. This latter class of models seems particularly relevant as agent-based models typically consist of some latent and some observable variables since not all the characteristics of agents would mostly be observable. Indeed, one might often not only be interested in estimating the parameters of a model, but also to infer the time development of some latent variable. However, agent-based models when interpreted as latent variable models would be typically characterized by non-linear dynamics and non-Gaussian fluctuations and, thus, would require a computational approach to statistical inference. Here we resort to Sequential Monte Carlo (SMC) estimation based on a particle filter. This approach is used here to numerically approximate the conditional densities that enter into the likelihood function of the problem. With this approximation we simultaneously obtain parameter estimates and filtered state probabilities for the unobservable variable(s) that drive(s) the dynamics of the observable time series. In our examples, the observable series will be asset returns (or prices) while the unobservable variables will be some measure of agents' aggregate sentiment. We apply SMC to two selected agent-based models of speculative dynamics with somewhat different flavor. The empirical application to a selection of financial data includes an explicit comparison of the goodness-of-fit of both models.

1. Introduction

Validation of agent-based models is currently an intense area of research. Various disciplines have developed distributed computational models of interacting agents that due to their inherent complexity appear to pose non-trivial demands when it comes to estimation of their parameters. Examples include the estimation of behavioral parameters in transportation networks (Molina et al., 2005), ecological models (Golightly and Wilkinson, 2011), sociological models of network formation (Snijders, 1996) and behavioral models of speculative behavior in financial markets. The latter has become a particularly burgeoning area of research. Recent contributions in this area have been surveyed by Lux and Zwinkels (2017). Various approaches can be found in this emerging literature: A large body of literature is concerned with models with a limited

E-mail address: lux@economics.uni-kiel.de

https://doi.org/10.1016/j.jedc.2018.01.021
0165-1889/

number of different strategies (mostly only two) whose dynamic evolution can be formalized by a system of typically highly nonlinear difference or differential equations in the limit of an infinite population of agents. This literature has been initiated by Day and Huang (1990), Chiarella (1992), and Brock and Hommes (1997, 1998), among others. Models of this class have been estimated by a variety of algorithms: for instance, Boswijk et al. (2007) estimate the regime-switching model of Brock and Hommes using a nonlinear least-squares approach, while Amilon (2008) estimates the same model using the efficient method of moments based on an auxiliary ARCH model.

More recently, Recchioni et al. (2015) and Lamperti et al. (2017) have resorted to calibration methods rather than econometric estimation to fit the same model to data. This alternative approach is motivated by the observation that even the relatively simple agent-based models of this class come along with a large number of parameters. When embedding a simulation model with many parameters into the loop of an iterative algorithm for parameter optimization, computational demands tend to increase easily beyond feasible boundaries. These computational limitations of 'traditional' statistical approaches have led researchers to select 'parameters of interest' for estimation and fix a priori the remaining ones. The calibration methods of Recchioni et al. (2015) and Lamperti et al. (2017) try to cope with the computational burden by using more efficient approaches: Recchioni et al. (2015) select parameters based on a constraint optimization using the squared residuals of the fit of the raw price data as the underlying loss function, while Lamperti et al. (2017) apply a machine learning algorithm based on a surrogate model that facilitates the exploration of new parameter values in successive rounds of the learning algorithm.

Other recent contributions have cut down on the number of parameters and have applied the simulated method of moments as a most generally applicable tool to estimate parameters, cf. Franke and Westerhoff (2011, 2012, 2016). Kukačka and Barunik (2017) have applied a simulated maximum likelihood approach, while Barde (2016) applies an information-theoretic criterion to calibrate and compare three closely related agent-based models that all can be expressed in a structural form as systems of difference equations.

Recent literature has also embarked on estimation of agent-based models not cast in structural form. Grazzini and Richiardi (2015) and Jang (2015) were probably the first who studied the estimation of a model in which a full set of agents is simulated. The underlying models are very simple ones with only few parameters: Grazzini and Richiardi illustrate minimum distance estimation of agent-based models by example of two simple models: an order book model with adaptive adjustment of limit prices, and a model of adaptation of a new product in a finite population of agents. Following Alfarano et al. (2008), Jang embeds Kirman's model of herding among speculators (Kirman, 1993) into a simple asset pricing framework. While this model can, in principle, be approximated by a stochastic difference equation (it is used by Barde, 2016, in this format), Jang (2015) uses microscopic simulations of the evolution of the strategies of the market participants. He shows that the objective function used in a simulated method of moments (SMM) algorithm is generically non-smooth, and often exhibits multiple equilibria as well as relatively flat areas over wide ranges of the parameter space. Exploring this issue further, Chen and Lux (2017) and Ghonghadze and Lux (2016) show that execution of a single SMM estimation from an arbitrary set of initial conditions could lead to virtually arbitrary results. On the base of extensive Monte Carlo simulations, they recommend to start the estimation with a comprehensive grid search followed by the application of a gradient-free optimization method for a sample of best fitting grid points. They also note strong correlations between the estimates of the three parameters of this model, a long preasymptotic range of the estimates before converging to $T^{1/2}$ consistency, and severe size distortions of the goodness-of-fit test based on the over-restrictions of the SMM algorithm. In conclusion, it appears that while SMM is a perfectly general approach to estimate full-fletched agent-based models, it suffers from important limitations: Firstly, finding the minimum of the SMM objective function can be computationally very challenging given the inconvenient properties mentioned above. Second, to obtain a decent signal-to-noise ratio of the estimated parameters, very large data sets (of order 10^5 to 10^6 in the above papers) would be needed. The latter problem reflects the generally lower efficiency of estimators based upon few moments compared to estimation methods that use more information. Chen and Lux (2017) argue that these problems might extend beyond the particular model explored in their paper, and might have as their principal cause the limited range of moments available for univariate financial data (basically measures of their heavy tails and clustering of volatility).

In view of this unsatisfactory scenario, the present paper aims at expanding the tool box available for the estimation of agent-based models into a new direction. While extant research has used a number of diverse estimation methods, there is, still an entire branch of statistical methods that should appear promising, but has to our knowledge never been applied so far to estimate agent-based models in economics and finance. This branch of statistical methods is Sequential Monte Carlo designed for state space models or more general models with latent variables. This class of models seems particularly relevant as agent-based models typically would consist of some latent and some observable variables since not all the characteristics of agents would mostly be observable. Hence, such models can be very naturally classified as hidden or latent variable models. Indeed, one might often not only be interested in estimating the parameters of a model, but one would also be interested to infer the time development of some latent variable from the observable ones. Extracting information on unobservable variables is a classical *filtering* problem for which state-space models are prototypical examples. The classical approach to inference in state-space models is the Kalman filter which is the optimal solution to the tasks of state filtering and parameter estimation for linear systems with Gaussian noises in both the dynamic laws governing the latent and observable variables (Grewal and Andrews, 2008).

However, agent-based models when interpreted as latent variable models would be typically characterized by non-linear dynamics and non-Gaussian fluctuations and, thus, would require a more general approach to statistical inference. Here

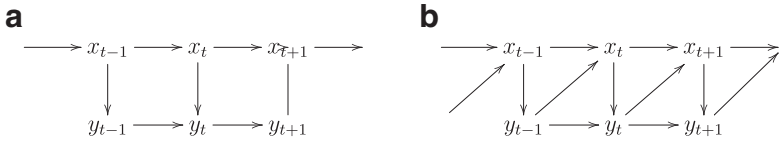

Fig. 1. a. Relationship between the hidden (latent) variable x_t and the measurement y_t in a standard state-space model, b. Relationship between the latent variable and the observable variable in a more general observation-driven model, or dynamic model with latent states.

we resort to an approach that is extremely versatile and has become very popular in a variety of applications: Sequential Monte Carlo (SMC) estimation based on a particle filter. This approach is used here to numerically approximate the conditional densities that enter into the likelihood function of the problem. With this approximation we simultaneously obtain filtered state probabilities for the unobservable variable(s) that drive(s) the dynamics of the observable time series when estimating the parameters of a model. In our examples below, the observable series will be asset returns (or prices) while the unobservable variables will be some measure of agents' aggregate sentiment. In contrast to many alternative approaches to the estimation of such models, we are thus able to identify in each single period the relative contribution of fundamental news and non-fundamental sentiment dynamics in the observed price changes, while in the previous GMM/SMM estimation of such models (e.g. Ghonghadze and Lux 2016; Chen and Lux 2017) the authors could only diagnose the relative importance *on average* over the whole time series of these two factors on the base of estimated parameters.

The main advantages of SMC methods are (i) a more efficient use of available data since SMC is used to numerically approximate the likelihood, and, therefore, uses more information than an SMM approach based on selected moments. In comparison with the results of Chen and Lux (2017), we will indeed find a dramatic increase in the precision of the parameter estimates for the same underlying model, (ii) the identification of unobservable variables which are often in the center of interest of agent-based models, e.g. how many agents adopt what type of strategy. Estimated trajectories of such hidden variables are obtained as a natural by-product of the filtering step involved in any SMC algorithm. Filtered estimates of the hidden variables could be based on different information sets, e.g. based only on past observations or based upon the complete available time series of the observed variable, (iii) sequential Monte Carlo could be embedded into different estimation methodologies, i.e. both frequentist maximum likelihood and Bayesian estimation are possible. Of particular interest given the computational demands of agent-based models is also the possibility of *on-line* estimation, i.e. identification of parameters after only one sweep over the data rather than in-the-loop estimation. All these possibilities will be illustrated on the base of two simple models. We note that by introducing Bayesian methods in the estimation of agent-based models, part of the present research program overlaps with that of Grazzini et al. (2017). However, the latter paper does not use sequential Monte Carlo methods.

In the following, we will proceed by providing a short introduction to latent variable models and their estimation via sequential Monte Carlo methods. Section 2 provides a classification of important classes of models with latent variables, and details on the application of Monte Carlo methods to state filtering and estimation. It also explains in detail the working of sequential Monte Carlo estimation via a particle filter, together with its use in both frequentist and Bayesian estimation and reviews theoretical results available in the statistical literature for this approach. Section 3 will shortly introduce two agent-based models of speculative dynamics with somewhat different flavor that will subsequently serve as our work horses for illustration of the statistical methodology. Section 4 offers details on some Monte Carlo experiments designed to shed light on the performance of sequential Monte Carlo methods for the two competing models, and Section 5 moves on to their empirical application to a selection of financial data including a model contest. Section 6 provides conclusions and avenues for future research.

2. Statistical models and methods of inference

2.1. Types of latent variable methods

The classical version of a latent variable model is the general state-space model. A state-space model is defined by the stochastic evolution in time (or along some other dimension) of the state vector, x_t, and the vector of observable measurements, y_t. In the most general case, this can be written as:

$$\begin{aligned} x_t &= f(x_{t-1}, \epsilon_t) \\ y_t &= y(x_t, \eta_t) \end{aligned} \tag{1}$$

where ϵ_t and η_t are noise factors that could be correlated, and $f(\cdot)$ and $g(\cdot)$ are stochastic processes. One could alternatively describe the state-space-model by the conditional densities of x_t and y_t but such conditional densities need not to be known. The minimum requirement of a general state-space model is that the recursions $f(\cdot)$ and $g(\cdot)$ are uniquely determined by some algorithm, i.e. they can be simulated.

The dynamics of the state-space model, thus, corresponds to the sequence of events illustrated on the left-hand side of Fig. 1. The measurements y_t are, then, just noisy signals for the underlying states and have no effect on the state recursion

itself. Hence, the baseline state-space model can be applied whenever one considers the empirical data as noisy measurements of some 'true' latent (i.e. unobservable) variable of interest. This is, for example, how macroeconomic models of the DSGE (dynamic stochastic general equilibrium models) type are framed as state-space models: macroeconomic variables are assumed to be measured with error only and, thus, take the role of the observables y_t (cf. Fernández-Villaverde and Rubio-Ramírez (2007)).

It turns out that agent-based models will sometimes obey this format; an example will be provided in the next section. Fitting a model in this format is extremely convenient for the plethora of statistical methods that have been developed for this setting over the last decades (only some of which we will illustrate below).

However, not all agent-based models fit into this format. A more general class of models will allow for feedback from the observation to the state, cf. the right-hand side illustration of Fig. 1. We will see that broad classes of financial agent-based models rather fall into this category. This more general framework is denoted an observation-driven system if there is no noise component in the state equation or more generally a dynamic system with latent variables if the state and observation equations are both stochastic (Liu and Chen, 1998). Recently, both variants have attracted some attention: The statistical properties of observation-driven models with deterministic state dynamics have been investigated by Douc et al. (2013). A leading example of an economic application of this class of models is the family of GARCH models. Their stochastic counterpart, denoted stochastic volatility models, in contrast, falls into the traditional framework of state-space models (as they usually lack the autoregressive dependency of volatility on past squared returns besides the autoregressive dynamics of the latent past volatility of the GARCH framework).

Stochastic volatility gives, however, rise to dynamics beyond the state-space formalism if it allows for a leverage effect between the first and second moment of returns. Dependency of volatility on past price changes then leads to a state dynamics that is not autonomous anymore so that a more general system of two stochastic differential equations is obtained with one observable (returns) and one latent variable (volatility) and feedback from the first to the second.[1] More recently, general models with latent variables have been studied in a microeconomic context by Blevins (2016) and Gallant et al. (2016). Both papers apply a so-called *particle filter*, the standard approach for sequential Monte Carlo estimation of complex non-linear models, which is also known as bootstrap filter or sampling-importance resampling (SIR) approach and has been introduced by Gordon et al. (1993) and Kitagawa (1996). Both Blevins (2016) and Gallant et al. (2016) elaborate on the theoretical properties of their estimation approach which is not covered by the rich body of results available for standard state-space models. Gallant et al. (2016) show that maximum likelihood estimation based upon a particle filter approximation is unbiased while Blevins (2016) demonstrates its asymptotic consistency.

2.2. Sequential Monte Carlo estimation of models with latent variables

Sequential Monte Carlo methods attempt to approximate unknown, evolving probability distributions via simulations. The distributions one wishes to approximate in dynamic models with latent variables are different forms of posterior distributions. Denoting by $\{y_t\}$, $t = 1, \ldots, T$ the observed variable, by $\{x_t\}$, $t = 1, \ldots, T$ the hidden variable, and by θ the vector of possibly unknown parameters, the entities of interest might be the predictive density $p(x_t|Y_{t-1}, \theta)$, the filtering density $p(x_t|Y_t, \theta)$ or the smoothing density $p(x_t|Y_T, \theta)$ where Y_t is the set of all observations from $y_1, \ldots$ to y_t. In a Bayesian framework, additional objects of interest might be defined by conditioning on the posterior for the parameters. Sequential Monte Carlo (SMC) methods provide approximations of these densities. These consist of discrete, weighted samples that provide a point-wise approximation of the density of interest that is not available in closed form. With these approximations, also functions involving these densities can be approximated. Because of the point-wise nature of the approximation, SMC methods are also known as particle filters (i.e. the relevant density is approximated by a set of discrete points or *particles*). In the following, we illustrate the evolution of such a particle filter for the task of approximating the likelihood function of a state space model or a dynamic system with latent variables.

The present implementation of the particle algorithm very much follows the principles laid out in the seminal contributions by Gordon et al. (1993) and Kitagawa (1996). A particle filter allows us to approximate via simulations the likelihood function which is:

$$L(\theta) = P(y_1, \ldots, y_T|\theta) = P(y_1|\theta) \prod_{t=2}^{T} P(y_t|y_{t-1}, \theta). \tag{2}$$

Considering the dependency of the observable variable on the latent state variable, the conditional probabilities in Eq. (2) can also be written as:

$$P(y_t|y_{t-1}, \theta) = \int P(y_t|x_t) P(x_t|y_{t-1}) dx_t. \tag{3}$$

[1] Surprisingly, available papers on stochastic volatility models with leverage effect use sequential Monte Carlo methods for parameter estimation referring to the pertinent formalism for state-space models and neglect the fact that their models go beyond the standard state-space formalism (Christoffersen et al., 2010; Pitt et al., 2014; Yu, 2005).

The state variable x_t and its conditional density in Eq. (3) will be approximated by a set of 'particles' B. Hence, denoting the B particles by $x_t^{(j)}$ we approximate Eq. (3) by

$$P(y_t|y_{t-1},\theta) \simeq \frac{1}{B}\sum_{j=1}^{B} P\left(y_t|x_t^{(j)}\right). \tag{4}$$

The filtering algorithm for the evolution of the particles works as follow:

1. A random collection of $j = 1,\ldots,B$ particles is sampled from the stationary distribution of the hidden variable x_t at time $t = 1$ and the densities $P(y_1|x_1^{(j)})$ are computed.
2. Particles are resampled using weights $\frac{P(y_1|x_1^{(j)})}{\sum_{l=1}^{B} P(y_1|x_1^{(l)})}$.
3. The so obtained new set of particles is iterated to form a new sample using the stochastic process defined by the first part of Eq. (1). This process can typically only be simulated in an agent-based model. A simulation is thus conducted for each particle and the resulting value of this particle after a unit time step is recorded, which we denote as $x_2^{(j)}$.
4. Steps 1 to 3 are repeated for $t = 2, 3, \ldots, T$.

Note that the requirement of sampling from the stationary distribution of the hidden variable(s) in *step 1* is not restrictive. In most cases this distribution might not be known in closed form, but in an agent-based or other causal model these draws can easily be obtained by conducting a sufficiently long simulation. The initialization can, then, be based on a sample of hidden variables taken from a simulation that has converged to a stationary distribution. As the next step, the densities of the first observation y_1 are evaluated conditional on each particle drawn from the stationary distribution. If y_t is linked to x_t in Eq. (1) via a simple probabilistic law, the conditional densities $P(y_t|x_t^{(j)})$ may be evaluated directly. If not, this part of the algorithm would also have to be implemented by simulations. In our examples below $P(y_t|x_t^{(j)})$ will be Gaussian and, therefore, their evaluation is trivial.

It is worthwhile to note that the evaluation of $P(y_1|x_1^{(j)})$ provides us already with the approximation of the first entry of the right-hand side of the likelihood function, Eq. (2) which is obtained as:

$$P(y_1|\theta) \simeq \frac{1}{B}\sum_{j=1}^{B} P\left(y_1|x_1^{(j)}\right), \tag{5}$$

a discrete approximation of the sought-after continuous density.

In step 1, the particles have been randomly drawn from the stationary distribution so that they all enter in Eq. (5) with unitary weights. However, the evaluation of their contribution to the likelihood shows that the likelihood of observing y_1 can be very different for different particles. Those that have a higher likelihood, are, of course, more likely to be in the vicinity of the unknown first realization of the unobserved variable, x_1. In order to take into account this variation, the particles are reweighted in *step 2* by their contributions to the likelihood. To obtain proper weights summing up to unity, the weights are normalized by the sum of all likelihoods. A new set of particles is then sampled from the old ones using multinomial draws on the base of these weights. Hence, the better fitting particles have a higher chance to be selected than those with small contributions.

In *step 3* the new set of particles is iterated from time $t = 1$ to $t = 2$ using the law of motion of the unobserved variable, $x_t = f(x_{t-1}, \epsilon_t)$. In extant applications of this SMC algorithm, this approach has been applied to many types of non-Gaussian, non-linear processes for which no closed from solution is known. In our case, the law of motion of the hidden variable(s) will be a complete agent-based model of which the hidden variable(s) are summary statistics. Note that we have to apply the law of motion, i.e. the agent-based model, to the iteration of each particle $x_t^{(j)}$, hence have to simulate the agent-based model B times using each resampled particle as the starting value of one of these simulations. While in many stochastic processes, the iteration of the particles just requires one random draw from some law of motion to move from t to $t+1$, the simulation part is very likely to be more complex in an agent-based setting. First, the unit time interval over which the observable variable y_t is recorded might not be identical to the time scale of the agents' activities of the model. Rather, agents might be active a number of times between unitary observation intervals, and many of them might take some decision at the same time. In our applications, the agent-based models are formulated in continuous time, and will be simulated using a discrete event algorithm for the determination of agents' decisions. In this case, the number of events happening within the agent-based model between unit time intervals is variable and depends on the random draws of event times. Note that all these complications are easily accommodated in the particle filter as *step 3* only assumes that the law of motion for the hidden variable(s) can be simulated.

Once we have obtained the new particles at time $t = 2$, these can again be evaluated on the base of the observed variable y_2 and the entire loop starts again. Running the filter for the complete time series y_t, $t = 1,\ldots,T$ provides us with approximations of all the terms entering the overall likelihood function, Eq. (2), all being of the form of Eq. (5). It also provides us with approximations of predictive densities $p(x_t|Y_{t-1},\theta)$ and filtering densities $p(x_t|Y_t,\theta)$ for the unobserved variable(s) which are simply obtained by the discrete distribution of the particles before or after the evaluation at time step t. Slightly more complex algorithms are required to obtain an approximation to a smoothing density $p(x_t|Y_T,\theta)$ but these can also be derived on the base of the output of the particle filter.

In an estimation framework, the particle filter would be embedded in the loop for the optimization of the relevant criterion (e.g. the likelihood function). Hence, each new trial set of parameters within this loop would require a new run of the complete particle filter for $t = 1$ through T. Since we simulate the agent-based model for each particle for each time increment, the overall number of activations of the agent-based model would be $B \times T \times$ (rounds of optimization). The computational demands can, therefore, be sizable although one should note that each one of these simulations only runs for a unit time interval. As we will see, it is possible to even conduct Monte Carlo simulations of this algorithm (adding another loop) for agent-based models that are not too complex. In the next section, we will use the particle filter within both a frequentist and a Bayesian approach. In addition to these traditional avenues to identification of parameters, the statistical literature has also developed so-called *online* filtering algorithms that attempt to identify parameters after only one sweep of the particle filter through the data. Since this would restrict the number of evaluations of the agent-based model to only $B \cdot T$, it might be worthwhile to explore the efficiency of such methods vis-à -vis frequentist and Bayesian estimation. We provide, therefore, also a simple illustration of an online estimation algorithm.

2.3. Theoretical results on sequential Monte Carlo methods

Before proceeding to our examples, it seems worthwhile to review some of the basic findings from the rich statistical literature on the properties of the particle approximation. When applying the particle filter to frequentist or Bayesian estimation based on numerical approximations of the likelihood function, three questions need to be addressed:

1. The consistency and the statistical properties of the maximum likelihood estimator for the model under scrutiny,
2. The consistency and statistical properties of the particle filter approximation to the 'true', intractable likelihood,
3. The convergence and statistical properties of the optimization algorithm applied to find the maximum of the approximate likelihood function.

Consistency of maximum likelihood estimators for state space models with finite number of states has been proven by Leroux (1992) and asymptotic normality has been proven by Bickel et al. (1998). Both papers require ergodicity of the Markov process plus a few 'mild' regularity conditions. While ergodicity is easy to demonstrate for our first example below, some of the relatively innocent regularity conditions would already be harder to prove for a complex agent-based model. As it has been mentioned, the second model will belong to the more general class of dynamic models with latent variables. This class has only become a subject of research relatively recently. For this class, consistency of the maximum likelihood estimator has been shown by Douc et al. (2013, 2015), again under ergodicity and additional regularity conditions. It is worthwhile mentioning that within the more general setting of a dynamic model with latent variables, new sources of non-ergodicity exist. In practical terms, the feedback of the observed variable on the hidden one (for example, of price changes on some behavioral regularities) could give rise to unstable, and, hence, non-ergodic processes. Ergodicity might, therefore, often only apply for a restricted set of parameter values of such systems.

Moving on to the second issue, we note that the particle filter is a point-wise approximation to the density of interest. Hence, the obvious question here is whether this approximation converges to the true underlying density with increasing number of particles. Because of the dependency structure inherent to the particle evolution, the evolving particles are *not* identically and independently distributed. Hence, standard asymptotic results for the Monte Carlo simulations are not applicable. Research on this issue is summarized by Crisan and Doucet (2002) and Kantas et al. (2015). Basically, they show that under certain mixing conditions, the particle approximation converges uniformly to the true density with rate of convergence of the order $1/B$, independent of the dimension of the state space. Mixing here means that the particle filter forgets its initial conditions sufficiently fast, so that the approximation errors will not increase over time. Note that exponential forgetting leads to geometric ergodicity, so that this assumption is stronger than the conditions for consistency and asymptotic normality of the maximum likelihood estimator.

The next question is about the asymptotic properties of estimators based upon the particle filter. Again, there is a sizable literature on the asymptotics of the maximum likelihood estimator based upon particle approximations (cf. Cappé et al. 2005; Kantas et al. 2015). First, under relatively weak regularity conditions, the maximum likelihood estimator can be shown to be consistent for a given sample size and a central limit law applies as the number of particles goes to infinity. However, the error of the approximation is increasing with sample size T. If the filtering density for the hidden variables obeys exponential forgetting of initial conditions[2], the relative variance of the likelihood estimate can be shown to be bounded above by a factor $D_\theta \frac{T}{B}$ where D_θ is exponential in the number of parameters in the set θ (cf. Kantas et al. 2015). Tadić and Doucet (2005) show that exponential forgetting of the filter density applies to a large class of Markov processes of the unobserved variable(s) under mild regularity conditions. They show that these conditions are typically met if the Markov process is constraint to values in a compact space. Since in our first model the latent variable is indeed constraint to a compact set, the exponential forgetting property should typically be satisfied (although it might again be difficult to provide a formal proof of the specific set of regularity conditions established in Tadić and Doucet (2005)). Experimentally, we

[2] Exponential forgetting is defined by the condition $||p_\theta(dx_t|Y_t, x_0) - p_\theta(dx_t|Y_t, x_0')|| \leq B_\theta \lambda^t$ for some $B_\theta < \infty$ and $\lambda \in [0, 1)$ where $p_\theta(dx_t|Y_t, x_0)$ is the filtering density for the unobservable variable given the observations $y_1, \ldots y_t$ and initial conditions x_0 and x_0', respectively and the notation $||\cdot||$ denotes the total variation distance.

can easily shed light on the mixing property by choosing different initial conditions. In all such experiments, the influence of the initial draws have been found to be washed out within a few time steps. To our knowledge, the only result available so far for the more general class of models with latent variables are those derived in two recent microeconomic applications: Gallant et al. (2016) show that maximum likelihood estimation based on the particle filter is unbiased while Blevins (2016) demonstrates its asymptotic consistency.

A practical obstacle to ML estimation based on a particle filter is that the resulting likelihood function is non-smooth in the underlying parameters. This is so because the multinominal draws in the resampling step would lead to discrete changes under continuous variations of parameters even if the same random numbers were used during the optimization process. Malik and Pitt (2011) have proposed a simple transformation of the filtering density to obtain a smooth approximation to the likelihood. Unfortunately, this approach is restricted to univariate state spaces. In typical agent-based models like the ones presented in Section 3, it would also not be sufficient. The reason is that the discrete variation brought about by multinominal draws is not only restricted to the particle filter, but a similar step already occurs when simulating the discrete choice of agents in the state dynamics. Hence, it seems practically unavoidable to try and estimate parameters on the base of a non-smooth approximation of the likelihood that does not allow easy determination of its optimum with gradient-based optimization routines.

This leads us to the question 3 above on the convergence of some optimization algorithm using the particle filter approximation of the likelihood. In the statistical literature, Olsson and Rydén (2008) have addressed this problem. They use a grid-based approach and ask the question how the number of grid points and particles should increase with T to guarantee consistency and asymptotic normality of the estimated parameters. When extending this point-wise approximation to the complete parameter space via piecewise constant functions or smoothing via interpolation by splines they obtain the following requirements: In the first case consistency is obtained if the grid size goes to zero at an arbitrary rate and B increases faster than $M^{2/r}$ for some integer $r \geq 1$ (M the number of grid points). For asymptotic normality the grid size has to decrease faster than $1/T$ and the number of particles has to increase faster than $M^{2/r}T^2$. In the case of a spline approximation, the requirements for consistency are the same while asymptotic normality is obtained if the grid size decreases faster than $1/\sqrt{T}$ with the same increase of the number particles exceeding $M^{2/r}T^2$ as in the previous case. Besides certain regularity conditions, the proofs assume multinominal resampling as described in the particle filter scheme above. Another theoretical result can be found in Ionides et al. (2011). These authors show that a particular iterative filtering algorithm (meaning that one repeats a number of times an *online* estimation approach) provides parameter estimates that converge to the maximum likelihood estimates.

Empirical applications of particle filters in economics have not used these particular approaches, but have resorted to simulated annealing (e.g.Fernández-Villaverde and Rubio-Ramírez 2007) or related optimization techniques for a non-smooth objective function. In the present application, the time honored Nelder-Mead simplex method is used[3]

The numerical approximation of the likelihood function cannot only be used for frequentist estimation of the parameters, but can also be used within a Bayesian approach. Indeed, Bayesian estimation is often used in extant literature in combination with a particle filter. A number of applications have adopted the approach by Andrieu et al. (2010) who propose to use the particle filter within a Metropolis-Hastings sampler of the posterior density of the parameters. In this approach, one attempts to simulate a Markov chain that converges to a stationary distribution that is the posterior distribution of the parameters. For such a Bayesian approach, one needs a prior, say $p(\theta)$, and a proposal density, say $g(\theta_\delta|\theta_{\delta-1})$ where δ is the sequential order of the chain. New draws θ^* from $g(\cdot)$ are then accepted with probability:

$$\alpha(\theta^*|\theta_{\delta-1}) = \frac{P_{\theta^*}(y)p(\theta^*)g(\theta_{\delta-1}|\theta^*)}{P_{\theta_{\delta-1}}(y)p(\theta_{\delta-1})g(\theta^*|\theta_{\delta-1})} \tag{6}$$

with $P_\theta(y)$ the marginal likelihood of the observed data given θ, and $0<\alpha<1$ is imposed as additional constraint. Andrieu et al. (2010) have shown that this chain converges to the posterior density under very general conditions on the likelihood and the proposal density. We will illustrate the implementation of this approach alongside with frequentist ML in the next section.

Still another way to estimate parameters within a particle filter approach is what is called *on-line estimation*, i.e. one attempts to estimate the parameters together with the state filtering *in a single sweep* through the available observable time series. The simplest such approach consists in expanding the state space by the parameters (c.f. Kitagawa (1998)). One then initiates a particle swarm consisting of $\{x_t^{(j)}, \theta_t^{(j)}\}$ and iterates it via multinominal resampling in exactly the same way as in the general scheme outlined above. The drawback of the proposal of Kitagawa (1998) is that the particles are stationary $(\theta_{t+1}^{(j)} = \theta_t^{(j)})$ and so over time the second part of the particle swarm will eventually become degenerate because more and more of the original particles will get lost. Despite this problem, this approach might sometimes yield very competitive estimates given the very limited computational demands of this approach (see below). Different improved versions of on-line estimators can be found in Carvalho et al. (2010) and Ionides et al. (2011).

In economics, particle filters have been applied for estimation of DSGE models (Amisano and Tristani, 2010; Fernández-Villaverde and Rubio-Ramírez, 2007), stochastic volatility models (Bao et al., 2017; Pitt et al., 2014; Yu, 2005) and

[3] We have also experimented with the stochastic Nelder–Mead algorithm proposed by Chang (2012). The latter algorithm replaces the shrinkage step of the original Nelder–Mead approach by an adaptive random search that either draws a new coordinate for the parameter values from a neighborhood of the current simplex or globally from the admissible range of parameter values. Results turned out to be very similar.

regime-switching models (Billio and Casarin, 2010). Very recently, particle filters have also been adopted to estimate dynamic microeconomic models (Blevins, 2016; Gallant et al., 2016). However, their models are of a more general format than the baseline state-space formalism underlying most of the available asymptotic results in the statistical literature. Like our second example, their models allow for feedback from the observed variable on the hidden states. While these papers show consistency of the likelihood approximation based on a particle filter, it is not known in how far the larger body of theoretical results reviewed above still applies for the more general format of a dynamic system with latent variables.

3. Two simple agent-based models

In the following we will present the prototype models. The first model is the agent-based approach to sentiment dynamics proposed by Alfarano et al. (2008) that had already been estimated via GMM by Ghonghadze and Lux (2016) and via SMM by Chen and Lux (2017). In this model, the asset price dynamics is determined by fundamental news and sentiment. In continuous time, log price changes are defined as:

$$\frac{dp}{dt} = \beta(T_f(p_{f,t} - p_t) + NT_c x_t) \tag{7}$$

with: $p_{f,t}$ the log fundamental value at time t, p_t the current log market price, and x_t a measure of sentiment. The parameters T_f and T_c denote the overall trading volume of fundamentalist and sentiment-prone agents. The latter is a per-head measure and is, therefore, multiplied by the number of non-fundamental traders, N. The parameter β, finally, is the price adjustment speed. If one assumes instantaneous price adjustment to clear the market, β drops out as a free parameter and the current equilibrium price would be attained as:

$$p_t = p_{f,t} + \frac{NT_c}{T_f} x_t. \tag{8}$$

It follows that returns, r_t, can be decomposed into fundamental news and sentiment innovations:

$$r_t = p_{t+1} - p_t = p_{f,t+1} - p_{f,t} + \frac{NT_c}{T_f}(x_{t+1} - x_t). \tag{9}$$

The log fundamental news is assumed to follow standard Brownian motion:

$$p_{f,t+1} - p_{f,t} = \sigma_f \varepsilon_{f,t}, \qquad \varepsilon_{f,t} = N(0,1). \tag{10}$$

In (10) and (11), we have defined returns over a unit time interval as we will estimate our continuous-time model on the base of discretely observed data whose sampling frequency we define as a unit time interval. The above framework is in line with a large literature in financial economics that investigates the influence of sentiment as an important asset pricing factor. This literature would typically attempt to identify good proxies for overall market sentiment x_t and assess its explanatory power for asset price movements, e.g. Brown and Cliff (2004), while here we treat sentiment as a latent factor we attempt to retrieve from filtered state probabilities.

The model of Alfarano et al. (2008) adds a theoretical module explaining the time evolution of sentiment itself. This component follows essentially the seminal herding model of Kirman (1993). Inspired by a variety of related phenomena in the social and natural sciences (most notably recruitment processes in ant colonies), it supposes that sentiment-prone agents are either in an optimistic or pessimistic mode and switch between both types of sentiment both due to idiosyncratic random factors and systematic conversion after meetings with agents of different disposition. In continuous time, this amounts to a formalization by Poisson transition rates that are dependent on the current state of the population of sentiment traders (whose number is N). Denoting by $\pi_{+,t}$ and $\pi_{-,t}$ the transition rates from the pessimistic into the optimistic camp and *vice versa*, and by n_t the current number of optimistic agents, the 'ant process' formalizes the transition rates as

$$\pi_{+,t} = a + n_t b, \qquad \pi_{-,t} = a + (N - n_t)b. \tag{11}$$

In Eq. (12), a is an idiosyncratic rate of change of opinion and b is the susceptibility to conversion when meeting an agent of the opposite attitude. The chances to meet agents from the other camp are proportional to their current numbers in the population, i.e. n_t and $(N - n_t)$. As transition rates, both a and b are restricted to non-negative values. On the base of the current distribution of the optimistic/pessimistic disposition, overall sentiment is defined as:

$$x_t = \frac{n_t - (N - n_t)}{N} = \frac{2n_t - N}{N}, \tag{12}$$

so that positive (negative) values indicate a dominance of optimistic (pessimistic) traders, and sentiment is normalized to be measured over the interval [−1,1]. Various properties of the sentiment dynamics itself and the asset price process including such sentiment formation are known (Kirman 1993; Alfarano et al. 2008). For instance, the stationary distribution of x_t is either unimodal or bimodal depending on the strength of the herding component vis-à-vis the idiosyncratic change of the opinion. The precise condition for unimodality (bimodality) is $a > (<)\ b$. Returns obtained by Eq. (10) share the universally observed fat tails and clustered volatility of empirical asset returns.

The second model takes its inspiration from Franke and Westerhoff (2012). It adds what is a standard feature of many agent-based models of asset price formation, but is missing in the Alfarano et al. (2008) framework: influence of price trends

on chartists' demand. Note that the absence of this factor is what guarantees independence of the state dynamics from the observations and, therefore, makes the former model fit into the standard state-space format. However, most agent-based behavioral asset pricing models would allow for some feedback from past price histories on current equilibrium prices via chartists' demand function (cf. Chiarella (1992), for one of the first formalisations and the many variations of chartist-fundamentalist models surveyed in the contributions to Hens and Schenk-Hoppé (2009)). We again assume that there are two components of excess demand, but in contrast to the previous model we now allow agents to switch between the chartist and fundamentalist types.

The number of the former is denoted by $N_{c,t}$, the number of the latter by $N_{f,t}$. We assume the simplest forms of their trading strategies leading to excess demand $ED_{f,t}$ and $ED_{c,t}$ as follows:

$$ED_{f,t} = N_{f,t}\ a\ (p_{f,t} - p_t),\ \ ED_{c,t} = N_{c,t}\ b\ (p_{t-1} - p_{t-2}) \tag{13}$$

with reaction coefficients a and b.

Assuming again instantaneous equilibration of the market, the log asset price in equilibrium develops analogously to Eq. (9):

$$\begin{aligned} p_t &= p_{f,t} + \frac{N_{c,t}}{N_{f,t}}\frac{b}{a}(p_{t-1} - p_{t-2}) \\ &= p_{f,t} + \frac{z_t}{1 - z_t}\ c\ (p_{t-1} - p_{t-2}) \end{aligned} \tag{14}$$

with $c = \frac{b}{a}$ and $z_t = \frac{N_{c,t}}{N}$ (N being the overall number of agents).

Similarly as in the model of Alfarano et al. (2008), we allow for time-variation of the two fractions of agents. However, in the present framework, the herding formalization of Eq. (12) would lead to unrealistic and degenerate results: In the bimodal case, when z_t fluctuates at high values close to one, we would enter into an explosive price trend that would soon reach the limits of computability. There are various ways to modify or embellish the herding dynamics to avoid such an outcome. Here, we just modify the herding process by using a different, also popular formalization. Namely, we adopt the seminal herding formalism introduced by Weidlich and Haag (1983) and adopted by, for example, Lux (1995) and Franke and Westerhoff (2012). Denoting the transition rates from the fundamentalist to chartist strategy and *vice versa* by $\pi_{c,t}$ and $\pi_{f,t}$, these are defined as:

$$\pi_{c,t} = \nu\ exp^{\alpha_1 x_t},\ \ \pi_{f,t} = \nu\ exp^{-\alpha_1 x_t} \tag{15}$$

with $x_t = \frac{N_{c,t} - N_{f,t}}{N} = \frac{2N_{c,t}}{N} - 1 = 2z_t - 1$.

This formalization also leads to unimodal or bimodal behavior of the population, now depending on whether the herding intensity α_1 is ≤ 1 or > 1. However, in the bimodal case, the two modes are not located at the extremes but at some finite values x_+^* and $x_-^* = -x_+^*$ with a dominance of one of both groups. Furthermore, even if $\alpha_1 < 1$, the resulting series of returns are characterized by fat tails and clustering of volatility due to the repeated fluctuations of agents between both camps. In terms of the latent variable framework, we can interpret x_t and z_t as the hidden variables. Via Eq. (14), these variables enter excess demand so that using the latter as the unobserved state variable, we have a structure as it has been depicted on the right-hand side of Fig. 1.

We note that we simulate both models as 'true' agent-based models. For both of them, we assume that the market consists of a pool of $N = 100$ speculators switching between strategies according to Eq. (11) or Eq. (15), respectively. The simulations are exact simulations of the continuous-time Poisson processes characterizing the agents' movements between groups. Such an exact simulation is obtained by drawing random numbers for the discrete events of agents switching from one group into another. After each such change, the transition rates are updated according to the changing influences on agents' decisions that come along with the resulting change of the overall composition of the ensemble of agents. Using the new transition rates, the next discrete event is determined. In each discrete event, the agent undergoing a change of strategy and the pertinent change are determined. While both models could be approximated by structural equations and simulated with less computational effort (cf. Alfarano et al. 2008; Franke and Westerhoff, 2012; Barde, 2016), we have chosen the discrete event approach here for two different reasons: First, since this is an exact simulation of the model, the resulting Monte Carlo runs do not suffer from an approximation bias. If we would use an approximation via stochastic difference or differential equations, a certain bias would be caused by this approximation which we could not disentangle from the estimation bias and variability of our estimators. In an exact simulation, no distortion of parameters is to be expected, and, hence, all potential biases can be attributed to the performance of the estimation methods. Second, using a system of 'autonomous' agents in the simulations, we obtain a more realistic perception of the computational demands of SMC methods when applied to typical agent-based models.

4. Monte Carlo experiments

In the following, we provide results for estimation of the parameters of the two agent-based models of Section 3 using both frequentist and Bayesian approaches as well as the elementary on-line estimation procedure proposed by

Table 1
Sequential Monte Carlo estimation of ALW model: estimation via maximum likelihood.

Parameter	a	b	σ_f	tsec	a	b	σ_f	tsec	a	b	σ_f	tsec
Parameter set I:	a = 0.3,	b = 1.4,	$\sigma_f = 30$									
	T = 1000 B = 1000				T = 2000 B = 250				T = 2000 B = 1000			
Mean	0.330	1.708	30.981	250.760	0.368	1.693	31.478	71.320	0.370	1.600	30.411	508.430
FFSE	0.116	0.532	3.564	29.603	0.091	0.400	2.262	11.935	0.080	0.347	1.919	54.664
RMSE	0.119	0.612	3.679		0.113	0.494	2.693		0.106	0.399	1.953	
	T = 1000 B = 2000				T = 2000 B = 500				T = 2000 B = 2000			
Mean	0.350	1.674	30.552	804.030	0.362	1.624	31.030	178.850	0.358	1.564	30.192	1601.200
FSSE	0.117	0.529	2.903	67.068	0.076	0.331	1.970	23.009	0.095	0.361	1.604	107.014
RMSE	0.126	0.593	2.940		0.098	0.398	2.215		0.111	0.395	1.608	
param. set II:	a = 1.4,	b = 0.3,	$\sigma_f = 30$									
	T = 1000 B = 1000				T = 2000 B = 250				T = 2000 B = 1000			
Mean	1.876	0.329	29.655	245.540	1.931	0.288	31.196	65.390	1.878	0.316	30.277	486.620
FSSE	0.363	0.067	2.352	10.570	0.481	0.080	1.846	3.944	0.419	0.082	2.302	22.065
RMSE	0.598	0.073	2.366		0.715	0.080	2.191		0.634	0.084	2.302	
	T = 1000 B = 2000				T = 2,000 B = 500				T = 2000 B = 2000			
Mean	1.864	0.323	29.577	787.780	1.971	0.312	30.409	168.680	1.918	0.306	30.238	1562.850
FSSE	0.343	0.066	2.303	28.244	0.367	0.063	2.137	8.721	0.362	0.066	2.175	53.723
RMSE	0.576	0.070	2.330		0.678	0.064	2.165		0.631	0.066	2.177	

Notes: The table shows the means, finite sample standard errors (FFSE) and root-mean squared errors (RMSE) of 100 replications of each scenario. The underlying parameters of the simulated models are multiplied by 10^3 for better readability. *tsec* denotes the time needed for the execution of one of the estimations in seconds.

Kitagawa (1998). We start with the model of Alfarano et al. (2008). Previous studies (Ghonghadze and Lux, 2016; Jang, 2015) have found it difficult to estimate T_c/T_f and N, the former because of near colinearity with other variables, the latter because of its discrete nature. We, therefore, follow these studies by fixing these parameters at $T_c/T_f = 1$ and $N = 100$ so that the set of parameters to be estimated is $\theta = \{a, b, \sigma_f\}$. One detail worth mentioning is the initialization of the particle filter. The initial set of particles should best be generated using the stationary distribution of the hidden variable. Unfortunately, this distribution is not known for $z_t = x_t - x_{t-1}$ in the model of Alfarano et al. (2008). However, at least an approximate solution is known for the stationary distribution of x_t (cf. Alfarano et al. 2008):

$$P_e(x) = \frac{\Gamma(2\varepsilon)}{2^{2\varepsilon-1}\Gamma(\varepsilon)^2}(1-x^2)^{\varepsilon-1} \tag{16}$$

where $\Gamma(\cdot)$ denotes the Gamma function and $\varepsilon = \frac{a}{b}$. Hence, we might proceed in this case by first generating draws of x_t at time $t = 0$ from its stationary distribution (16), and apply the stochastic process of Eq. (11) to generate a stationary ensemble of particles $z_1^{(j)} = x_1^{(j)} - x_0^{(j)}$ that can be plugged into Eq. (9) for the observable variable returns.

For the model of Franke and Westerhoff, no results on the stationary distribution of any of the state variables are available. We, therefore, initiate the particle filter via uniform random draws of its sentiment index x_t. We have actually also conducted experiments with the same random initialization for the Alfarano et. al. model and found no perceptible differences in the Monte Carlo results. When performing comparisons of the goodness-of-fit of candidate models, one might take into account this slight asymmetry by discarding a certain number of observations at the beginning of a time series to allow the particle filter to wash out the influence of the initializations.

We now turn to the results of the Monte Carlo simulations. We start with the model of Alfarano et al. (2008). Table 1 exhibits the results of 400 Monte Carlo simulations of each of a variety of settings. First, we have selected two different sets of parameter values following Chen and Lux (2017) and Ghonghadze and Lux (2016): The first has a high interactive component in agents' opinion formation ($b/a > 1$) and, hence, leads to pronounced fat tails and volatility clustering, while in the second parameter set, agents are rather autonomous in their opinions ($b/a < 1$) and so the time series behavior of the model is closer to Brownian motion. Second, we consider sample sizes of both $T = 1000$ and 2000 observations as well as a number of particles ranging from $B = 250$ to $B = 2000$.

Given the theoretical results reviewed in Section 2, we would not expect square-root consistency with respect to sample size under the same number of particles. Nevertheless, we find a clear improvement in the quality of our estimates for at least one of the parameters (σ_f) when moving from $T = 1000$ to $T = 2000$ in the first set of parameters. In contrast, there is little improvement in the quality of our estimates when increasing the number of particles for a given length of the sample. This holds even more for the second set of parameters where the differences between all six scenarios with different numbers of T and B are minor. The table also shows average computation times across the 100 replications of each

Table 2
Estimation of ALW model via self-organizing state space.

Parameter	a	b	σ_f	tsec	a	b	σ_f	tsec	a	b	σ_f	tsec
Parameter set I:	a = 0.3,	b = 1.4,	$\sigma_f = 30$									
	T = 1000 B = 200				T = 2000				T = 2000, stratified			
Mean	0.923	0.727	28.202	0.380	1.024	0.817	27.830	0.620	0.921	0.805	30.166	0.650
FSSE	0.542	0.535	9.509	0.488	0.574	0.552	8.934	0.488	0.537	0.528	7.875	0.479
RMSE	0.824	0.858	9.631		0.922	0.801	9.268		0.820	0.794	7.838	
	T = 1000 B = 2000				T = 2000				T = 2000, stratified			
Mean	0.509	0.937	30.097	0.940	0.482	1.005	30.235	1.970	0.528	1.037	29.036	9.760
FSSE	0.235	0.537	6.450	0.312	0.257	0.547	5.936	0.559	0.263	0.494	8.415	1.478
RMSE	0.314	0.706	6.419		0.314	0.672	5.910		0.347	0.611	8.428	
	T = 1000 B = 20,000				T = 2000				T = 2000, stratified			
Mean	0.489	1.118	29.037	6.980	0.482	1.324	29.493	13.740	0.512	1.275	28.933	751.220
FSSE	0.227	0.462	4.412	1.504	0.262	0.394	3.429	2.841	0.276	0.486	8.296	105.575
RMSE	0.294	0.540	4.495		0.318	0.400	3.450		0.347	0.500	8.323	
	T = 1000 B = 200,000				T = 2000							
Mean	0.479	1.255	28.900	67.840	0.429	1.369	29.583	133.000				
FSSE	0.132	0.404	3.163	13.973	0.176	0.266	1.680	27.636				
RMSE	0.222	0.427	3.334		0.218	0.266	1.723					

Notes: The table shows the means, finite sample standard errors (FSSE) and root-mean squared errors (RMSE) of 100 replications of each scenario. The underlying parameters of the model are multiplied by 10^3 for better readabilty. The first two panels use binomial sampling, the third uses stratified sampling. For the latter, the case $B = 200,000$ has not been implemented as it turned out to be too costly in terms of computation time. *tsec* denotes the time needed for the execution of one of the estimations in seconds.

scenario. As it would have been expected, computational demands increase almost perfectly linearly with sample size T, but when changing the number of particles the computation time increases more than proportionally.

The precision of our estimates seems overall better than with moment-based estimates. Indeed, in Chen and Lux (2017) and Ghonghadze and Lux (2016) no results for time series of length smaller than $T = 5000$ are shown and results for their smallest sample size seem to be inferior to those of Table 1. We also note, however, that Monte Carlo simulations with $T = 5000$ or higher would be very expensive for the present approach which illustrates the higher need in computing time vis-à-vis moment based estimates. To illustrate the much higher computational demands of sequential Monte Carlo estimation compared to SMM, we conducted Monte Carlo exercises similar to the ones reported in Chen and Lux (2017). Using 15 moment conditions we considered sample sizes $N_1 = 10,000$, $N_2 = 20,000$ and $N_3 = 100,000$, together with $S = 4N$ as the size of the simulated samples. Average computation time (standard deviation in brackets) of 100 replications of these scenarios were: 74.310 (9.070) seconds for sample size N_1, 134.900 (12.815) seconds for N_2 and 594.320 (31.769) seconds for N_3. Scaling these down to sample sizes $T = 1000$ and 2000, we would get estimation done much faster with SMM than SMC but with large sacrifices in terms of precision.

Table 2 presents results using the on-line approach of Kitagawa (1998). Since here the particle filter always only is applied over one single sweep through the data, we could afford to use up to $B = 200,000$ particles. In the maximum likelihood approach of Table 1, we have set the maximum number of iterations of the Nelder–Mead algorithm to 100. Since in every iteration, a three-digit number of evaluations of the likelihood function is conducted, the time needed for one estimation with the standard frequentist approach is about three orders of magnitude more compared to the one pseudo-evolutionary sweep in the 'self organizing state space' approach. We see that for the largest size of $B = 200,000$, the results are of about the same quality for the parameters b and σ_f as with frequentist maximum likelihood albeit with different performance for different parameters. This is also the only case in which we observe a more pronounced difference in the Monte Carlo results for sample sizes $T = 1000$ and 2000.

With smaller numbers of B, the set of auxiliary particles for the parameters, $\theta_t^{(j)}$, just becomes degenerate long before the end of the time series, and, hence, more observations cannot change the outcome of this on-line estimation. In the right-hand part of Table 2, we also show results using what is called stratified sampling rather than multinomial draws from a distribution of weights. Stratified sampling is one among a number of methods that that could be used to reduce sampling variability in the resample step of the particle filter. It consists in partitioning the interval (0, 1] into disjoint sets and sampling independently from each of these sets. In this way, closer proximity of the overall sample to its target distribution can be inforced. This should guarantee a more representative distribution of the particles, and, thus, should help to eliminate sampling noise. Because of the construction of the subsets, this approach, however, includes higher computational costs than multinomial sampling. As it turns out, this modification does not have a clearly positive effect on the outcome of the Monte Carlo experiments. Because of its higher computational demands, we have also executed the above experiments only for up to $B = 20,000$ particles.

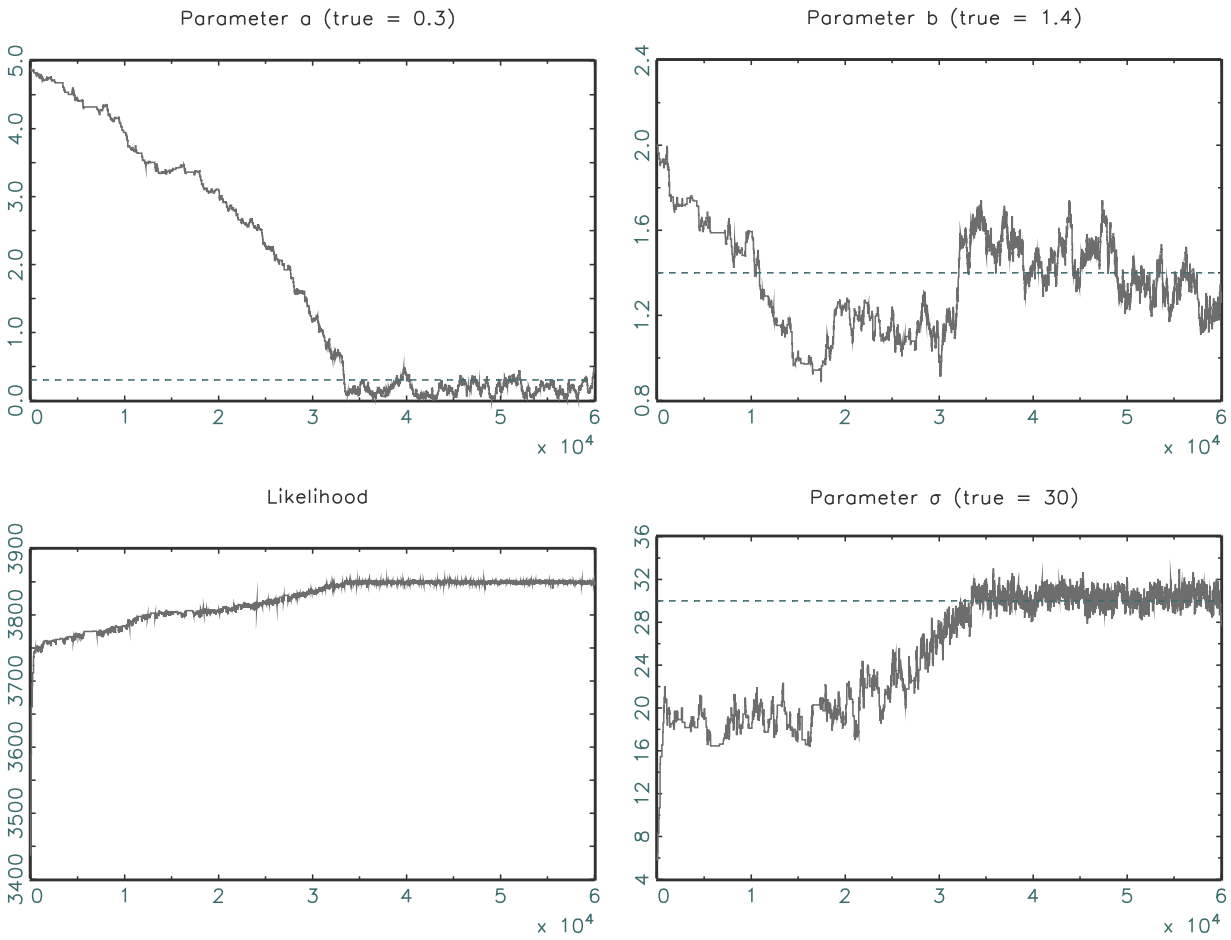

Fig. 2. Bayesian estimation using the Particle Markov Chain Monte Carlo (PMCMC) algorithm applied to the model of Alfarano et al. (2008). The four panels show the evolution of the three parameters a, b and σ_f (clockwise from the upper left panel) and the filtered likelihood (lower left panel) over 60,000 iterations of the Markov Chain. The present example shows a more extreme realisation of the length of the transient period prior to convergence to the stationary distribution than the one used for the statistics of Table 3.

Finally, Table 3 and Fig. 2 turn to the Bayesian approach, denoted as Particle Markov Chain Monte Carlo (PMCMC) by Andrieu et al. (2010). Fig. 2 illustrates the evolution of the likelihood and the parameters in the Markov Chain generated according to Eq. (6) over a length of $N = 60,000$ iterations. Since in empirical applications, we would have hardly any good a priori clue on the range of the expected parameter estimates, wide uniform priors have been chosen: For parameters a and b we have assumed a uniform prior over the interval [0, 5], while for σ_f the prior was taken to be uniform over $[0, \sigma_{emp}]$ where σ_{emp} was the empirical standard deviation of returns in the underlying sample. The reason for this choice is that this is the highest value that σ_f can attain if the sentiment process were absent and all of the variation of returns were explained by changes of fundamentals. The 'pseudo-empirical' sample used in Fig. 2 had 2000 observations. What is immediately obvious from Fig. 2 is the long transient part of the evolution: It takes more than 30,000 iterations until eventually the stationary posterior distribution is attained. While this illustration exhibits an extreme example, the potentially very long transient apparently could lead to a huge waste of computational resources, or, if resources are limited, they might not allow to arrive at the stationary distribution of the posterior. Because of the strong increase of computation time, we only conducted this experiment with particle numbers up to $B = 2000$. In Table 3 we report the statistics of a less extreme case for which we have based the estimation on the last 30,000 out of 60,000 iterations (with sample sizes $T = 1000$ and $T = 2000$). Depending on the randomly chosen starting value, convergence can also indeed occur much earlier.

Table 3 shows the mean of the posterior distribution in the so chosen asymptotic regime. While the means appear close the their 'true' values, the computed standard errors appear relatively large. One should note, however, that these errors are the standard errors of one single estimation, while the FSSE and RMSE are reported for the whole set of Monte Carlo runs in Tables 1 and 2. One notes that the standard errors taking into account the persistence of the chain are much higher than the naive ones that one would compute under independence of the draws. While one finds smaller standard errors for the

Table 3
Estimation of ALW model via Bayesian PMCMC algorithm.

Parameter set I: a = 0.3, b = 1.4, σ_f = 30,							
	T = 1000,	N = 30,000					
	mean	s.e	s.e.naive	ineff.	cov. matrix		
a	0.309	1.510	0.190	63.16	1.000	0.029	−0.460
b	1.431	1.416	0.222	40.68	0.029	1.000	−0.008
sig	31.359	8.302	1.318	39.68	−0.460	−0.008	1.000
logl	1786.785	8.435	1.582				
tsec	128,976						
	T = 2000,	N = 30,000					
a	0.204	0.767	0.110	48.31	1.000	0.081	−0.535
b	1.419	1.433	0.202	50.09	0.081	1.000	−0.022
sig	30.187	4.749	0.847	31.47	−0.535	−0.022	1.000
logl	3849.316	8.852	1.679				
tsec	241,809						

Notes: The table shows the means and standard errors of two Monte Carlo simulations of the posterior distribution using the PMCMC algorithm. The 'inefficiency' is the ratio between the autocorrelation consistent estimate of the standard error in the second column and a naive estimate assuming independent draws. The sample statistics are based on a Monte Carlo simulation of 30,000 draws after discarding an initial transient of 30,000 iterations. The pseudo empirical sample had T = 1000 (2000) observations, and B = 2000 particles have been used. The acceptance rate of new draws was 0.448 and 0.380. *tsec* denotes the time needed for the execution of one of the estimations in seconds.

Table 4
Sequential Monte Carlo estimation of FW model: estimation via maximum likelihood.

Parameter set: v = 1, α_1 = 0.85, b = 2.0, σ_f = 30										
Parameter	v	α_1	b	σ_f	tsec	v	α_1	b	σ_f	tsec
	T = 1000,	B = 1000				T = 2000,	B = 1000			
Mean	1.490	0.692	1.881	30.187	5827.710	1.518	0.688	1.880	30.086	11718.650
FSSE	0.539	0.619	0.205	1.628	1703.660	0.512	0.761	0.140	1.325	3290.842
RMSE	0.726	0.636	0.236	1.631		0.727	0.774	0.184	1.321	

Notes: The table shows the means, finite sample standard errors (FSSE) and root-mean squared errors (RMSE) of 100 replications. The underlying parameter σ_f has been multiplied by 10^4 for better readability. *tsec* denotes the time needed for the execution of one of the estimations in seconds.

larger sample throughout, these are results extracted from only two samples so that it is not obvious in how far this would hold more generally.

The main drawback of the Bayesian approach is its need of computational resources: A little reflection reveals that producing the one example of Fig. 2 and Table 3 takes about as long as a complete set of 100 Monte Carlo replications of the frequentist approach before. What is more: to design a Markov chain with reasonable properties, a preliminary series of computations is necessary to fine-tune the number of particles B and the parameters of the proposal distribution $g(\theta^*|\theta_{\delta-1})$ in Eq. (6). The usual target is an acceptance rate of about 0.4. As Table 3 shows, we get relatively close to this target in our present setting. This, however, has only been possible by another preliminary computation intensive set of experiments with varying number of particles B, and varying standard deviations of the proposal distribution $g(\cdot)$. For the latter we followed the popular choice of Gaussian proposals. Note that the target acceptance rate of ≈ 0.4 should provide for good mixing properties of the Markov chain. This should, on the one hand, prevent that the evolution gets stuck for long time at parameters far away from those that maximize the likelihood, and should, on the other hand, also keep the number of 'wasted', i.e. non-accepted trials in bounds. Nevertheless, despite our fine-tuning, we sometimes need a very long simulation run to observe convergence to a seemingly stationary distribution of the posterior as illustrated in Fig. 2. With sub-optimal choices, this problem would be exacerbated and we would often not find convergence even after 100,000 simulation steps although theoretically the Markov chain's eventual convergence to the posterior might still be guaranteed.

Tables 4 and 5 report similar results on the base of the model by Franke and Westerhoff which because of the dependency of the agents' strategies on past price records is an example of a dynamic model with latent states (right-hand side of Fig. 1). While the particle filter has been developed for state space models, the sequence of operations remains exactly the same for more general systems of stochastic processes with latent variables. When estimating our second model, we, thus, simply take into account in step 3 the dependency of the state variable (which we might define as $ED_t = ED_{f,t} + ED_{c,t}$ as introduced in Eq. (13)) on the observable variable p_t. Note that the resulting system can be brought into standard Markov form by defining two observables: p_t and $\tilde{p}_t = p_{t-1}$. Due to the increase of the number of parameters, computations become even more time consuming, and so we provide details on a limited set of examples only. In all the subsequent simulations, we have again fixed the number of agents at $N = 100$, and have set the parameter $a = 1$ as it is colinear with b and, hence,

Table 5
Estimation via self-organizing state space of FW model.

Parameter set: v = 1, $\alpha_1 = 0.85$, b = 2.0, $\sigma_f = 30$										
Parameter	v	α_1	b	σ_f	tsec	v	α_1	b	σ_f	tsec
T = 1000						T = 2000				
B = 200										
Mean	2.403	0.597	2.269	32.660	5.310	2.579	0.622	2.541	32.336	11.050
FSSE	1.435	0.315	0.843	5.840	2.960	1.518	0.360	0.975	5.733	6.333
RMSE	2.002	0.403	0.881	6.390		2.185	0.425	1.098	6.163	
B = 2000										
Mean	2.231	0.720	2.218	30.379	48.180	1.888	0.749	2.185	30.942	82.900
FSSE	1.455	0.312	0.753	2.465	25.916	1.342	0.280	0.767	3.191	52.813
RMSE	1.900	0.336	0.781	2.482		1.604	0.296	0.786	3.312	
B = 20,000										
Man	1.599	0.835	2.191	30.054	403.640	1.668	0.828	2.046	30.257	781.160
FSSE	0.982	0.135	0.618	1.678	149.373	1.024	0.140	0.399	1.867	352.432
RMSE	1.100	0.136	0.644	1.671		1.219	0.141	0.399	1.875	
B = 200,000										
Mean	1.256	0.836	2.033	29.930	3637.100	1.252	0.847	1.990	29.978	6331.790
FSSE	0.511	0.106	0.281	1.232	1047.180	0.503	0.081	0.233	0.988	1743.945
RMSE	0.569	0.106	0.281	1.228		0.560	0.081	0.232	0.983	

Notes: The table shows the means, finite sample standard errors (FSSE) and root-mean squared errors (RMSE) of 100 replications of each scenario. The underlying parameter σ_f has been multiplied by 10^4 for better readability. *tsec* denotes the time needed for the execution of one of the estimations in seconds.

cannot be estimated independently. The remaining parameters have been chosen so as to obtain realistic time series, i.e. returns exhibiting clustered volatility and fat tails.

Table 4 exhibits results of the frequentist estimation using again 100 replications and considering only the case $B = 1,000$ together with $T = 1,000$ and $T = 2,000$. Again, parameter estimates appear reasonably close to their 'true' values. We also observe that estimating a model with four rather than three parameters leads to an increase of computation time of roughly one order of magnitude.

Table 5 provides results on the performance of the on-line approach, again showing averages over 100 Monte Carlo runs, and using up to $B = 200,000$ particles as in the case of the Alfarano et al. model. Here we observe a monotonic improvement with increasing number of particles, and for $B = 20,000$ and $B = 200,000$ a longer time series also leads to better identification of some of the parameters on average. With $B = 200,000$ all parameters are estimated practically without bias and (except for ν) signal-to-noise ratios are relatively small across Monte Carlo runs. Again, computation times increase by at least one order of magnitude when compared to those for three-parameter model documented in Table 2. It appears remarkable that for the Franke and Westerhoff model, the self-organizing state space approach almost completely dominates the Nelder-Mead estimation. For instance, using 200,000 particles, the naive filtering approach provides both more precise estimates of almost all parameters, being at the same time much cheaper in computation time than the Nelder-Mead optimization with 1000 particles.

We do not show results of the Bayesian approach as a preliminary series of experiments with varying B and varying parameters of the proposal distribution did not produce any combination that would have resulted in an acceptance ratio anywhere near the targeted 40 percent. Indeed, we could hardly infer any clear tendency for the parameters involved which did not provide an avenue for further experimentation.

5. Empirical application

We finally proceed to an empirical application: We used frequentist maximum likelihood to estimate the parameters of both models for a number of selected empirical data sets including stock market indices, exchange rates and the price of gold. The stock market indices are: the S&P500, the German DAX and the Japanese NIKKEI, all represented by daily returns, all sampled from 01/01/1980 to 12/31/2005. The exchange rates are: U.S. Dollar against Euro (with data ranging from 01/01/1999 to 12/31/2010), Japanese Yen against U.S. Dollar (from 01/02/1986 to 12/31/2005) and the Swiss Franc against the Euro (07/15/2003 to 12/31/2010). Finally, the price of gold is also used in the form of daily returns over the same time horizon like the stock indices. Exactly the same time series have been used by Ghonghadze and Lux (2016) and Chen and Lux (2017) who have estimated the model of Alfarano et al. using GMM and SMM.

Table 6 shows the estimated parameter values and their standard errors and the maximized likelihoods. We note that these are positive as for both models the average conditional density exceeds unity. We note that the likelihood values of both models are very close with the Alfarano et al. (2008) model coming out with a slightly higher value throughout despite having one parameter less than its competitor.

Table 6
Empirical parameter estimates.

Parameter	ALW				FW					
	a	a	σ_f	logl	v	α_1	a	σ_f	logl	Vuong
S&P500	0.223 (0.001)	0.982 (0.001)	0.839 (0.002)	20134.087	1.281 (0.008)	2.571 (0.016)	0.030 (0.000)	1.216 (0.008)	19115.619	12.494 (0.000)
DAX	0.283 (0.000)	1.080 (0.005)	0.737 (0.003)	21386.671	0.829 (0.001)	1.234 (0.002)	5.246 (0.003)	1.008 (0.001)	20740.043	5.976 (0.000)
NIKKEI	0.282 (0.001)	0.760 (0.001)	0.700 (0.001)	20411.168	0.551 (0.003)	3.289 (0.006)	0.013 (0.000)	1.138 (0.009)	19510.904	13.480 (0.000)
USD/EUR	0.057 (0.000)	1.391 (0.004)	5.955 (0.004)	10552.723	0.678 (0.278)	1.713 (0.199)	1.988 (0.392)	0.633 (0.010)	10439.419	3.024 (0.001)
YEN/USD	0.113 (0.001)	0.304 (0.001)	0.600 (0.001)	17784.405	0.201 (0.306)	2.596 (1.784)	5.676 (4.429)	0.719 (0.005)	17422.916	4.754 (0.000)
CHF/EUR	0.149 (0.003)	1.729 (0.109)	0.265 (0.007)	7443.652	1.643 (0.026)	2.982 (0.047)	0.015 (0.000)	0.270 (0.002)	7412.537	0.283 (0.389)
Gold	0.386 (0.001)	0.855 (0.002)	0.666 (0.001)	21533.262	2.456 (0.003)	2.617 (0.005)	0.044 (0.000)	1.147 (0.006)	19960.242	18.033 (0.000)

Notes: The table shows the estimates of the parameters of the models of Alfarano et al. and Franke and Westerhoff, obtained by frequentist maximum likelihood together with the maximized log likelihood of both models. Standard errors of the parameters are shown in parentheses. Parameters a and b in ALW have been multiplied by 1000 and parameter σ_f has been multiplied by 100 for both models for better readability. The test statistics of the Vuong test is the normalized average log likelihood difference between both models. In parentheses the probabilities of rejection of the null hypothesis are shown that the first model (ALW) is at least as good as the second model (FW).

The last column of Table 6 shows the result of a Vuong Test (Vuong, 1989). This test allows to compare the likelihood of non-nested models. Its test statistic is a standardized likelihood difference which is exhibited in the table together with the probability of rejection of the null that the first model (ALW) dominates the second one (FW). As we can see, there is only one instance in which this statistics is not significant at all traditional levels of confidence. However, if we would use the adjusted Vuong test that adds a correction factor for different numbers of parameters, the result would be as strong as for the other cases. We note that because of the slight disadvantage of the Franke/ Westerhoff model mentioned above (our inability to initialize it with the stationary distribution of the latent variable) we have discarded the first 500 observations of each series as a burn-in phase of the particle filter in our specification test.

What insights can we infer from the results displayed in Table 6? First, in all cases, we find the parameters of both models in their respective bimodal regimes. For the Alfarano et al. approach this means that the herding part of the opinion formation process is stronger than its autonomous component, i.e. $b > a$. For the model of Franke and Westerhoff, we also always find strong herding which in this case is obtained by $\alpha_1 > 1$. However, by construction of the pertinent models, the bimodal case in Alfarano et al. leads to 'realistic' dynamics with clustered volatility and fat tails, while this is not necessarily so with $\alpha_1 > 1$ in the second model. Indeed, close inspection shows that for most of the time series, the filtering of the opinion process on the base of the second model yields relatively trivial results: Because of the strong herding, the opinion dynamics converges to a strong dominance of one strategy within the population. In all cases except for the German DAX (which has the smallest α_1) the filtered fraction of fundamentalists is practically equal to 1 for the entire time series. Hence, prices should follow the fundamental innovations closely which is reflected in a higher estimate of σ_f than for the model of Alfarano et al. The estimates for the standard deviation of the fundamental innovations are indeed very close to the standard deviation of the underlying returns series to that the dynamics of asset prices would be entirely explained by fundamental factors. In contrast, the estimate of σ_f is always smaller for the first model indicating that this one leaves a role for speculative price movements caused by its sentiment process.

Fig. 3 shows details of the only 'interesting' case, the German DAX. This is the only time we see some fluctuations of the fraction of fundamentalists and chartists in the model of Franke and Westerhoff: While most of the time, the index x_t of the herding process fluctuates around the positive mode of the bi-modal distribution ($x_t^* \approx 0.6$), there are frequent excursions downward that coincide with phases of strong volatility of the index (right-hand side panel). We also see that the opinion index of both models (indicating the relative fraction of optimists compared to pessimists in Alfarano et al.) move very much in tandem with each other. Hence, both models seem to pick up some element of sentiment-based causation of high volatility, while in calm phases they diagnose a strong dominance of either the fundamentalists or the optimistic noise traders. In this respect, it is worthwhile to remember that the Alfarano et al. model would only predict an influence of changes of sentiment, cf. Eq. (9). With a prevailing majority of either optimists or pessimists, price changes would also be dominated by fundamental innovations.

The upper right-hand panel shows, however, that mostly the Franke/Westerhoff model performs less successfully during volatile phases, i.e. its likelihood often decreases more sharply during an outburst of volatility which explains its inferior overall performance. This is much more pronounced in the other time series where the fraction of fundamentalists is stuck at about 1 and the model, thus, does not react in phases of excess volatility.

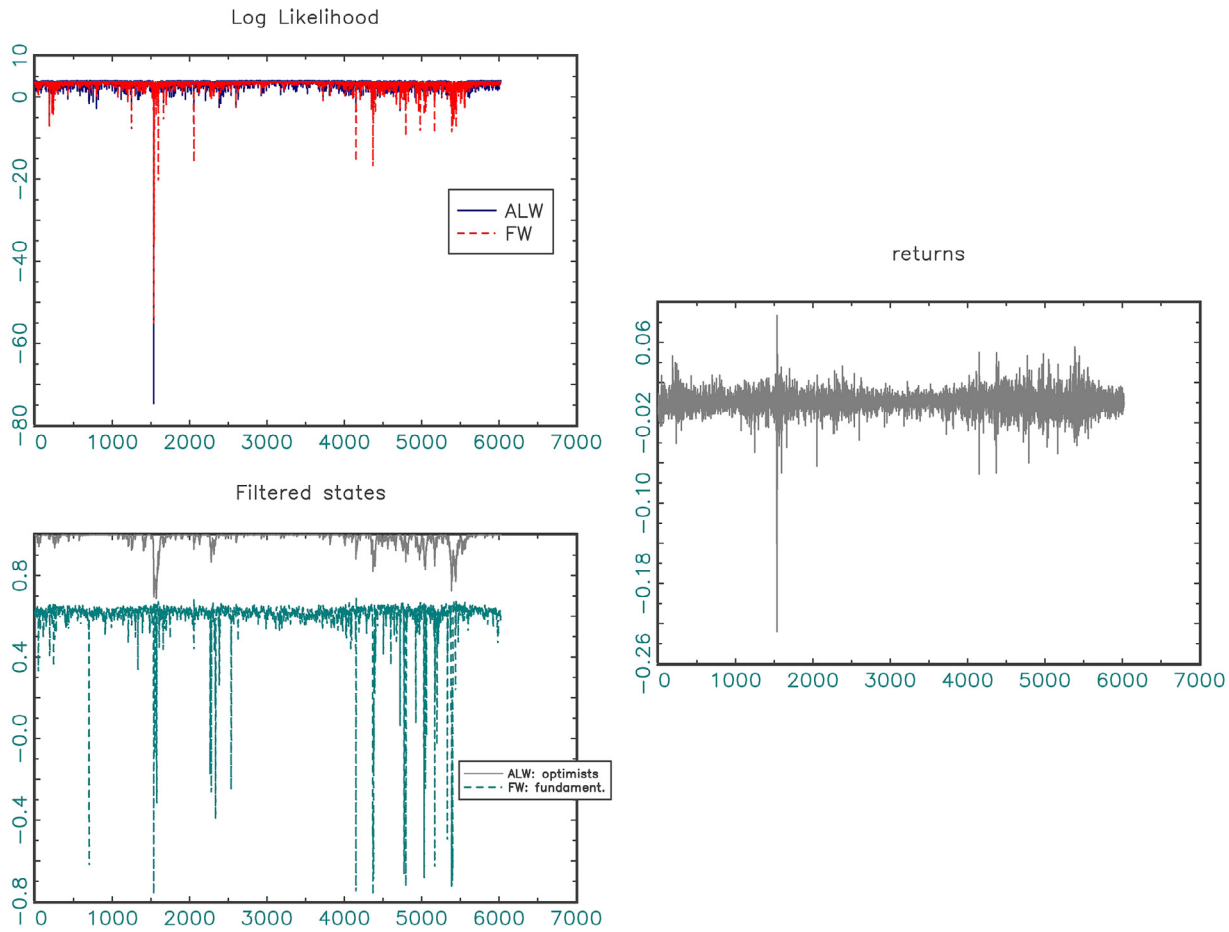

Fig. 3. Results of estimation of the Alfarano et al. (2008) and Franke/Westerhoff models using returns of the German stock index DAX. The upper left-hand panel shows the log likelihoods of both models while the lower left-hand panel exhibits the filtered trajectories of their sentiment variables. The right-hand side shows DAX returns over the time horizon 1980–2005.

Overall, we see from our estimation exercise how behavioral models can contribute to an understanding of different market phases. In the Alfarano et al. (2008) model, excess volatility is caused by *changes* in sentiment, and we clearly see how our filter based on this model tracks such hypothetical sentiment changes. The same can be observed for the model of Franke and Westerhoff for the case of the DAX. The less satisfactory overall performance of this model might simply be an indication of a too rigid formalization of trend-following behavior. Fluctuations caused by the excess demand function assumed for chartists might simply become so violent, that the filter is urged to 'avoid' a significant fraction of chartists at all times and, hence, we arrive at the not-so-interesting result of permanent fundamentalist dominance (which leaves deviations of the time series from homoskedastic, Normally distributed returns unexplained). Alternative specifications might lead to different results and could be explored with the methods introduced in this paper.

6. Conclusion

Agent-based models can be interpreted as special cases of state space models or more general models with hidden (latent) variables. Taking stock of the immense literature on statistical methods for such models, many extant approaches appear suitable and could be appropriately adapted for the estimation of agent-based models. On the positive side, using advanced filtering methods could provide parameter estimates of much higher precision compared to hitherto popular moment estimators, and this approach has the added advantage of providing the possibility of identification of the trajectories of the hidden variables that are of crucial conceptual importance in most agent-based models. On the downside, many of the methods reviewed here are very computation intense. Particle-based maximum likelihood and Bayesian estimation both come with the computational burden of a simulation approach that uses in each of its iterations the simulation of an agent-based model. Given that this raises the computation time to a multiple of what is needed in other applications, it seems worthwhile to particularly explore the various methods proposed for on-line estimation in subsequent research.

Acknowledgment

This article is dedicated to the memory of Carl Chiarella, who has been an inspiring colleague and a role model as a devoted scholar to me for more than 25 years. This article combines two of his major areas of interest: agent- based modeling and nonlinear filtering. If he would have had more time, he would probably have combined both strands of his research interests himself in due course. For many helpful comments, I am grateful to the two anonymous reviewers and participants at various presentations of the present material. I am also most grateful to Deakin University's Business School for its excellent hospitality during the time when this version of the paper was finalized.

References

Alfarano, S., Lux, T., Wagner, F., 2008. Time variation of higher moments in a financial market with heterogeneous agents: an analytical approach. J. Econ. Dyn. Control 32 (1), 101–136.
Amilon, H., 2008. Estimation of an adaptive stock market model with heterogeneous agents. J. Empirical Finance 15, 342–362.
Amisano, G., Tristani, O., 2010. Euro area inflation persistence in an estimated nonlinear dsge model. J. Econ. Dyn. Control 34, 1837–1858.
Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc.: Ser. B (Statistical Methodology) 72, 269–342.
Bao, J., Chiarella, C., Kang, B., 2017. Particle filters for Markov switching stochastic volatility models. In: Chen, S.-H., Kaboudan, M. (Eds.), Handbook on Computational Economics and Finance. Oxford University Press. forthcoming.
Barde, S., 2016. Direct calibration and comparison of agent-based models of herding in financial markets. J. Econ. Dyn. Control (73) 329–353.
Bickel, P.J., Ritov, Y., Rydén, T., 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. Ann. Stat. 26 (4), 1614–1635.
Billio, M., Casarin, R., 2010. Identifying business cycle turning points with sequential monte carlo methods: an online and real-time application to the Euro area. J. Forecast. 29, 145–167.
Blevins, J.R., 2016. Sequential Monte Carlo methods for estimating dynamic microeconomic models. J. Appl. Econom. 31, 773–804.
Boswijk, H., Hommes, C., Manzan, S., 2007. Behavioral heterogeneity in stock prices. J. Econ. Dyn. Control 31, 1938–1970.
Brock, W., Hommes, C., 1997. A rational route to randomness. Econometrica 65 (5), 1059–1095.
Brock, W., Hommes, C., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J. Econ. Dyn. Control 22, 1235–1274.
Brown, G.W., Cliff, M.T., 2004. Investor sentiment and the near-term stock market. J. Empirical Finance 11 (1), 1–27.
Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Chains. Berlin, Springer.
Carvalho, C.M., Johannes, M.S., Lopes, H.F., Polson, N.G., et al., 2010. Particle learning and smoothing. Stat. Sci. 25, 88–106.
Chang, K.-H., 2012. Stochastic Nelder–Mead simplex method–a new globally convergent direct search method for simulation optimization. Eur. J. Oper. Res. 220, 684–694.
Chen, Z., Lux, T., 2017. Estimation of sentiment effects in financial markets: A simulated method of moments approach. Comput. Econ. in press
Chiarella, C., 1992. The dynamics of speculative behavior. Ann. Oper. Res. 37, 101–123.
Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices. Rev. Financial Studies 23, 3141–3189.
Crisan, D., Doucet, A., 2002. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Signal Process. 50, 736–746.
Day, R.H., Huang, W., 1990. Bulls, bears and market sheep. J. Econ. Behav. Organiz. 14, 299–329.
Douc, R., Doukhan, P., Moulines, E., 2013. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. Stoch. Process. Appl. 123 (7), 2473–2920.
Douc, R., Rouett, F., Sim, T., 2015. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. Manuscript, Telecom Paris.
Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2007. Estimating macroeconomic models: a likelihood approach. Rev. Econ. Stud. 74, 1059–1087.
Franke, R., Westerhoff, F., 2011. Estimation of a structural stochastic volatility model of asset pricing. Comput. Econ. 38, 53–83.
Franke, R., Westerhoff, F., 2012. Structural stochastic volatility in asset pricing dynamics: estimation and model contest. J. Econ. Dyn. Control 8 (36), 1193–1211.
Franke, R., Westerhoff, F., 2016. Why a simple herding model may generate the stylized facts of daily returns: explanation and estimation. J. Econ. Interact. Coord. 11 (1), 1–34.
Gallant, A. R., Hong, H., Khwaja, A., 2016. A Bayesian approach to estimation of dynamic models with small and large number of heterogeneous players and latent serially correlated states. Manuscript, Penn State University.
Ghonghadze, J., Lux, T., 2016. Bringing an elementary agent-based model to the data: Estimation via GMM and an application to forecasting of asset price volatility. J. Empirical Finance 37, 1–19.
Golightly, A., Wilkinson, D., 2011. Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. Interface Focus 1, 807–820.
Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear / non-Gaussian Bayesian state estimation. IEE-Proceedings-F 140, 107–113.
Grazzini, J., Richiardi, M., 2015. Estimation of ergodic agent-based models by simulated minimum distance. J. Econ. Dyn. Control 51, 148–165.
Grazzini, J., Richiardi, M., Tsionas, M., 2017. Bayesian estimation of agent-based models. J. Econ. Dyn. Control 77, 20–47.
Grewal, M., Andrews, A., 2008. Kalman Filtering: Theory and Practice with MATLAB, 3rd ed. Hoboken NJ, John Wiley.
Hens, T., Schenk-Hoppé, K., 2009. Handbook of Financial Markets: Dynamics and Evolution. Amsterdam, North Holland.
Ionides, E.L., Bhadra, A., Atchadé, Y., King, A., et al., 2011. Iterated filtering. Ann. Stat. 39, 1776–1802.
Jang, T.-S., 2015. Identification of social interaction effects in financial data. Comput. Econ. 45, 207–238.
Kantas, N., Doucet, A., Singh, S., Maciejowski, J., Chopin, N., et al., 2015. On particle methods for parameter estimation in state-space models. Stat. Sci. 30, 328–351.
Kirman, A., 1993. Ants, rationality, and recruitment. Q. J. Econ. 79 (2), 171–175.
Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J. Comput. Gr. Stat. 5, 1–25.
Kitagawa, G., 1998. A self-organizing state-space model. J. Am. Stat. Assoc. 93, 1203–1215.
Kukačka, J., Barunik, J., 2017. Estimation of financial agent-based models with simulated maximum likelihood. J. Econ. Dyn. Control 85, 21–45.
Lamperti, F., Roventini, A., Sani, A., 2017. Agent-based model calibration using machine learning surrogates. Working Paper, Scuola Superiore Sant'Anna.
Leroux, B.G., 1992. Maximum-likelihood estimation for hidden Markov models. Stoch. Process. Appl. 20 (1), 127–143.
Liu, J.S., Chen, R., 1998. Sequential Monte Carlo methods for dynamic systems. J. Am. Stat. Assoc. 93, 1032–1044.
Lux, T., 1995. Herd behaviour, bubbles and crashes. Econ. J. 105 (431), 881–896.
Lux, T., Zwinkels, R. C. J., 2017. Empirical validation of agent-based models. Chapter prepared for Hommes, C. and LeBaron, B., eds., Handbook of Computational Economics. Elsevier, Amsterdam, forthcoming.
Malik, S., Pitt, M.K., 2011. Particle filters for continuous likelihood evaluation and maximisation. J. Econom. 165 (2), 190–209.
Molina, G., Bayarri, M., Berger, J., 2005. Statistical inverse analysis for a network microsimulator. Technometrics 47 (4), 388–398.
Olsson, J., Rydén, T., 2008. Asymptotic properties of particle filter-based maximum likelihood estimators for state space models. Stoch. Process. Appl. 118, 649–680.

Pitt, M.K., Malik, S., Doucet, A., 2014. Simulated likelihood inference for stochastic volatility models using continuous particle filtering. Ann. Inst. Stat. Math. 66, 527–552.
Recchioni, M.C., Tedeschi, G., Gallegati, M., 2015. A calibration procedure for analyzing stock price dynamics in an agent-based framework. J. Econ. Dyn. Control.
Snijders, T., 1996. Stochastic actor-oriented models for network change. J. Math. Sociol. 21, 149–172.
Tadić, V.B., Doucet, A., 2005. Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models. Stoch. Process. Appl. 115 (8), 1257–1436.
Vuong, Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica (57) 307–333.
Weidlich, W., Haag, G., 1983. Concepts and Models of a Quantitative Sociology: The Dynamics of Interacting Populations. Berlin, Springer.
Yu, J., 2005. On leverage in a stochastic volatility model. J. Econom. 127, 165–178.