Thomas Lux*

Approximate Bayesian inference for agent-based models in economics: a case study

https://doi.org/10.1515/snde-2021-0052
Received May 26, 2021; accepted June 8, 2022; published online June 27, 2022

Abstract: Estimation of agent-based models in economics and finance confronts researchers with a number of challenges. Typically, the complex structures of such models do not allow to derive closed-form likelihood functions so that either numerical approximations to the likelihood or moment-based estimators have to be used for parameter inference. However, all these approaches suffer from extremely high computational demands as they typically work with simulations (of the agent-based model) embedded in (Monte Carlo) simulations conducted for the purpose of parameter identification. One approach that is very generally applicable and that has the potential of alleviating the computational burden is Approximate Bayesian Computation (ABC). While popular in other areas of agent-based modelling, it seems not to have been used so far in economics and finance. This paper provides an introduction to this methodology and demonstrates its potential with the example of a well-studied model of speculative dynamics. As it turns out, ABC appears to make more efficient use of moment-based information than frequentist SMM (Simulated Method of Moments), and it can be used for sample sizes of an order far beyond the reach of numerical likelihood methods.

Keywords: agent-based models; Bayesian estimation; sequential Monte Carlo.

JEL Classification: G12; C15; C58.

1 Introduction

Approximate Bayesian Computation (ABC) has been developed for parameter estimation in cases for which no closed-form likelihood function is available, and even numerical approximations do not exist or are too costly to compute. This will often be the situation one faces when confronting complex agent-based models (ABMs) with data, so that ABC should be a welcome extension of the statistical toolbox for ABM researchers. Indeed, ABC enjoys quite a high popularity in applications in biology, ecology and epidemiology (e.g. Csilléry et al. 2010), but seems not to have been discovered yet by agent-based modellers in economics. While estimation and validation of ABMs has become an important area of research activity over the last decade (cf. Lux and Zwinkels 2018), most available contributions have used a traditional frequentist approach, and only a few have used Bayesian estimation methods (Grazzini, Richiardi, and Tsionas 2017; Lux 2022). The later approaches are, however, also based on evaluations of the likelihood and would, therefore, not be applicable if the likelihood could not be obtained at least by numerical approximations. Since availability of an analytical or approximate likelihood function cannot always be taken for granted, it should be welcome to have alternative methods available for more complex models.

In order to get an idea of the potential of ABC-based inference for ABMs in economics, the present paper will evaluate the performance of various algorithms developed in the ABC literature to a well-studied

***Corresponding author: Thomas Lux**, Department of Economics, University of Kiel, Olshausenstr. 40, Kiel 24118, Germany, E-mail: lux@economics.uni-kiel.de

asset-pricing model with interacting agents for which comparable results on the efficiency and computational demands of alternative estimation methods (method of simulated moments, simulated maximum likelihood) have been documented in previous papers.

To set the stage the remainder of the paper starts with a short introduction into the relevant methods we adopt from the flourishing literature on Approximate Bayesian Computation (Section 2). Section 3 reviews methods for post-processing of the immediate output of an ABC application that are routinely applied in the literature, and that typically provide a promising avenue for reducing the bias inherent in this approximate estimation approach. Section 4 presents the structure of the relatively simple ABM that we use as a test case of the application of various ABC algorithms. Section 4 also provides a detailed analysis of the comparative performance of different ABC specifications and algorithms for this model. Section 5 explains how to filter information on unobserved variables in a state-space framework using ABC, and Section 6 applies the previously introduced methods for estimation of the parameters of the exemplary financial ABM with data of the S&P 500. In this section, we also demonstrate how ABC could be applied for filtering information on hidden state variables, and we conduct a specification test based on the ABC estimates that appears surprisingly capable of identifying subtle discrepancies between certain features of the simulated data and their empirical counterparts. Section 7 concludes.

2 Principles of approximate Bayesian computation

Rubin (1984) has been the first to propose an ABC algorithm for applied Bayesian inference, while the first example of an empirical application of an ABC algorithm can be found in Tavaré et al. (1997). Both papers propose a likelihood-free rejection sampler to approximate the posterior distribution of the parameter vector θ of a model. Denoting by y an observed sample of data, by z samples simulated from a hypothesized data generating mechanism for y, i.e. $z \sim f(\cdot|\theta)$ where $f(\cdot)$ stands for the computational model that generates output samples z, and by $s(\cdot)$ some *summary* of features that is both applicable to data and simulated samples, the rejection sampler is best summarized by pseudo code, cf. Algorithm 1.

The rejection sampler repeatedly draws parameter values from the prior and accepts those draws with summary function $s(\cdot)$ sufficiently close to that of the empirical data. In the above algorithm, $\rho(\cdot,\cdot)$ is a distance function that aggregates the differences in the possibly multivariate summaries $s(\cdot)$ between observed and simulated data, and ϵ is a tolerance level up to which simulations are accepted as sufficiently similar to the data. In economic applications $s(\cdot)$ would typically consist of a set of appropriately chosen moments of the data such as unconditional moments or autocovariances as they have been used in frequentist applications of the Simulated Methods of Moments to univariate financial data (e.g. Chen and Lux 2018; Franke and Westerhoff 2012), or both auto- and crosscorrelations in the case of macroeconomic applications (Jang and Sacht 2016). The most common choice in the ABC literature is to take ρ to be the Euclidian norm. In Algorithm 1, all draws of θ for which the simulated summaries, $s(z)$ are within an ϵ neighbourhood of the observed summaries, $s(y)$, are accepted. However, one could also replace the binary acceptance step by a probabilistic

Algorithm 1: Rejection sampler

1: **for** $i = 1$ to N **do**
2: **repeat**
3: draw θ' from the prior $\pi(\theta)$
4: generate z from $f(\cdot|\theta')$
5: **until** $\rho(s(z), s(y)) \leq \epsilon$
6: set $\theta_i = \theta'$
7: **end for**

acceptance step with a kernel function that assigns a weight to the distance $\rho(\cdot,\cdot)$. Typical kernel functions are the uniform (which enters implicitly in Algorithm 1), triangular, Gaussian and others.

The accepted samples (θ, z) from Algorithm 1 are distributed according to:

$$\pi_\epsilon(\theta, z|s(y)) = \frac{\pi(\theta) f(z|\theta) I_{\rho(s(z),s(y))\leq\epsilon}(z)}{\int I_{\rho(s(z),s(y))\leq\epsilon} \pi(\theta) f(z|\theta) \mathrm{d}z \mathrm{d}\theta} \tag{1}$$

with $I_{\rho(s(z),s(y))\leq\epsilon}(z)$ the indicator function counting those simulations that fulfill the acceptance criterion. One would hope that the posterior for the parameters

$$\pi_\epsilon(\theta|s(y)) = \int \pi_\epsilon(\theta, z|s(y)) \mathrm{d}z \tag{2}$$

is a sufficiently close approximation to $\pi(\theta|s(y))$ which is intractable or too costly to simulate. The posterior $\pi(\theta|s(y))$ itself, serves as an approximation to the posterior $\pi(\theta|y)$. It is well known that the posterior conditional on moments $s(y)$ is equal to the posterior conditional on data y only if the moments are sufficient statistics. In general, sufficient statistics only exist for the exponential family of distributions. Hence, ABC replaces the exact full posterior density $\pi(\theta|y)$ by what one could call the exact partial posterior $\pi(\theta|s(y))$. Fearnhead (2019) summarizes recent results on ABC asymptotics by Frazier et al. (2018) and Li and Fearnhead (2018) that demonstrate posterior concentration at the true value in the limit of a sample size $n \to \infty$ of the empirical data and an ABC design that decreases the tolerance level with increasing sample size, $\epsilon_n \to 0$. In addition, the limit of the summaries has to satisfy an identifiability condition for the true parameter vector. In particular, as shown by Frazier et al. (2018) if the summary functions satisfy a central limit theorem with rate $\sqrt{n}$, the tolerance level ϵ_n has to be decreased as $\epsilon_n = o(1/\sqrt{n})$ for the ABC posterior to converge to the exact partial posterior. If not, i.e. $\epsilon_n\sqrt{n} \to c > 0$, it will not have a Gaussian limit and will not converge to the partial posterior.

Unfortunately, the simple ABC rejection sampler appears computationally very burdensome in many applications (as illustrated below). If the prior distribution is not very informative, many trials will be rejected so that most of the computation time would be wasted in so far as the simulated samples would not enter into the computation of the posterior. Hence, it would seem beneficial to find ways to move the proposals from an uninformative prior towards subsets of the parameter space with higher probability and, therefore, higher acceptance rates. Adaptations of traditional Monte Carlo methods, i.e. sequential Monte Carlo (SMC) and Markov Chain Monte Carlo (MCMC) to the ABC setting provide possible avenues for shifting the proposal distribution towards the posteriors.

Sequential Monte Carlo methods seem to have become the most often applied algorithms in the available literature. The first SMC algorithm within an ABC context has been proposed by Sisson, Fan, and Tanaka (2007), and improved by Beaumont et al. (2009). As usual, SMC propagates a set of realizations (particles) from the prior distribution through a sequence of intermediate steps to the sought posterior. In ABC-SMC the key role is assumed by a sequence of tolerance levels ϵ_t, $t = 1, \ldots, T$. The pseudo-code of Algorithm 2 describes the approach developed by Beaumont et al. (2009) and Toni et al. (2009) which has been used in many empirical applications.

Lines 11 to 20 in Algorithm 2 show how the intermediate distributions of particles are generated by weighted draws from the preceding pool of particles together with random variation in the form of a transition kernel $K_t(\theta^*|\theta')$. The relatively complex weights in line 19 are necessary in order to make sure that the weights are unbiased in the progression of intermediate distributions in the presence of weighted resampling of the particles used in the propagation step with kernel $K(\cdot)$. The first version of ABC-SMC by Sisson, Fan, and Tanaka (2007) using simpler weights has been shown to lead to biased outcomes (Beaumont et al. 2009; Toni et al. 2009).

The general approach to SMC-ABC discussed so far provides little overall guidance on the choice of the sequence of tolerance levels ϵ_t. If the decay factor is chosen in a too ambitious way, many proposals together with the generated samples of the underlying model would again be rejected and, therefore, computation time would probably not be used efficiently. It seems, therefore, sensible to determine the sequence ϵ_t in

Algorithm 2: ABC-SMC

1: **for** decreasing sequence $\epsilon_1, \ldots, \epsilon_T$ **do**
2: **for** $t = 1$ **do**
3: **for** $i = 1$ to N **do**
4: **repeat**
5: draw θ' from the prior $\pi(\theta)$
6: generate z from $f(\cdot|\theta')$
7: **until** $\rho(s(z), s(y)) \leq \epsilon_1$
8: set $\theta_i = \theta'$; assign weight $w_1^{(i)} = \frac{1}{N}$
9: **end for**
10: **end for**
11: **for** $t > 1$ **do**
12: **for** $i = 1$ to N **do**
13: **repeat**
14: draw θ' from previous distribution $\left\{\theta_{t-1}^{(i)}\right\}$ using weights $w_{t-1}^{(i)}$
15: generate $\theta^* \sim K_t(\theta^*|\theta')$ with $K(\cdot)$ a transition kernel
16: generate z^* from $f(\cdot|\theta^*)$
17: **until** $\rho(s(z^*), s(y)) \leq \epsilon_t$
18: set $\theta_t^{(i)} = \theta^*$
19: assign weight $w_t^{(i)} = \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^{N} w_{t-1}^{(j)} K_t(\theta_t^{(i)}|\theta_{t-1}^{(j)})}$
20: **end for**
21: **end for**
22: **end for**

a data-driven way. Del Moral, Doucet, and Jasra (2012) propose to determine ϵ_t relative to the α quantile of the distribution of observed distances $\rho(s(z^{(i)}), s(y))$ calculated from the particles $\left\{\theta_{t-1}^{(i)}\right\}$ at step $t-1$ of the algorithm. Together with adaptive adjustment of the propagation kernel, such a choice should lead to a relatively high acceptance rate as a subset of the particles generated at round $t-1$ did already satisfy this tolerance level. Del Moral, Doucet, and Jasra (2012) design an alternative ABC-SMC algorithm based upon multiple simulations $\left\{z_1^*, \ldots, z_M^*\right\}$ for each proposed parameter vector θ^* that exploits the so obtained insight into the variability of the simulations.

An alternative algorithm with data-driven choice of the threshold has been proposed by Drovandi and Pettitt (2011). These authors also propose to set the threshold ϵ_t as the α-quantile of the distribution of observed distances calculated at the previous round, $t-1$, of the algorithm. Sorting the accepted samples (θ, z) by the discrepancy, i.e. the distances $\rho(s(z), s(y))$, the $100(1-\alpha)\%$ fraction of largest distances is dropped and the remaining $100\alpha\%$ fraction that satisfies this threshold is carried over to the next round. Subsequently, $(1-\alpha)N$ new draws are generated via a Markov Chain Monte Carlo (MCMC) move from the preserved pool of the αN particles obeying the current threshold ϵ_t (with αN and $(1-\alpha)N$ rounded to integer values). In order to avoid accumulation of identical particles, Drovandi and Pettitt (2011) propose to perform R independent MCMC draws at each iteration with R adaptively determined by the formula

$$R_t = \frac{\ln(c)}{\ln(1 - P_{\text{acc},t-1})} \tag{3}$$

with c a tolerable probability that the particle remains unchanged (e.g. $c = 0.01$), and $P_{\text{acc},t-1}$ the acceptance frequency at the previous round, given the previous choice R_{t-1}. Both the preservation of part of the previous swarm and the adaptive choice of the sequence of thresholds should help to reduce the number of wasteful computations. Note that because of the MCMC move used in generating the new particles, the weight of all active particles always remains $w_t^{(i)} = \frac{1}{N}$ in all iterations, so that no reweighting as in line 19 of Algorithm 2 is required.

Another scheme, due to Lenormand, Jabot, and Deffuant (2013), also selects the threshold in the same data driven way as the previous one, and keeps a fraction αN of the previous set of particles. The remaining $(1-\alpha)N$ particles are generated with a kernel function based on the accepted subset, and the so obtained new particles are all accepted. As a modification of the ABC-SMC of Beaumont, Zhang, and Balding (2002), the new particles are assigned the weights

$$w_t^{(i)} = \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^{\alpha N}\left(w_{t-1}^{j}/\sum_{l=1}^{\alpha N} w_{t-1}^{(l)}\right) * K_t(\theta_t^{(i)}|\theta_{t-1}^{(j)})} \tag{4}$$

taking into account that their ancestry only stems from the αN maintained particles. The authors also propose a data-driven stopping rule, namely to not continue with the iterations if the number of new particles among the $(1-\alpha)N$ that satisfy the new threshold ϵ_t falls below a certain lower boundary. Obviously, dispensing with rejection reduces the number of wasteful simulations in each round, but slows down the progression of thresholds so that more rounds will be needed to reach a certain target. The stopping rule also guarantees that the algorithm will not try to generate samples with unrealistically low thresholds that fall below the sample variability of the distances with given parameters. Lenormand, Jabot, and Deffuant (2013) show that their algorithm is more efficient than the previous ones in certain applications.

3 Post-processing of ABC output: regression adjustment

The acceptance of sets of parameters with distance function obeying $\rho(s(z), s(y)) \leq \epsilon$ leads to a posterior distribution $\pi_\epsilon(\theta|s(y))$ which serves as an approximation to the sought posterior. While convergence of $\epsilon_n \to 0$ with increasing sample size n would guarantee concentration of the posterior $\pi_\epsilon(\theta|s(y))$ at the exact partial posterior $\pi(\theta|s(y))$, in empirical studies, the sample size is usually given and a certain deviation would, therefore, have to be accepted.

To correct for this discrepancy, various forms of regression adjustments have been proposed to extrapolate from finite ϵ_n to the limiting case $\epsilon_n = 0$ (cf. Beaumont, Zhang, and Balding 2002). Assuming a linear relationship between accepted draws and summaries, the adjustment of the posterior could be based on a sequence of regressions performed for each element of the set of parameters:

$$\theta_i = \alpha + \beta'(s_i - s(y)) + \eta_i \tag{5}$$

with θ_i the vector of parameters underlying one of the samples accepted to the posterior $\pi_\epsilon(\theta|s(y))$, $s_i - s(y)$ the vector of differences between the simulated summaries $s_i = s(z^{(i)})$ for the parameter vector θ_i and the measurements obtained with the empirical data, and η_i an error term. The regression of Eq. (5) is performed separately for each parameter (for the sake of better readability, we did not introduce notation to distinguish between different parameters). To concentrate the regression in the vicinity of $\epsilon = 0$, instead of a simple linear regression a weighted regression could be applied, with the weights being determined by the distance to $\epsilon = 0$, i.e. $\omega_i \propto K_\epsilon\{\rho(s(z_i), s(y))\}$ with K_ϵ a bounded kernel function.

After estimation of the regression coefficients, $\hat{\beta}$, $\hat{\alpha}$ and the residuals $\hat{\eta}_i = \theta_i - \hat{\beta}' s(z_i)$, the corrected sample is obtained as:

$$\hat{\theta}_i = \theta_i - \hat{\beta}'(s_i - s(y)) = \hat{\alpha} + \hat{\eta}_i. \tag{6}$$

Such corrections appear by now standard in applied research in ecology and evolutionary genetics. Extant literature also considers nonlinear adjustments in order to control for heteroskedasticity in the regression. A nonlinear heteroskedastic model proposed in Blum and François (2010) assumes a more flexible form:

$$\theta_i = \alpha + \beta'(s_i - s(y)) + \sigma(s_i)\eta_i. \tag{7}$$

Blum and François (2010) propose to estimate the factors $\sigma(s_i)$ by a second auxiliary regression:

$$\ln(\hat{\theta}_i - \hat{\alpha})^2 = \gamma_0 + \gamma_1'(s_i - s(y)) + \xi_i \tag{8}$$

Again, Eqs. (7) and (8) are to be understood as separate regressions conducted for each parameter line-by-line. In our subsequent applications we will make use of both the standard linear correction of Eq. (5) as well as the nonlinear version despicted in Eqs. (7) and (8). There also exist contribution in the literature using more flexible approaches such as neural networks or ridge regressions (cf. Blum et al. 2013).

4 An application: the Alfarano et al. asset pricing model

4.1 The model

We apply the above ABC estimation algorithms to the simple asset pricing model proposed by Alfarano, Lux, and Wagner (2008) in order to shed light on the performance of various proposed estimation schemes.

Using this particular agent-based model as a test bed provides several advantages: First, it is 'truely' agent-based in that the underlying data-generation process simulates the behavior of a set of distinct agents who all are characterized by their own stochastic process and as a group define the sentiment variable driving stock prices. Second, this model is known to be able to generate time series with 'stylized facts', i.e. fat tails and clustered volatility, thus resembling closely the behavior of empirical data. Third, since this model has already been used to test alternative estimators, we can compare the present results to those obtained with estimation by Generalized Method of Moments (Ghonghadze and Lux 2016), Simulated Method of Moments (Chen and Lux 2018), Simulated Maximum Likelihood using a particle filter approximation (Lux 2018) and Bayesian Markov Chain Monte Carlo (Lux 2022).

The following are the basic ingredients of this model: There is an ensemble of speculative market participants with size N_c each one of whom at any point in time is either bullish (+) or bearish (−), and these agents change their attitude in continuous time according to two transition rates under the influence of other traders whom they meet randomly:

$$\pi_{+,t} = a + bn_{+,t}; \pi_{-,t} = a + bn_{-,t}, \tag{9}$$

with $n_{+,t}$ ($n_{-,t}$) the number of bullish (bearish) investors at time t. This formalization allows for idiosyncratic switches via the parameter a and pair-wise interaction with intensity b. It is well known that this model generates an interesting bi-modal distribution of attitudes if $b > a$ holds, while it would yield a more innocuous uni-modal distribution for $a > b$.

As explained in detail in Appendix A, we consider a second group of fundamentalists together with the speculators following bullish and bearish sentiment as described above. We also assume that changes of fundamentals are white noise and restrict the aggregate trading volumes of both fundamentalists and speculators to equal unity. Under these assumptions, market returns, are obtained as:

$$r_t = u_{f,t} + x_t - x_{t-1} \tag{10}$$

with $u_{f,t} \sim N\left(0, \sigma_f^2\right)$ the fundamental innovations assumed to be Gaussian, and $x_t = \frac{n_{+,t} - n_{-,t}}{N_c}$ the current realization of the sentiment index. It is further assumed that the underlying unit time steps correspond to trading days and the returns are extracted as discrete observations of an underlying continuous-time process. Indeed, the agents' switches between attitudes will be simulated with an exact algorithm for discrete events taking place in continuous time. More details on the economic assumptions underlying Eqs. (9) and (10) are provided in Appendix A.

4.2 An example

We start by exploring the potential of ABC by simply simulating one time series from the model for which we attempt to recover the underlying parameters using the algorithms presented in Section 2. The parameters used to generate the sample in Figure 1 have been: $a = 0.0003$, $b = 0.0014$, $\sigma_f = 0.03$ and $N_C = 100$, the sample size is $n = 20,000$. These parameters are in the vicinity of what has been estimated with empirical data in previous papers, and as can be seen from the figure, the simulated series shares the salient properties of empirical returns from stock markets and other financial instruments.

The application of ABC requires a number of decisions. The most important one is which moments or measurements to use. For univariate financial returns, we can select various unconditional or conditional empirical moments, or moments based on the scores of auxiliary models as in the indirect inference literature. In the literature on estimation of agent-based models, a standard set of moments used in many empirical studies includes unconditional moments of returns up to the forth as well as autocorrelations of various powers of returns over a selection of lags. Other choices include coefficients of auxiliary ARCH models (Gilli and Winker 2001) or statistics related to value-at-risk (Tubbenhauer, Fieberg, and Poddig 2021). Since we want to compare the performance of ABC with previous results for GMM/SMM we follow Chen and Lux (2018) by selecting four different sets with increasing numbers of empirical moments: the smallest has 4 moments, and the others come with 7, 11 and 15 moments. Smaller sets are subsets of the larger ones, and we use the following complete list of moments:

$$\underbrace{r_t^2, r_t r_{t-1}, r_t^2 r_{t-1}^2, r_t^4}_{M4}, \underbrace{|r_t||r_{t-1}|, r_t^2 r_{t-5}^2, |r_t||r_{t-5}|}_{M7\setminus M4},$$
$$\underbrace{|r_{t+10}||r_{t-10}|, r_t^2 r_{t-10}^2, |r_t||r_{t-15}|, r_t^2 r_{t-15}^2}_{M11\setminus M7},$$
$$\underbrace{|r_t||r_{t-20}|, r_t^2 r_{t-20}^2, |r_t||r_{t-25}|, r_t^2 r_{t-25}^2}_{M15\setminus M11}.$$

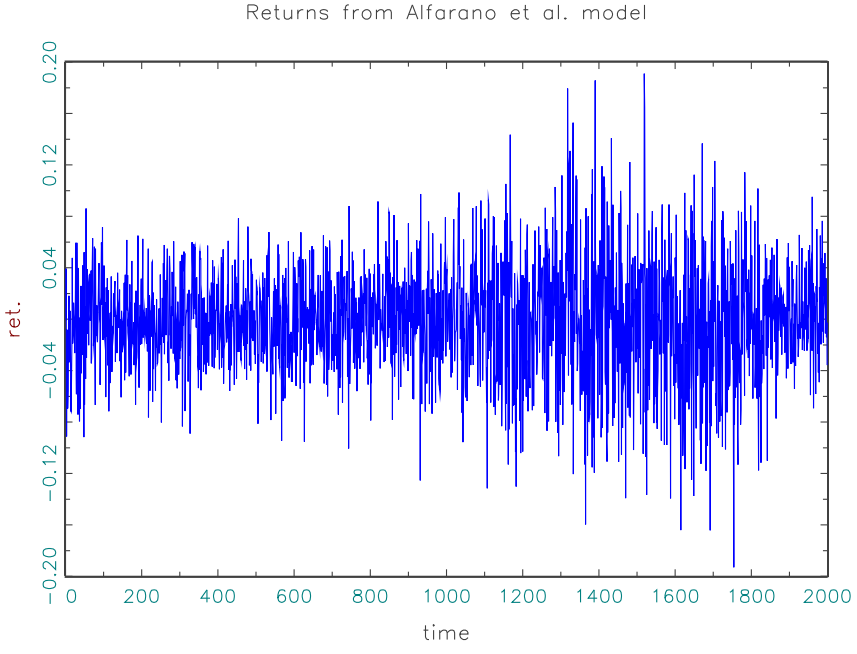

Figure 1: An example of simulated returns over unit time intervals from a continuous-time simulation of the model of Alfarano, Lux, and Wagner (2008).

These moments cover the scale of fluctuations $\left(r_t^2\right)$, the typical excess kurtosis of financial returns $\left(r_t^4\right)$, the absence of raw autocorrelation $(r_t r_{t-1})$ and the slow decay of their absolute and squared transformations.

In order to let the data speak on the preferred set of moments, we need a selection criterion. Blum et al. (2013) propose to use standard information criteria for this purpose as one possibility. We apply here the Akaike criterion, and we do not only apply it to the different sets of moments themselves, but also to the question of whether post-processing via linear or nonlinear regressions (as presented in Section 3) improves upon the 'raw' estimates. The Akaike criterion reads

$$\mathrm{AIC} = N \ln \prod_{j=1}^{q} \hat{\sigma}_j^2 + d \tag{11}$$

with $\hat{\sigma}_j^2$ the standard errors of the estimated parameters θ_j, $j = 1, \ldots, q$ (here: $q = 3$ with the estimated parameters being a, b and σ_f). d is a penalty term for the additional parameters needed for the regression adjustment with $d = q(M+1)$ for the linear regression and $d = 2q(M+1)$ for the regression with heteroscedastic variance, M denoting the number of moments used.

In a pilot run of four methods applied to the synthetic time series of Figure 1 we try to obtain some first insights into the efficiency of ABC estimates and their computational costs.

For all methods we set the same goal: We attempt to generate a set of $N = 1000$ particles for which the simulated summaries $s(z^{(i)})$ are all within a prespecified distance from their pseudo empirical counterparts. As summaries $s(z^{(i)})$ and $s(y)$ we choose the weighted Euclidean norm of the M distances taking some inspiration from the approach of Simulated Methods of Moments. Namely, we define:

$$\rho(s(z^{(i)}), s(y)) = \|X - Y\|^2 A_n = (X - Y)' A_n (X - Y) \tag{12}$$

with $X = s(z^{(i)})$, and $Y = s(y)$ the vectors of summaries of the simulated and pseudo-empirical samples, and $X - Y$ denoting the vector of deviations between simulated moments of the ith simulation and the moments of the pseudo-empirical series. Since there is no need in our case to restrict the simulations $z^{(i)}$ to the length of the targeted time series, we have chosen a length $4n = 80,000$ for each $z^{(i)}$. To capture the (co-) variation of the various moments, the weight function A_n is chosen as the inverse of the covariance matrix of the moments of the pseudo-empirical series y. As we will see, the realisations of $\rho(\cdot, \cdot)$ vary over several orders of magnitude in our initial sample of 1000 realisations from the prior of $\pi(\theta)$. To reflect a high degree of uncertainty, we assume a uniform prior with $a \in [0, 0.05]$, $b \in [0, 0.05]$ and $\sigma_\epsilon \in [0, 5 * stdc(ret)]$. This reflects our knowledge that all parameters should be positive and sets upper limits, mostly to prevent very computation-intensive simulations (the higher a and b, the more discrete changes of agents happen during a unit time interval). $stdc(ret)$ is the standard deviation of returns in the pseudo-empirical sample. Note that the 'true' parameters are all much smaller than the upper boundaries of their priors which allows to conveniently asses the influence of the information from the sample on the posterior.

In Table 1 we report parameter estimates obtained from four different ABC algorithms. For all cases we have computed an initial sample of 1000 simulations with parameters drawn randomly from their priors. The task for the algorithms was then to produce another set of 1000 simulations so that their differences $\rho(\cdot, \cdot)$ were all smaller than the second smallest difference of summaries in the initial sample. This choice is motivated by the following consideration: On the one hand, such a low distance is feasible as demonstrated by the set of initial particles. On the other hand, it is likely not an extreme outlier as the sample contained another entry with an even smaller distance. The four methods were the following:

- Method 1 is the baseline rejection sampling algorithm depicted as Algorithm 1 in the pseudo code above: This simply works as repeated sampling from the prior without any further guidance until the target sample of 1000 trials all with their $\rho(\cdot, \cdot) \leq \epsilon_{\min}$ is complete.
- Method 2 is the ABC-SMC approach depicted as Algorithm 2 above. As recommended by Beaumont et al. (2009), we have used as transition kernel a random walk kernel with innovations drawn from a trivariate Gaussian with covariance matrix equal to two times the covariance matrix of the accepted particles. This method requires a predefined sequence for the threshold values ϵ_t. We set $T = 10$ and

Table 1: ABC estimation of ALW model.

Method 1				**Method 3**			
	Estimates	**Linear corr.**	**Heterosc. corr**		**Estimates**	**Linear corr.**	**Heterosc. corr**
a	0.565 (0.365)	**0.343 (0.131)**	0.343 (0.088)	a	0.626 (0.309)	**0.358 (0.098)**	0.358 (0.129)
b	1.539 (0.681)	**1.834 (0.519)**	1.888 (0.838)	b	1.342 (0.222)	**1.410 (0.145)**	1.428 (0.260)
σ_f	23.662 (9.978)	**30.639 (2.585)**	30.644 (2.749)	σ_f	21.843 (8.143)	**29.476 (2.375)**	29.477 (1.472)
AIC	1816.728	−3385.088	−3010.623	AIC	−1166.705	−6669.051	−5823.042
tsec	25,861			tsec	586		
No. of evaluations	875,787			No. of evaluations	10,999		
Method 2				**Method 4**			
	Estimates	Linear corr.	Heterosc. corr		Estimates	Linear corr.	Heterosc. corr
a	0.533 (0.344)	**0.351 (0.106)**	0.337 (0.215)	a	0.636 (0.352)	**0.337 (0.100)**	0.337 (0.089)
b	1.463 (0.421)	**1.587 (0.376)**	1.569 (0.339)	b	1.285 (0.178)	**1.421 (0.149)**	1.450 (0.295)
σ_f	24.173 (9.587)	**30.125 (2648)**	30.779 (5351)	σ_f	21.271 (9.007)	**27.824 (2.834)**	28.056 (4.703)
AIC	656.072	−4411.038	−1689.667	AIC	−1141.954	−6222.072	−3991.966
tsec	2060			tsec	839		
No. of evaluations	83,405			No. of evaluations	19,000		

Example of ABC estimation of the parameters of the model of Alfarano et al., using four different ABC algorithms with distance functions composed of all 15 moments, with and without correction of the initial estimates via a standard weighted regression as well as via a weighted nonlinear regression allowing for heteroscedasticity. The sample size has been $n = 20{,}000$. The table shows the means and standard deviations of the posteriors of the parameters together with the value of the AIC criterion for each estimation. Bold numbers indicate the estimates preferred by the AIC. The estimates of the parameters are all multiplied by 10^4 for better readability. The table also provides information on the computation time in seconds (tsec), and the number of evaluations needed to arrive at the final set of accepted particles.

use as steps between successive rounds 10 equidistant differences in logarithms between the maximum distance of the first sample and the targeted final minimum $\epsilon_{\min}$.

- Method 3 is the algorithm proposed by Drovandi and Pettitt (2011) outlined also in Section 2. We choose $\alpha = 0.5$ and again use random walk proposals with Gaussian innovations with mean zero and covariance matrix equal to two times the sample covariance matrix of the αN maintained particles. In contrast to methods 2 and 4, the new proposals are generated, however, in the form of MCMC moves.
- Method 4 is the method of Lenormand, Jabot, and Deffuant (2013). Since this algorithm does not reject new simulations, we start here with a larger population $N = 1,500$ and set $\alpha = 2/3$. Hence, the algorithm will also stop its iterations once 1000 particles satisfy $\rho(\cdot,\cdot) \leq \epsilon_{\min}$. Again, random walk proposals are used between SMC iterations as with Method 2 and 3.

The target $\epsilon_{\min}$ has been the same for all methods for a given set of moments, namely the 0.2 percent quantile of the distances of the original sample of 1000 draws from the prior (i.e. the second smallest distance). In method 1, this target is imposed immediately, while in the other methods it is approached over a sequence of SMC steps. In method 2, we start with ϵ_1 set equal to the median of the initial sample, and then move down to $\epsilon_{\min}$ in 10 rounds with equidistant reductions (in logs) of ϵ_t. In method 3 and 4, the progression of ϵ_t occurs in the data-driven way described in Section 2 which comes to a halt if it reaches $\epsilon_{\min}$. For all methods we show in Table 1 the mean and standard errors of the original posteriors as well as those obtained after correction by linear and nonlinear heteroscedastic regressions. We also show the computation times needed and the number of evaluations of the model during the estimation. The reported AIC criterion allows us to identify which of the many alternative results would have to be preferred under this information criterion. The table reports results based on $M = 15$ moments which in this experiment always were preferred under the AIC criterion over those with smaller number of moments. Results for $M = 4$ to $M = 11$ can be found in Appendix B. Overall, we can infer the following insights from this exercise:

i. The minimum number of 4 moments seems to generate relatively unsatisfactory estimates (cf., Appendix B): The means of the posteriors are often quite different from the 'true' parameters, and the AIC criterion never selects an estimate with $M = 4$,
ii. Starting with $M = 7$, results become always more promising. In this particular case, all methods prefer $M = 15$ together with linear regression adjustment. However, results obtained with any number of moments between 7 and 15 are not very different in all cases, and indeed we obtain quite some variation between the preferred set of moments when running additional Monte Carlo simulations.
iii. Adjustment of the original swarm of particles through linear regressions typically improves the AIC values dramatically and shifts the posterior distribution evidently towards the 'true' values of the parameters. The heteroscedastic correction, in contrast, does not add value on top of linear extrapolation.
iv. Computational demands in terms of both computing time and number of evaluations of the model differ dramatically for the four algorithms: As expected, Method 1 is the least efficient. Method 3 is the most efficient one followed relatively closely by Method 4. Method 2 comes third with significantly higher computational demand.

Inspecting, for instance, the results for $M = 15$, we find that Method 1 needs 44 times as long as Method 3 to arrive at the final sample of 1000 particles with moment distance smaller than $\epsilon_{\min}$. Against Method 2, the advantage of Method 3 is still 1:4 while Method 4 only requires 0.5 times more in computation time. The comparisons are even better against Method 2 in terms of number of evaluations. This is unexpected as Method 3 also does not have any reweighting of the particles to do and so would have an advantage even for the same number of evaluations. It is also noteworthy that for some methods, the computational effort is non-monotonic in the number of moments: For Methods 1 and 2 the effort is lower for $M = 7$ against $M = 4$.

Figures 2 to 4 provide additional insights into the operations of the algorithms. Figure 2 shows the progression of the particle distribution for Method 2 with $M = 15$. The SMC algorithm iterates over 10 rounds until the stopping criterion is met. The development of means and 95 percent intervals shows that initially there is more pronounced adjustment in the parameters b and σ_f while a shows sizable variations only from

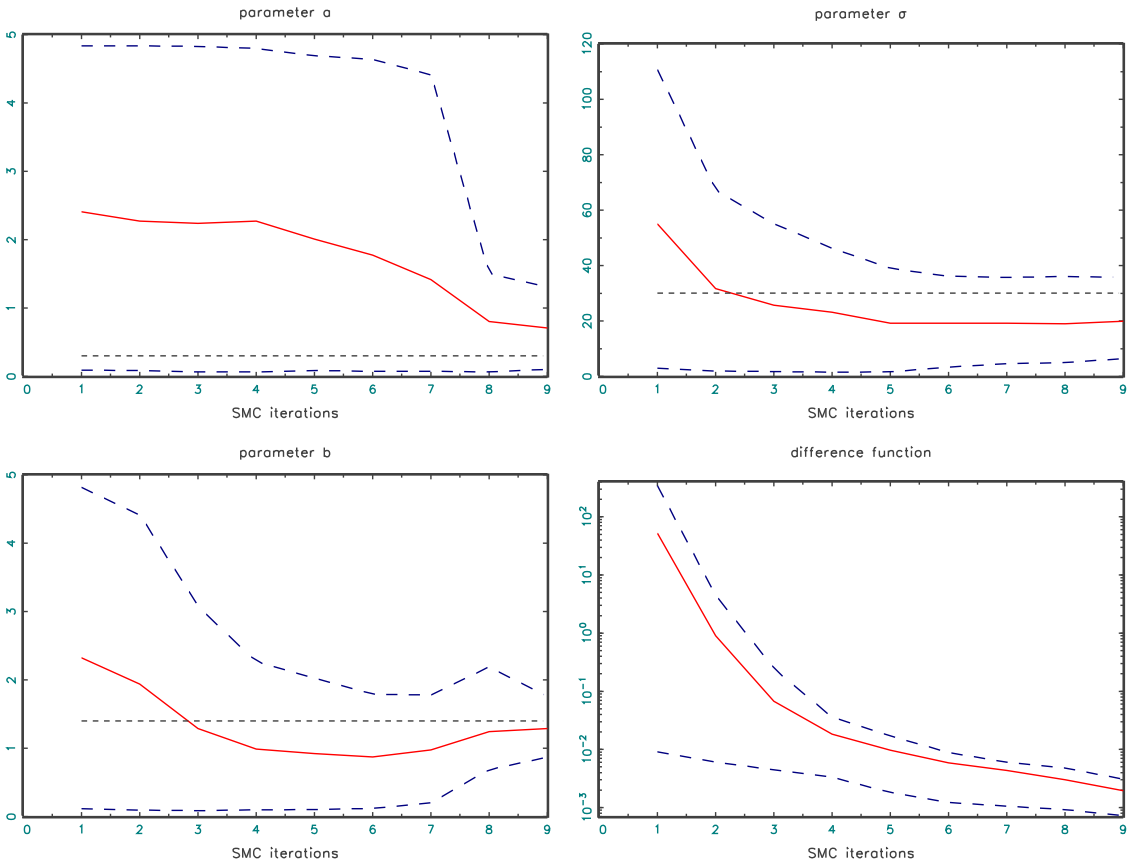

Figure 2: Evolution of the mean and inner 95 percent quantile of the particles of the parameters a, b and σ_f and of the distance function $\rho(\cdot,\cdot)$ in one exemplary SMC run using Method 2. The parameter values have all been multiplied by 10^4 for better visibility. The dashed straight lines indicate the 'true' parameter values.

about round 7, a reaction pattern that we also found in the other experiments. Since in this case, Method 3 needs the same number of iterations like Method 2, we can easily compare their computational demands per iteration. Figure 3 shows that Method 2 requires more evaluations and computation time in every round, but the needed computational resources increase particularly strongly towards the end of the simulation when it becomes increasingly difficult to find parameter sets satisfying the small target threshold ϵ_t. In contrast, Method 3 shows no significant change in computation times across iterations which demonstrates the advantage of a flexible adaptation of the sequence ϵ_t versus some predefined set of thresholds. The disadvantage of Method 4 vis-à-vis Method 2 is explained by the much longer number of iterations. In the present case with $M = 15$, Method 4 actually needs 35 rounds for convergence. Figure 4 illustrates a certain difference between the posteriors generated by Method 1 against the SMC algorithms: By drawing parameters always from the prior, it sometimes finds realisations far away from the 'true' parameter values that match the distance criterion. As a consequence, its distributions are typically broader and show a particularly heavy tail that in the present example assigns non-negligible probability over the entire support of the prior.

Since one simple estimation exercise can only serve to illustrate tendencies, we have also conducted a small Monte Carlo exercise repeating this estimation 50 times with different realisations of the underlying pseudo-empirical series (same parameters but different random number seeds) for Methods 2 to 4. Here we choose the intermediate number of moments, $M = 11$. Table 2 exhibits the results of this set of Monte Carlo experiments. The means of the posteriors are averaged over the 50 runs and the means (over these 50 runs) of their 95 percent credible intervals are given in parentheses. Also shown are the coverage ratios of these

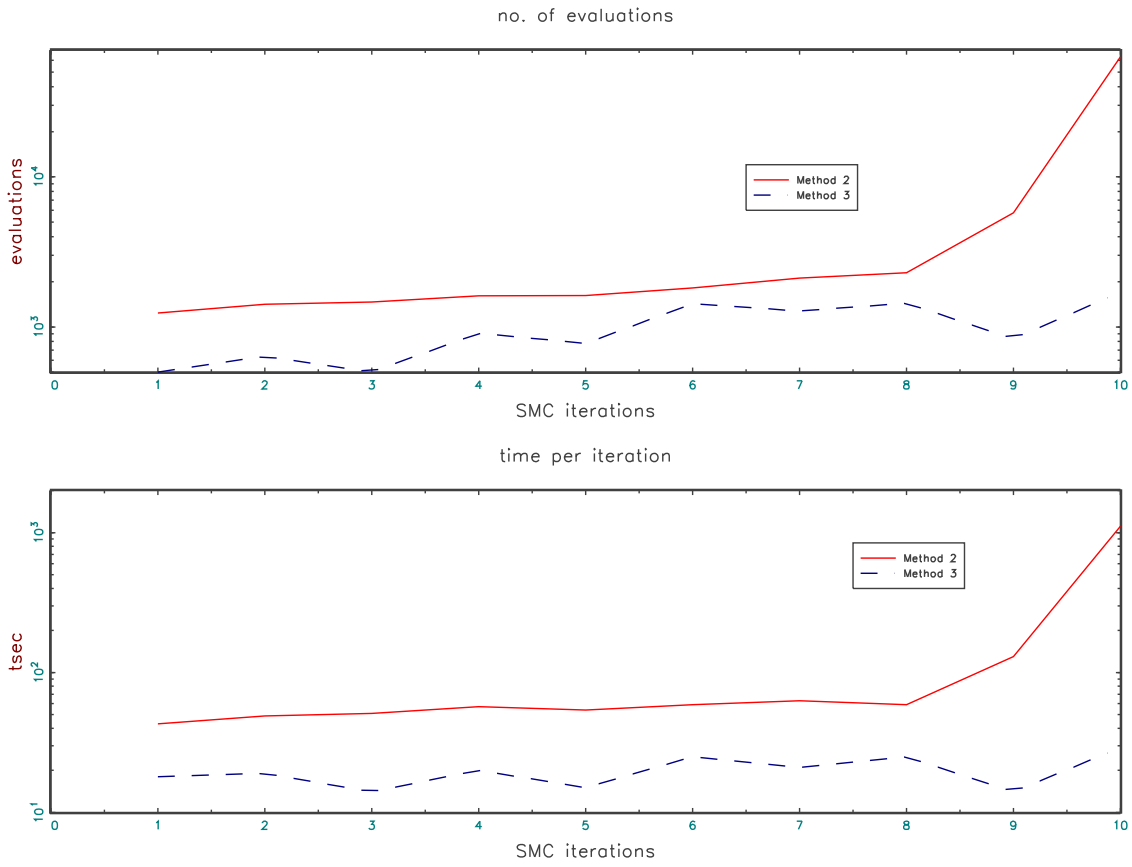

Figure 3: Comparison of the number of evaluations of the model and the time per round of SMC iteration for Methods 2 and 3.

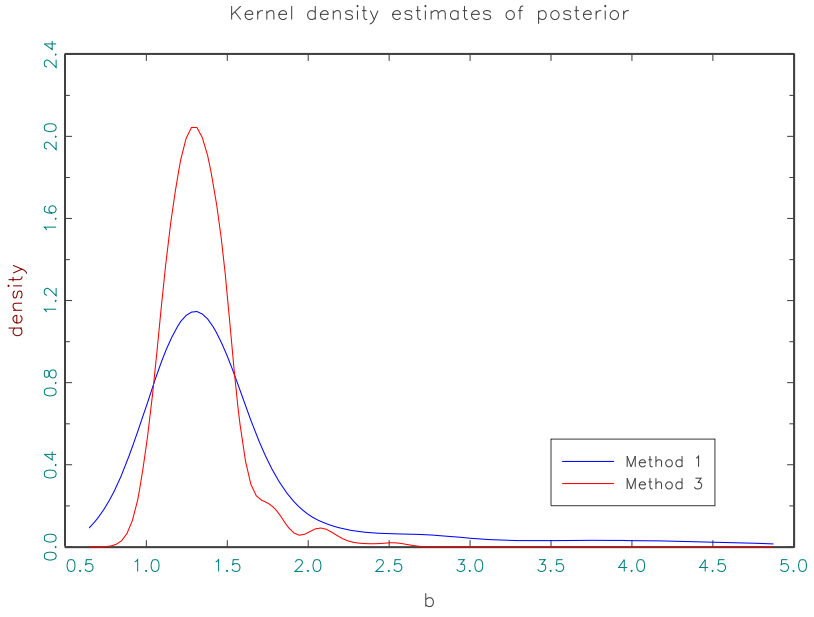

Figure 4: Comparison of the posterior density of parameter *b* obtained with Methods 1 and 3. The more diffuse posterior and the long right tail are characteristic for the results obtained with Method 1, while differences between the posteriors of the other methods are minor.

95 percent credible intervals. Again, we show the original parameter estimates and the ones obtained after linear and nonlinear corrections. Results confirm all the insights from the previous experiments. Indeed, we see that Method 3 is both more efficient in computational demands, and also provides more precise parameter estimates, at least after linear adjustment. Regression adjustment appears very valuable in all cases, but

Table 2: Monte Carlo results for ABC estimation of ALW model.

	Method 2		Method 3		Method 4	
	Mean	Cov.	Mean	Cov.	Mean	Cov.
a	0.632	1.00	0.690	1.00	0.646	0.92
	(0.062, 1.470)		(0.107, 1.331)		(0.110, 1.326)	
a_R	0.363	0.98	0.381	0.98	0.354	0.92
	(0.040, 0.725)		(0.118, 0.690)		(0.077, 1.072)	
a_{HR}	0.366	1.00	0.383	0.90	0.353	0.92
	(0.062, 0.705)		(0.137, 0.672)		(0.121, 0.982)	
b	1.477	1.00	1.336	0.98	1.467	0.98
	(0.896, 2.659)		(0.956, 2.038)		(0.947, 2.489)	
b_R	1.624	0.92	1.462	0.80	1.607	0.82
	(1.245, 2.549)		(1.218, 1.929)		(1.292, 3.167)	
b_{HR}	1.584	0.92	1.478	0.92	1.639	0.80
	(1.217, 2.254)		(1.189, 2.059)		(1.259, 3.334)	
σ_f	23.023	1.00	20.819	1.00	22.284	0.92
	(6.245, 38.395)		(7.009, 36.344)		(7.643, 36.805)	
σ_R	30.053	0.88	28.794	0.88	28.887	0.88
	(23.479, 35.172)		(23.274, 34.386)		(23.839, 37.088)	
σ_{HR}	30.072	0.86	28.797	0.92	28.935	0.86
	(23.035, 35.521)		(23.122, 34.601)		(24.017, 36.722)	
tsec	1072.680 (608.971)		176.820 (65.464)		730.940 (49.456)	
Evaluations	76,474.080 (52868.571)		11,866.040 (2424.730)		19,521.000 (3087.103)	

The table shows the posterior means and the means of the 95 percent credible intervals (in brackets) averaged over 50 Monte Carlo runs of parameter estimation via ABC algorithms for the model of Alfarano et al. as well as the coverage ratio (the relative number of cases in which the 'true' value resides inside the 95 percent credible interval). The sample size has been $n = 20{,}000$ in every run, and the number of moments used in all estimations was $M = 11$. The indices R and HR denote estimates obtained after post-processing using a linear regression and heteroscedastic regression, respectively. The last two lines give the means and standard deviations (in brackets) of the computation time and number of evaluations across the 50 Monte Carlo runs.

at least in this application, heteroscedastic corrections do not add any visible value on top of the linear adjustment. We also find that, at least for Methods 2 and 3, the credible intervals of the original posterior are very wide with coverage ratios of 1 indicating a posterior that is too broad. Regression adjustment leads to more narrow credible intervals but its effect on the coverage ratio is generally ambiguous. Overall, the computational demands of Method 3 appear unexpectedly modest: On average one run of this algorithm with about 12,000 evaluations all with a length of 80,000 time periods takes only about 3 min. Since the main body of the algorithm could be parallelized easily, we could therefore arrive at a computation time of less than 30 s with only 10 parallel threads.

We can compare the performance of the SMC-ABC approaches in Table 2 to similar results obtained for a Simulated Method of Moments (SMM) estimator, with identical sample size and set of moments, reported in Chen and Lux (2018, Table 4). While we are using a Bayesian approach here, the uninformative prior using a uniform distribution over a wide interval for all parameters brings the present estimation exercise close in spirit to frequentist SMM (but note that the present approach still obtains an estimate of the posterior distribution of the parameters, not only a point estimate). Comparing the mean values across the 50 replications in Table 2 with those reported in Chen and Lux, in particular Method 3 performs better with mean estimates closer to the 'true' values. Inspection of the standard deviations of these estimates (not shown in Table 2) indicates that those of the ABC posterior are smaller than the standard errors of the GMM estimates for two out of three parameters. Since ABC is approximate by nature and accepts a certain bias from the outset, one might even entertain the hope of improving on the results displayed in Table 2 by imposing a smaller threshold for the

distances and employing more computational resources to generate a sample satisfying this constraint, while in SMM there is no obvious way to improve upon the outcome of the initial estimations. This squares well with the results of Frazier et al. (2018) who show that the posterior mean obtained from ABC is asymptotically equivalent to the point estimates obtained via GMM with optimal weight matrix when identical moments are used.

It is worthwhile to also note that for the present sample size of $n = 20,000$ approximations for the likelihood by a particle filter (used in a frequentist and Bayesian context in Lux 2018, 2022) are completely out of reach due to their much higher computational demands.

4.3 Alternative distance functions

The distance used so far in our computational experiments is not a typical choice in extant ABC applications. More often, a standardized Euclidian distance is used, i.e. weights are reduced to their diagonal entries discarding the covariance terms. Replicating the Monte Carlo analysis with this alternative choice (first columns of Table 3, where both Methods 3 and 4 are applied with this alternative weight function), we find overall an inferior performance. Particularly poor are the results for parameters a and σ_f without post-processing. After post-processing, the means of the posterior distributions are closer to their 'true' values, but the dispersion is much larger than under the distance with non-diagonal weight function and we also find undercoverage of the 95 percent credible intervals for at least one of the parameters.

Another alternative proposed by Prangle (2017) uses a Mahalanobis or a standardized Euclidian distance with weights given by the inverse of the covariance matrix of the set of single Euclidian distances between sample moments and empirical moments obtained in the first round of the SMC algorithm (in the spirit of frequentist GMM, cf. Hansen 1982). We have run two versions of this approach: One again with only the diagonal weights used, and one with the full weight matrix, i.e. the Mahalanobis distance of the difference between simulated and pseudo-empirical moments. Results again fall back behind our benchmark of Table 2. The posteriors show a particularly strong bias for parameter a, while for the other parameters the dispersion is also mostly larger than for the choice of weight matrix used in Table 2. Not surprising, results from the diagonal weight matrix are worse again than those from the Mahalanobis version. There also appears to be some interaction between the chosen distance function and the estimation method. In particular, Method 4 generates outcomes with severe undercoverage for the alternative distance measures.

A different approach leading to a very different choice of distance function has been proposed by Fearnhead and Prangle (2012): Their objective is to determine optimal summaries under a quadratic loss function, and they show that the posterior mean is an optimal statistic under their criterion. Exploiting these insights, they propose to target a regression for the inference from the moments to the parameters as it is usually applied to post-process ABC output.

Hence, the distance function would consist of the distances between the parameters and their predicted values on the base of a regression using the moments as independent variables. We implement this approach using Method 3 which so far has been found to be the most parsimonious and efficient one.

To determine the regression coefficients, first a large initial set of $10N$ ABC particles is simulated. The weight function consists of a Mahalanobis type distance of the three differences between parameters and their predictions on the base of the regression, i.e. the inverse of the variance-covariance function of these differences (the prediction errors) in the initial large sample. With the so-defined distance functions, SMC with Method 3 is invoked with N particles until convergence is obtained. The convergence target is set as a maximum distance for all accepted particles below the 0.2 percent quantile of the distances of the initial sample. 11 moments again are used in the regression, in three different implementations: only the raw moments, and with their squares and powers of three added as independent variables.

Again, this approach does not appear to be competitive against our benchmark. The proximity of the posterior means to the true parameters is very mixed and the 95 percent credible intervals are very narrow often excluding the 'true' parameter values. The coverage ratios are very small showing that the credible

Table 3: Monte Carlo results for ABC estimation of ALW model, alternative weights used in distance function.

Weights: Method 3	Diag. mean	Cov.	GMM mean	Cov.	GMM-Diag mean	Cov.	Diag. mean Method 4	Cov.	GMM mean	Cov.	GMM-Diag. mean	Cov.
a	1.175	0.98	2.127	0.64	2.118	0.62	1.266	0.82	2.003	0.38	1.777	0.16
	(0.165, 2.731)		(0.265, 4.522)		(0.275, 4.533)		(0.191, 3.003)		(0.888, 3.272)		(0.775, 3.192)	
a_R	0.361	0.96	0.717	0.98	0.636	0.98	0.330	0.80	0.869	0.64	0.954	0.46
	(−0.141, 1.050)		(−0.982, 2.768)		(−1.010, 2.630)		(−0.343, 1.691)		(−0.248, 3.239)		(0.230, 4.416)	
a_{HR}	0.368	0.82	0.732	0.94	0.656	0.90	0.357	0.88	0.751	0.62	1.021	0.46
	(−0.085, 0.960)		(−1.791, 3.598)		(−1.797, 3.465)		(−0.264, 2.222)		(−0.456, 5.792)		(0.143, 14.968)	
b	1.272	0.98	1.038	0.98	1.070	1.00	1.241	0.94	1.067	0.54	1.346	0.52
	(0.902, 1.969)		(0.221, 2.310)		(0.228, 2.388)		(0.833, 1.732)		(0.537, 1.609)		(0.704, 2.005)	
b_R	1.463	0.74	1.326	0.86	1.347	0.82	1.614	0.80	1.291	0.58	1.148	0.68
	(1.251, 1.765)		(0.787, 1.835)		(0.846, 1.817)		(1.256, 2.465)		(0.980, 2.541)		(1.116, 3.552)	
b_{HR}	1.492	0.86	1.325	0.90	1.343	0.88	1.597	0.80	1.351	0.64	1.129	0.70
	(0.989, 2.216)		(0.346, 2.316)		(0.361, 2.336)		(1.184, 2.364)		(0.741, 2.976)		(1.023, 3.664)	
σ	15.943	0.98	34.371	1.00	31.967	1.00	16.421	0.94	34.554	0.60	20.425	0.60
	(1.061, 34.945)		(20.790, 48.252)		(17.607, 46.324)		(1.703, 34.948)		(25.923, 43.653)		(9.729, 32.384)	
σ_R	28.896	0.94	28.015	0.96	28.418	0.92	31.866	0.92	27.979	0.80	24.612	0.84
	(22.020, 36.094)		(19.371, 36.568)		(19.801, 36.856)		(21.749, 37.462)		(23.407, 38.130)		(22.953, 48.799)	
σ_{HR}	28.915	0.76	28.033	0.80	28.445	0.78	31.683	0.90	27.911	0.72	24.411	0.80
	(21.149, 37.310)		(19.550, 36.437)		(20.041, 36.684)		(21.049, 40.587)		(24.295, 36.591)		(22.655, 97.892)	
tsec	246.340 (47.755)		455.980 (199.750)		464.740 (202.833)		946.520 (119.805)		1672.640 (542.408)		1273.940 (344.699)	
Evaluations	16,908.600		29,328.160		29,832.300		26,161.000		57,561.000		46,651.000	
	(3796.377)		(12,950.788)		(13,344.152)		(6375.527)		(28,441.067)		(16,447.892)	

The table shows the posterior mean estimates, means of the 95 percent credible intervals (in brackets) and coverage ratios of 50 Monte Carlo runs of parameter estimation via ABC algorithms for the model of Alfarano et al. The sample size has been $n = 20{,}000$ in every run, and the number of moments used in all estimations was $M = 11$. The distance functions in this set of experiments use the following weights (1) the inverse of the variances of the moments of the data (i.e., the diagonal elements of variance–covariance matrix used in the previous set of experiments), denoted as Diag., (2) the inverse of the variance–covariance matrix of the differences between simulated and pseudo-empirical moments form the first round of simulations, denoted GMM, and (3) the inverse of the variances of the differences between simulated and pseudo-empirical moments, denoted GMM-Diag.

Table 4: Monte Carlo results for regression-based distance functions.

	mom		*mom*, mom^2		*mom*. mom^2. mom^3	
	Mean	**Cov.**	**Mean**	**Cov.**	**Mean**	**Cov.**
a	1.459 (1.312, 1.615)	0.02	0.554 (0.427, 0.687)	0.10	0.367 (0.255, 0.484)	0.22
a_w	1.461 (1.325, 1.601)	0.02	0.558 (0.444, 0.670)	0.08	0.370 (0.269, 0.474)	0.20
b	1.129 (1.053, 1.206)	0.16	1.625 (1.567, 1.677)	0.10	1.847 (1.801, 1.890)	0.06
b_w	1.129 (1.060, 1.197)	0.14	1.624 (1.575, 1.670)	0.10	1.847 (1.806, 1.887)	0.06
σ	27.581 (26.917, 28.253)	0.14	29.200 (28.614, 29.784)	0.06	29.567 (29.035, 30.098)	0.10
σ_w	27.590 (26.980, 28.193)	0.08	29.202 (28.681, 29.709)	0.06	29.564 (29.104, 30.041)	0.10
tsec	3970.920 (832.358)		4314.540 (463.153)		4904.480 (1119.068)	
# eval.	87,943.340 (56463.140)		91,654.140 (29210.679)		109,927.420 (68678.096)	

The table shows the posterior means and their 95 percent credible intervals (in brackets) averaged over 50 Monte Carlo runs of parameter estimation via ABC algorithms for the model of Alfarano et al. using the regression-based distance functions proposed by Fearnhead and Prangle (2012). Also shown are the coverage ratios, i.e. the percentage of cases in which the 'true' parameter value falls into the credible intervals. This approach has been implemented using the raw moments as well as their squares and third powers used as additional regressors. Parameter estimates are obtained by performing the regression of Eq. (5) with the final sample of particles. The SMC algorithm used in this exercise was Method 3 as defined in the main text. The algorithm has been initiated by a large sample of 10^*N observations using the inverse of the variance-covariance matrix of the moments of the data (with $M = 11$). From this initial sample, the 10 percent best cases were used for a regression and the variance-covariance matrix of the regression residuals of this sample has been adopted as the new weight matrix. Subsequently, Method 3 was applied to the regression distances until the distances were all smaller than 0.2 percent of the initial ones. The sample size has again been $n = 20{,}000$ in every run. The final regression has been performed either as unweighted or weighted regression (the estimates from the latter identified by the index w).

intervals indeed convey a false impression of accuracy of the estimates. Computation time and number of iterations are also much higher than with the previous choices (Table 4).

My conjecture at this stage for the reason of the dismal performance of the regression-based distance function is that the reliance on a regression over the entire set of initial estimates might not be a useful approach in the presence of the bi-modal behaviour of the present process, as the dependency between parameters and moments might be highly non-monotonic. For the same reasons the choice of GMM-type weights in the distance function based on the weighted initial sample (as suggested by Prangle 2017) might bring in too much sensitivity to realisations that are located far from the true values.

5 Filtering

In a state-space framework, ABC can also be used for filtering information on the hidden variables from the measurement equation. Again, ABC would achieve this goal without the necessity of being able to compute an approximate likelihood function. Filtering within ABC has first been proposed by Jasra et al. (2012) using a very straightforward algorithm.

Algorithm 3: Filtering via ABC

1: create an initial particle swarm with elements x_i, $i = 1, \ldots, N$ of the hidden variable(s)
2: let n be the sample size of the measurement variable
3: **for** $\tau = 1, \ldots, n$ **do**
4: **for** $i = 1, \ldots, N$ **do**
5: generate $z_{i,\tau}$ from $f(\cdot|y_{i,\tau-1})$
6: compute $d(z_{i,\tau} - y_\tau)$ where y_τ is the observed variable
7: **end for**
8: resample the hidden variables using multidimensional draws based on their relative distances
9: **end for**

It essentially consists in replacing the state probabilities of a particle approximation to the likelihood by distances between the prediction of the observed variable and their realisations as detailed in Algorithm 3. The filtered values of unobserved variable x_τ are obtained as:

$$\hat{x}_\tau = \frac{x_{i,\tau}\ d(z_{i,\tau} - y_\tau)}{\sum_{i=1}^{N} d(z_{i,\tau} - y_\tau)}. \tag{13}$$

With $d(z_{i,\tau} - y_\tau)$ the deviation between the prediction of the i'th element of the ABC filter and the empirical observation at time τ.

Figure 5 shows an example of the hidden variable in a simulation of the Alfarano et al. model together with its filtered reconstruction using both the standard particle filter and the ABC filter. As it can be seen both filters work well for most of the time except for the ABC filter showing a longer period of deviation from the underlying hidden variable at the beginning of the series. In this model, when the sentiment variable x_t is close to one of the extremes, relatively little information on x_t is contained in the measurement y_t, and, hence, when the filter does not get the initial switch towards one of the extremes right, this 'misinterpretation' is hardly corrected as long as sentiment remains in the same mode. While not obvious from Figure 5, also the particle filter is not immune to this aberrance. Table 5 evaluates the root mean squared error of both filters over 50 replications of this filtering experiment. We have estimated the parameters as above via ABC using both methods 2 and 3, and we have alternatively also performed the filtering operations with known rather than estimated parameters. With known parameters we find a more pronounced advantage of the particle filter which we would expect to be better as it approximates the optimal filter numerically. However, with the added uncertainty of the estimated parameters, the ranking is not so clear anymore, and the errors of both methods are relatively similar. Since filtering on the base of information up to time $t-1$ is equivalent to out-of-sample forecasting, these results also provide some promising perspective for forecasting based on ABC estimation. In a recent paper, Frazier et al. (2019) show that asymptotically, for a sufficiently large sample size, there is no difference in accuracy between exact predictions and those based on an ABC approximation, and they also show with various examples that poor estimation of the posterior of the parameters is not necessarily reflected in poor out-of-sample forecasts.

6 An application

Table 6 shows the outcome of the estimation of the ALW model for the complete available historical data set of daily observations of the S&P500 from the beginning of 1928 to the end of the year 2020 (a total of 24,142 data points) using either $M = 7$,11 or 15 moments. Parameter estimates after regression adjustment all fullfill the bimodality condition $a < b$. While parameter b shows a certain tendency to increase with the number of moments, a, σ_f in contrast show no clear pattern across moments.

Overall, the numbers are in good agreement with results for shorter portions of the S&P500 record reported in Lux 2018, 2022.

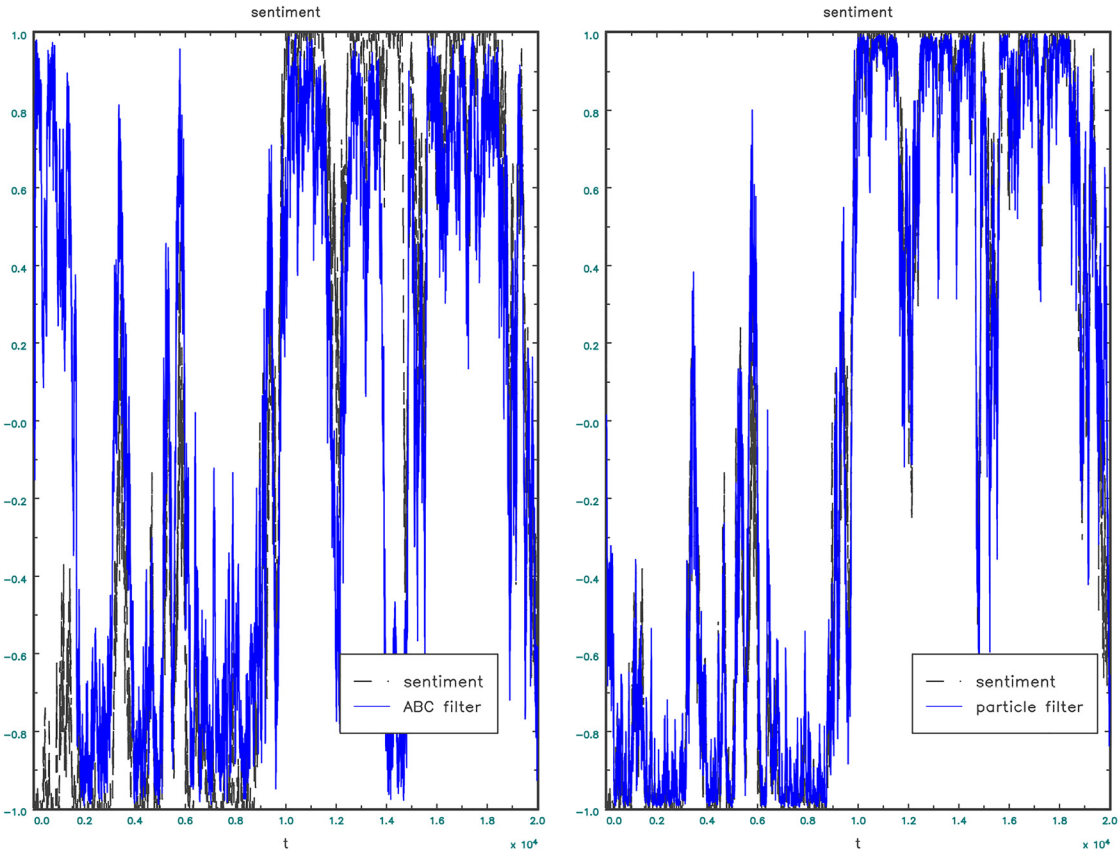

Figure 5: Reconstruction of the hidden sentiment variable x_t in one exemplary simulation using both the ABC and the particle filter.

Table 5: RMSE of filtered sentiment.

Estim. param.	Method 2	Method 3
ABC filter	0.507 (0.179)	0.542 (0.203)
Particle filter	0.512 (0.283)	0.512 (0.262)
True param		
ABC filter	0.550 (0.209)	
Particle filter	0.465 (0.296)	

The table shows the average root mean squared errors (RMSEs) and their standard deviations (in brackets) of 50 Monte Carlo runs of filtered sentiment, both either using the estimated parameters on the base of ABC algorithms 2 and 3 and the true underlying parameters. The ABC filter is computed as described in Algorithm 3. The ABC estimates and true parameters have been both combined with the ABC filter and with a particle filter (the later providing a stochastic approximation of the likelihood and the state probabilities of the current realization of sentiment). To allow for a certain transient after initialization, the RMSE has been averaged over periods 2,001 through 20,000 of a simulation with 20,000 realizations.

In terms of the Akaike information criterion, the best performing estimation is the parsimonious version with only 7 moments, after heteroscedastic correction. However, the extremely large variation of the AIC values might indicate that this comparison is not too reliable. Since the standard deviation of S&P500 returns is 11.644, the estimates for the standard deviation of fundamental news indicate that the model preferred by

the AIC criterion would assign about fourty percent of the market's volatility to the influence of sentiment under the preferred specification.

Figure 6 shows the S&P500 returns together with filtered sentiment (using the ABC and the particle filter). The figure indicates a dominance of optimistic sentiment over most of the history of the S&P500. Given its secular trend, this is not too surprising. The only time, the reconstructed sentiment reaches out into the negative region is during the great depression (the fluctuations at the very beginning of the record

Table 6: ABC estimation of ALW model for S&P 500.

	Estimates	Linear corr.	Heterosc. corr
# of moments	7		
a	0.226	0.151	0.170
	(0.001, 1.905)	(−0.361, 0.958)	(−0.236, 0.807)
b	1.366	0.836	0.844
	(0.004, 3.861)	(−0.088, 2.508)	(0.329, 1.854)
σ_f	7.595	7.091	7.103
	(1.858, 12.221)	(4.664, 9.642)	(4.580,9.738)
AIC	16,051.718	−56,341.729	−74,474.921
tsec	10,623		
No. of evaluations	506,799		
Misspec. test	0.585 (0.824, 0.931)		
# of moments	11		
a	0.524	0.600	0.614
	(0.001, 2.632)	(−0.209, 1.586)	(−0.065, 1.522)
b	0.753	5.242	5.342
	(0.001, 3.829)	(4.388, 6.730)	(1.835, 12.675)
σ_f	7.740	10.072	10.064
	(1.976, 12.604)	(8.042, 11.902)	(7.985, 11.705)
AIC	28,279.902	−57,665.884	8598.393
tsec	11,745		
No. of evaluations	449,527		
Misspec. test	5.976 (0.705, 0.631)		
# of moments	15		
a	0.628	0.086	0.119
	(0.001, 2.780)	(-0.813, 1.196)	(-1.063, 1.691)
b	0.599	4.218	4.286
	(0.001, 3.225)	(3.408, 5.833)	(1.923, 9.630)
σ_f	7.544	5.480	5.467
	(2.015, 12.381)	(3.895, 6.822)	(0.036, 9.483)
AIC	26,284.282	−65,161.022	54,872.529
tsec	13,143		
No. of evaluations	435,108		
Misspec. test	6.109 (0.656, 2.756)		

The table exhibits posterior means together with their 95 pecent credible intervals (in brackets) for the ALW model applied to daily S&P500 records over the time horizon January 1928 to December 2020 (24,142 observations). All other settings are exactly as in the Monte Carlo experiments of Section 4. The label 'misspec. test' shows the results of the misspecification test detailed in the main text together with the simulated 95 percent boundaries under the assumption that either the original SMC estimates or the linearly corrected ones are the 'true' parameters. SMC-ABC method 3 has been used with $N = 20{,}000$ particles in the original estimation, while for the replications in the test for missspecification, only 1000 particles have been used in each run.

should be disregarded as they are presumably rather transitory adjustments of the filter upon initialisation). The second most important drop is during the recent crisis around 2008. Figure 6 also shows that the ABC filter and the traditional particle filter are visually almost indistinguishable, and the same applies if different sets of moments are used for the estimation for the underlying parameters. Since we have seen that the ALW model might not always pick up the right sign of the sentiment variation, its filtered sentiment variable is better interpreted as an indication of convergence versus divergence of opinions. High values would then signal a homogeneous market climate, while drops would indicate divergence of opinions. Another interpretation is, therefore, as a behavioral measure of volatility and it can indeed be applied to forecast volatility adding valuable input to out-of-sample forecasts that are not encompassed by, for example, standard GARCH predictions (as demonstrated by Ghonghadze and Lux 2016).

The large variation of parameter estimates before and after the regression adjustment, and also between different sets of moments might raise doubts whether the present model is indeed correctly specified. Indeed, it would be more than remarkable if such a simple framework would provide a correct specification for the dynamics of stock market returns over almost one century. The topic of model misspecification has been addressed in the ABC literature by Prangle et al. (2014) and Frazier, Robert, and Rousseau (2020), among others. Frazier et al. show that under model misspecification, credible sets may come with any level of coverage, and posteriors before and after regression adjustment could be concentrated in subsets of the parameter space that are far apart. Very narrow credible intervals of posteriors after regression adjustment could then provide a false impression of accuracy of the so obtained results. In order to detect model misspecification Frazier et al. propose various tests, one of which we adopt here. The misspecification test is

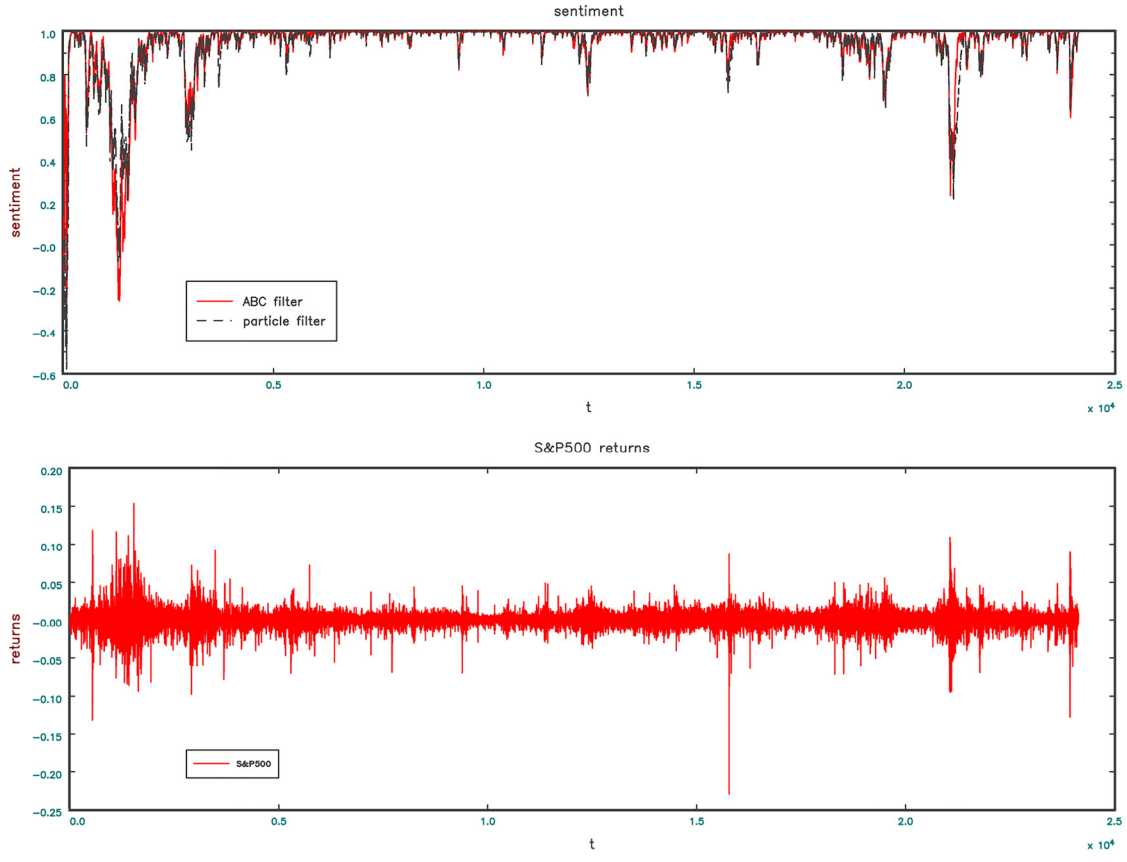

Figure 6: S&P500 daily returns from 1928 to 2020 (lower panel) and filtered sentiment x_f from the Alfarano et al. model using parameter estimates obtained by ABC-SMC with 15 moments (upper panel).

based on the insight that posterior distributions before and after regression adjustment will converge to the same limit for correctly specified models as long as the tolerance level is chosen appropriately. Under model misspecification, however, the two posteriors will diverge.

To diagnose misspecification, the difference between posteriors before and after regression adjustment can be simulated under correct specification, i.e., from the underlying model with parameters set to their posterior means. While this approach can be implemented for any function of interest, we apply it directly to the parameters of the model. The statistics shown in Table 6 with the label "misspec. test" gives the Euclidian distance between the estimates before and after linear regression adjustment. Because of the different magnitudes of the three parameters, relative distances are used. The pertinent numbers are compared with the 95 percent quantile obtained from Monte Carlo simulations of the Alfarano et al. model with parameters set equal to the posterior means. To this end, the ABC estimation plus linear regression adjustment has been repeated 100 times with simulated time series of length identical to the S&P 500 series. The brackets show two such 95 percent quantiles for each set of moments: The first assumes the posterior mean before regression adjustment is the 'true' parameter value, the second one assumes this for the posterior mean after the regression. If the empirical distance is larger than the 95 percent quantile, we find evidence for misspecification of the model.

Table 6 shows that the results of this test turn out to be different across different set of moments: while for $M = 7$, the test provides no indication of misspecification, the large differences between posterior means before and after regression adjustment for $M = 11$ and $M = 15$ are a strong signal for misspecification as the empirical distances by far exceed the simulated 95 percent quantiles. These divergent findings can be linked to the well-known stylized facts of financial returns, fat tails and volatility clustering. Elementary moments for both of these salient features are already represented in the first sets of moments with $M = 7$. The richer sets of moments add longer autocorrelations of squared and absolute returns. Adding such moments targets the hyperbolic decline of these autocorrelation functions that is found in the data, and that captures the long-term dependence in volatility. However, a simple model like the present one with Markovian behavior can mimic this decay pattern of the autocorrelation only to a certain extent. The misspecification does not become apparent with unconditional moments and short-run autocorrelations only, but is revealed more easily when higher-order lags are included. While it is well-known that agent-based models of the type presented above can only generate spurious long-term dependence of volatility (as demonstrated analytically by Alfarano and Lux 2007, and Alfarano, Lux, and Wagner 2008), the strong rejection of the present specification after consideration of a few long lags in the autocovariances is surprising. This sensitivity stands in contrast to estimates of statistics like the Hurst exponent that typically indicate spurious long-term dependence (i.e., hyperbolic decay of the autocorrelation functions) of such Markov processes. Against this background, the ABC-based specification test appears surprisingly sensitive.

7 Conclusions

The present paper has demonstrated the application of Approximate Bayesian Computation (ABC) for parameter inference for agent-based models in economics and finance. Our example of a well-studied model of speculative interaction suggests that when comparing the mean parameter estimates of the posterior of an ABC estimation with estimates obtained from GMM with the same moments as input, the former turn out to be more efficient. While GMM estimates can hardly be improved upon, ABC still provides the opportunity for higher precision estimates when one were willing to invest more computational resources to achieve a closer match to the moments of the data. In contrast to numerical approximations of the likelihood, ABC allows to cope with sample sizes that are far outside the reach of these methods. Besides parameter inference, ABC also provides a straight forward avenue for filtering information on hidden variables. At least in our example, the ABC filter turned out to perform practically on par with the asymptotically optimal particle filter. Since research on ABC has been flourishing in recent literature, and its applications in agent-based models in ecology and evolutionary genetics are mushrooming, we expect that it will gain a broader footing in economics and finance with the advancement of agent-based models.

Acknowledgments: The very detailed and helpful comments of an anonymous reviewer are gratefully acknowledged.
Author contribution: The author has accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The author declares no conflicts of interest regarding this article.

Appendix A: Details of Alfarano, Lux, and Wagner (2008) model

The model assumes that two groups of traders exist in a financial market: chartists and fundamentalists. The number of chartists is N_c with each one of them commanding a trading volume T_c per integer time unit (which might be a day). Chartists are subject to fluctuations of sentiment which either leads to an optimistic or pessimistic attitude. Sentiment formation is the agent-based part of the model, and its formalization via pairwise interaction is formalized by the transition rates of Eq. (9) in the main text, which are used to determine the switches of each member of this group from optimistic to pessimistic and *vice versa* as a Poisson process with time-varying intensity depending on the overall numbers of optimists or pessimists among their peers. Equation (9) thus, captures in a very direct way, how sentiment becomes contagious through herd behavior of investors. In the simulations of this model, the number of chartists is set equal to $N_c = 100$ and the behavior of each one of them is simulated with the Gillespie (1977) algorithm for discrete events in continuous time. In contrast to chartists who are considered as individuals, the group of fundamentalists is aggregated in a traditional way summing over the excess demand functions of its members. With N_f members of this group and trading volume T_f measured per unit of deviation between the assumed fundamental value (p_f) and the current (log) market price (p_t), their excess demand $\mathrm{ED}_{f,t}$ is:

$$\mathrm{ED}_{f,t} = N_f T_f (p_t - p_{f,t}).$$

Excess demand of chartists is simply assumed to be positive if they are optimistic, and negative if pessimistic, so that the second component of market excess demand amounts to:

$$\mathrm{ED}_{c,t} = T_c N_c x_t$$

with x_t the sentiment indicator $x_t \equiv \frac{n_{+,t} - n_{-,t}}{N_c}$ computed very much along the lines of such indicators as they are used in practice. Instantaneous market clearing leads to an equilibrium price

$$p_t = p_{t,f} + \frac{T_c N_c}{T_f N_f} x_t$$

which shows the potential for mispricing generated by the non-fundamental traders. We, finally, obtain the simple expression of Eq. (10) in the main text for returns over unit time intervals by assuming that the fundamental value follows Brownian motion with a variance σ_f^2 and setting $\frac{T_c N_c}{T_f N_f} = 1$. The later assumption is justified by the observation that the two parameters of the agent-based process depicted in eq. (9), a and b, are already capable of generating a wide variety of outcomes for the conditional and unconditional distribution of asset returns. Due to proximity to collinearity of the prefactor and the behavioral parameters, including the composite expression $\frac{T_c N_c}{T_f N_f}$ as another parameter for the estimation, would be cumbersome. The model above is attractive as a test example for an agent-based framework as its interesting properties are intrinsically linked to the interaction of the chartist agents, and cannot be replicated easily by any aggregation device.

Appendix B: Results of exemplary estimations of Table 2 with moment sets $M = 4, 7$ and 11

Table B1: ABC estimation of ALW model.

Method 1				**Method 3**			
# of moments	**4**			**# of moments**	**4**		
	Estimates	**Linear corr.**	**Heterosc. corr**		**Estimates**	**Linear corr.**	**Heterosc. corr**
a	0.830 (0.456)	0.723 (0.272)	0.722 (0.271)	a	0.852 (0.326)	0.733 (0.235)	0.733 (0.244)
b	1.410 (0.568)	1.472 (0.373)	1.499 (0.318)	b	1.292 (0.145)	1.313 (0.089)	1.313 (0.077)
σ_f	17.574 (11.847)	19.485 (7.784)	19.485 (7.701)	σ_f	16.324 (9.007)	19.257 (6.775)	19.263 (6.861)
AIC	2244.269	−438.777	−755.198	AIC	−1703.467	−3865.423	−4039.3293
tsec	13,009			tsec	331		
No. of evaluations	443853			No. of evaluations	11320		
Method 2				**Method 4**			
# of moments	**4**			**# of moments**	**4**		
	Estimates	Linear corr.	Heterosc. corr		Estimates	Linear corr.	Heterosc. corr
a	0.752 (0.403)	0.690 (0.252)	0.694 (0.253)	a	0.826 (0.424)	0.643 (0.220)	0.640 (0.219)
b	1.377 (0.249)	1.301 (0.182)	1.287 (0.139)	b	1.317 (0.202)	1.305 (0.178)	1.365 (0.272)
σ_f	18.652 (11.559)	18.150 (7.495)	18.137 (7.550)	σ_f	17.122 (11.055)	20.639 (6.325)	20.592 (6.346)
AIC	300.975	−2109.340	−2594.779	AIC	−109.081	2767.830	−1884.762
tsec	1446			tsec	513		
No. of evaluations	67439			No. of evaluations	18500		
Method 1				**Method 3**			
# of moments	**7**			**# of moments**	**7**		
	Estimates	Linear corr.	Heterosc. corr		Estimates	Linear corr.	Heterosc. corr
a	0.788 (0.505)	0.302 (0.221)	0.302 (0.166)	a	0.755 (0.355)	0.361 (0.165)	0.361 (0.165)
b	1.362 (0.672)	1.666 (0.488)	1.709 (0.540)	b	1.264 (0.226)	1.349 (0.121)	1.351 (0.096)
σ_f	20.847 (10.534)	31.259 (4.115)	31.266 (3.813)	σ_f	19.928 (8.474)	29.378 (3.686)	29.384 (3.215)
AIC	2547.091	−1572.991	−2054.828	AIC	−770.943	−5161.544	−5871.325
tsec	6228			tsec	356		
No. of evaluations	207649			No. of evaluations	9018		
Method 2				**Method 4**			
# of moments	**7**			**# of moments**	**7**		
	Estimates	Linear corr.	Heterosc. corr		Estimates	Linear corr.	Heterosc. corr
a	0.630 (0.435)	0.307 (0.169)	0.313 (0.153)	a	0.648 (0.420)	0.271 (0.165)	0.272 (0.156)
b	1.461 (0.570)	1.417 (0.428)	1.380 (0.289)	b	1.520 (0.582)	1.302 (0.431)	1.263 (0.323)
σ_f	23.014 (10.889)	29.022 (4.072)	29.045 (4.508)	σ_f	23.182 (11.460)	25.741 (5.510)	25.599 (5.354)
AIC	1988.859	−2394.745	−3133.938	AIC	2060.239	−1826.285	−2522.144
tsec	1021			tsec	554		
No. of evaluations	40071			No. of evaluations	17000		
Method 1				**Method 3**			
# of moments	**11**			**# of moments**	**11**		
	Estimates	Linear corr.	Heterosc. corr		Estimates	Linear corr.	Heterosc. corr
a	0.676 (0.411)	0.219 (0.149)	0.220 (0.118)	a	0.656 (0.346)	0.207 (0.134)	0.208 (0.201)
b	1.479 (0.727)	1.640 (0.561)	1.696 (0.702)	b	1.368 (0.288)	1.559 (0.199)	1.580 (0.254)
σ_f	21.280 (10.660)	33.055 (3.109)	33.058 (2.922)	σ_f	21.151 (9.137)	33.621 (2.853)	33.623 (5.751)
AIC	2316.248	−2618.748	−2693.632	AIC	−183.090	−5077.424	−2305.322
tsec	10,155			tsec	490		
No. of evaluations	334774			No. of evaluations	10523		

Table B1: (continued)

Method 2				**Method 4**			
# of moments	**11**			**# of moments**	**11**		
	Estimates	**Linear corr.**	**Heterosc. corr**		**Estimates**	**Linear corr.**	**Heterosc. corr**
a	0.603 (0.384)	0.247 (0.139)	0.255 (0.099)	a	0.733 (0.182)	0.228 (0.061)	0.196 (0.047)
b	1.453 (0.475)	1.614 (0.421)	1.671 (0.601)	b	1.483 (0.204)	1.677 (0.291)	1.849 (0.361)
σ_f	22.900 (10.295)	32.633 (3231)	32.624 (3237)	σ_f	18.800 (5.011)	27.454 (4.038)	28.333 (4.083)
AIC	1261.959	−3257.719	−3158.271	AIC	−3358.088	−5209.842	−5203.432
tsec	1265			tsec	587		
tsec	1021						
No. of evaluations	40071			No. of evaluations	12500		

Continuation of Table 1 from main text: Example of ABC estimation of the parameters of the model of Alfarano et al., using four different ABC algorithms with distance functions composed of all 4,7 and 11 moments, with and without correction of the initial estimates via a standard weighted regression as well as via a weighted nonlinear regression allowing for heteroscedasticity. The sample size has been $n = 20{,}000$. The table shows the means and standard deviations of the posteriors of the parameters together with the value of the AIC criterion for each estimation. Bold numbers indicate the estimates preferred by the AIC. The estimates of the parameters are all multiplied by 10^4 for better readability. The table also provides information on the computation time in seconds (tsec), and the number of evaluations needed to arrive at the final set of accepted particles.

References

Alfarano, S., and T. Lux. 2007. "A Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory." *Macroeconomic Dynamics* 11 (1): 80–101.

Alfarano, S., T. Lux, and F. Wagner. 2008. "Time Variation of Higher Moments in a Financial Market with Heterogeneous Agents: An Analytical Approach." *Journal of Economic Dynamics and Control* 32 (1): 101–36.

Beaumont, M. A., J.-M. Cornuet, J.-M. Marin, and C. P. Robert. 2009. "Adaptive Approximate Bayesian Computation." *Biometrika* 96 (4): 983–90.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. "Approximate Bayesian Computation in Population Genetics." *Genetics* 162 (4): 2025–35.

Blum, M. G., and O. François. 2010. "Non-Linear Regression Models for Approximate Bayesian Computation." *Statistics and Computing* 20 (1): 63–73.

Blum, M. G., M. A. Nunes, D. Prangle, S. A. Sisson. 2013. "A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation." *Statistical Science* 28 (2): 189–208.

Chen, Z., and T. Lux. 2018. "Estimation of Sentiment Effects in Financial Markets: A Simulated Method of Moments Approach." *Computational Economics* 52: 711–44.

Csilléry, K., M. Blum, O. Gaggiotti, and O. François. 2010. "Approximate Bayesian Computation (ABC) in Practice." *Trends in Ecology & Evolution* 25: 410–8.

Del Moral, P., A. Doucet, and A. Jasra. 2012. "An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation." *Statistics and Computing* 22 (5): 1009–20.

Drovandi, C. C., and A. N. Pettitt. 2011. "Estimation of Parameters for Macroparasite Population Evolution Using Approximate Bayesian Computation." *Biometrics* 67: 225–33.

Fearnhead, P. 2019. "Asymptotics of ABC." In *Handbook of Approximate Bayesian Computation*, edited by Y. F. Sisson, and S. M. Beaumont. London: Chapman & Hall. chapter 10.

Fearnhead, P., and D. Prangle. 2012. "Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic Approximate Bayesian Computation." *Journal of the Royal Statistical Society: Series B* 74 (3): 419–74.

Franke, R., and F. Westerhoff. 2012. "Structural Stochastic Volatility in Asset Pricing Dynamics: Estimation and Model Contest." *Journal of Economic Dynamics and Control* 36 (8): 1193–211.

Frazier, D., W. Maneesoonthorn, G. Martin, and B. McCabe. 2019. "Approximate Bayesian Forecasting." *International Journal of Forecasting* 35: 521–39.

Frazier, D. T., G. M. Martin, C. P. Robert, and J. Rousseau. 2018. "Asymptotic Properties of Approximate Bayesian Computation." *Biometrika* 105 (3): 593–607.

Frazier, D. T., C. P. Robert, and J. Rousseau. 2020. "Model Misspecification in Approximate Bayesian Computation: Consequences and Diagnostics." *Journal of the Royal Statistical Society: Series B* 82 (2): 421–44.

Ghonghadze, J., and T. Lux. 2016. "Bringing an Elementary Agent-Based Model to the Data: Estimation via GMM and an Application to Forecasting of Asset Price Volatility." *Journal of Empirical Finance* 37: 1–19.
Gillespie, D. T. 1977. "Exact Stochastic Simulation of Coupled Chemical Reactions." *Journal of Physical Chemistry* 81 (25): 2340–61.
Gilli, M., and P. Winker. 2001. "Indirect Estimation of the Parameters of Agent Based Models of Financial Markets." In *Working Paper*. University of Geneva.
Grazzini, J., M. G. Richiardi, and M. Tsionas. 2017. "Bayesian Estimation of Agent-Based Models." *Journal of Economic Dynamics and Control* 77: 26–47.
Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–54.
Jang, T.-S., and S. Sacht. 2016. "Animal Spirits and the Business Cycle: Empirical Evidence from Moment Matching." *Metroeconomica* 67 (1): 76–113.
Jasra, A., S. Singh, J. S. Martin, and E. McCoy. 2012. "Filtering via Approximate Bayesian Computation." *Statistics and Computing* 22 (6): 1223–37.
Lenormand, M., F. Jabot, and G. Deffuant. 2013. "Adaptive Approximate Bayesian Computation for Complex Models." *Computational Statistics* 28 (6): 2777–96.
Li, W., and P. Fearnhead. 2018. "On the Asymptotic Efficiency of Approximate Bayesian Computation Estimators." *Biometrika* 105 (2): 285–99.
Lux, T. 2018. "Estimation of Agent-Based Models Using Sequential Monte Carlo Methods." *Journal of Economic Dynamics and Control* 91: 391–408.
Lux, T. 2022. ""Bayesian Estimation of Agent-Based Models via Adaptive Particle Markov Chain Monte Carlo", Computational Economics, in press." In *Working Paper*. University of Kiel.
Lux, T., and R. C. Zwinkels. 2018. "Empirical Validation of Agent-Based Models." In *Handbook of Computational Economics*, 4, edited by C. Hommes, and B. Lebaron, 437–88. Amsterdam: Elsevier.
Prangle, D. 2017. "Adapting the ABC Distance Function." *Bayesian Analysis* 12: 289–309.
Prangle, D., M. G. Blum, G. Popovic, and S. Sisson. 2014. "Diagnostic Tools for Approximate Bayesian Computation Using the Coverage Property." *Australian & New Zealand Journal of Statistics* 56 (4): 309–29.
Rubin, D. B. 1984. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician." *Annals of Statistics* 12 (1): 1151–72.
Sisson, S. A., Y. Fan, and M. M. Tanaka. 2007. "Sequential Monte Carlo without Likelihoods." *Proceedings of the National Academy of Sciences of the United States of America* 104 (6): 1760–5.
Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. "Inferring Coalescence Times from DNA Sequence Data." *Genetics* 145 (2): 505–18.
Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf. 2009. "Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems." *Journal of The Royal Society Interface* 6 (31): 187–202.
Tubbenhauer, T., C. Fieberg, and T. Poddig. 2021. "Multi-Agent-Based VaR Forecasting." *Journal of Economic Dynamics and Control* 131. in press.