



Bayesian estimation and likelihood-based comparison of agent-based volatility models

Nils Bertschinger^{1,2}  · Iurii Mozzhorin²

Received: 1 July 2019 / Accepted: 26 May 2020 / Published online: 24 June 2020
© The Author(s) 2020

Abstract

The statistical description and modeling of volatility plays a prominent role in econometrics, risk management and finance. GARCH and stochastic volatility models have been extensively studied and are routinely fitted to market data, albeit providing a phenomenological description only. In contrast, agent-based modeling starts from the premise that modern economies consist of a vast number of individual actors with heterogeneous expectations and incentives. Observed market statistics then emerge from the collective dynamics of many actors following heterogeneous, yet simple rules. On the one hand, such models generate volatility dynamics, qualitatively matching several stylized facts. On the other hand, they illustrate the possible role of different mechanisms, such as chartist trading and herding behavior. Yet, rigorous and quantitative statistical fits are still mostly lacking. Here, we propose *Hamiltonian Monte Carlo*, an efficient and scalable Markov chain Monte Carlo algorithm, as a general method for Bayesian inference of agent-based models. In particular, we implement several models by Vikram and Sinha, Franke and Westerhoff and Alfarano, Lux and Wagner in Stan, an accessible probabilistic programming language for Bayesian modeling. We also compare the performance of these models with standard econometric models of the GARCH and stochastic volatility families. We find that the best agent-based models are on par with stochastic volatility models in terms of predictive likelihood, yet exhibit challenging posterior geometries requiring care in model comparison and sophisticated sampling algorithms.

Keywords Agent-based models · Stochastic volatility models · Bayesian estimation · Hamiltonian Monte Carlo

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11403-020-00289-z>) contains supplementary material, which is available to authorized users.

✉ Nils Bertschinger
bertschinger@fias.uni-frankfurt.de

¹ Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany

² Goethe University, Frankfurt am Main, Germany

JEL Classification C11 · C52 · C58 · G12

I wonder who it was defined man as a rational animal. It was the most premature definition ever given. Man is many things, but he is not rational.

Oscar Wilde, *The Picture of Dorian Gray and Other Writings*

1 Introduction

Financial markets exhibit some remarkable and often surprisingly stable statistical signatures, often referred to as *stylized facts* (Cont 2001; Lux 2009). Most notable and researched are the properties of asset price returns exhibiting fat-tailed distributions and volatility clustering. Volatility in particular has received much attention in the econophysics community for its autocorrelation decaying as a power law suggesting a long-memory process. Volatility also plays a prominent role in econometrics, risk management and finance. Correspondingly, phenomenological statistical models such as GARCH (Bollerslev 1986) and stochastic volatility models (Kim et al. 1998) are extensively studied and routinely fitted to market data.

Agent-based models consider the statistical signatures of financial markets as emergent properties, i.e., arising from the collective actions of many interacting traders. They provide a complement to standard economic models, which, presuming rational actors, are often unable to explain the rapid changes in volatility between calm market phases and highly volatile episodes. Shiller has coined the term excess volatility, hinting at these shortcomings (Shiller 1980). In contrast, agent-based models allow for bounded rational actors and can often reproduce the stylized facts presuming chartist trading and/or herding behavior (Samanidou et al. 2007).

Agent-based models of speculative behavior in financial markets are nowadays able to replicate many stylized facts simultaneously. They provide an alternative to standard econometric models, offering behavioral explanations of observed market statistics (Lux 2009; LeBaron 2000). Yet, estimation of such models is still challenging and has mostly resorted to simulation-based methods striving to match selected moments of the data (Franke and Westerhoff 2011; Ghonghadze and Lux 2016). More recently, direct comparisons between the probabilistic dynamics of simulated and observed time series have been proposed. To this end, transition probabilities of return time series, observed and simulated, are discretized and compared in terms of context-tree weighted Markov approximations (Barde 2016, 2017) or JS divergence (Lamperti 2018). Alternatively, predictive likelihoods are estimated on the continuous time series of returns via (Guerini and Moneta 2017) fitting a VAR model or (Kukacka and Barunik 2017) kernel density estimation and again compared and matched with model predictions. These methods go beyond moment matching and allow to generate and evaluate model predictions. They still fall short to recover latent dynamical states as estimates are not conditional to data but merely matched to their probabilistic structure. In order to recover latent dynamics, either maximum likelihood estimation or Bayesian methods are required. Estimating agent-based models in this fashion is challenging as their dynamics are often highly nonlinear. Yet recently, sequential Monte Carlo methods (Lux 2018) and the unscented Kalman filter (Majewski et al. 2018) have been successfully used in this context.

Here, we follow this line of research and utilize modern software tools from machine learning and statistics to fit agent-based market models. In particular, we employ *Stan* (2017), a probabilistic programming language for Bayesian modeling, to fit several different agent-based models, namely from Vikram & Sinha (2011), Franke & Westerhoff (2012) and Alfarano, Lux & Wagner (2008). We believe that Bayesian estimation has many advantages as it allows to access parameter uncertainties as well as to generate model predictions. Furthermore, being based on the full model probability, including the likelihood, different models can be systematically compared, e.g., based on their predictive likelihood on held-out data. Indeed, for similar reasons Bayesian estimation is popular in macroeconomics for estimating classical DSGE (An and Schorfheide 2007) as well as agent-based models (Grazzini et al. 2017).

Overall, our contribution is threefold: First, we discuss several agent-based models and the behavioral assumptions they are based on. In particular, this includes the model by Vikram & Sinha, which had not been fitted before, as well as a novel moving average specification for the model of Franke & Westerhoff. Secondly, we implement all models in *Stan*, a modern probabilistic programming language for Bayesian modeling. Thirdly, we provide a detailed pairwise comparison of all models based on cross-validated predictive likelihoods. In particular, we find that the best agent-based models are competitive with standard econometric models. While this has been observed previously for the Franke & Westerhoff model (Barde 2016), we provide evidence that other herding dynamics give comparable results provided that the model allows for persistent mispricing between fundamental and observed prices.

Our presentation is structured as follows: In Sect. 2, we introduce Bayesian data modeling and Markov chain Monte Carlo (MCMC) algorithms. In particular, we shortly explain Hamiltonian Monte Carlo (HMC) and how it is implemented in *Stan*. Then, in Sect. 3 we introduce all considered models and express them as probabilistic models for return data. Our results on simulated as well as actual S&P 500 stock return data are summarized in Sect. 4. Finally, we conclude by discussing our main findings in Sect. 5.

2 Stan and Hamiltonian MCMC

2.1 Bayesian modeling

In Bayesian modeling observed data,¹ $\mathbf{x} = (x_1, \dots, x_N)$ are related to unobserved parameters/latent variables $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ in terms of a joint probability distribution with density $p(\mathbf{x}, \boldsymbol{\theta})$. This density is usually factorized as $p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, i.e., into the parameter likelihood and prior density. Inference then rests on Bayes rule to obtain the density of the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})}$$

¹ Vectors are denoted with bold symbols throughout the text.

where the normalization is given by $p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. The posterior summarizes the information obtained about the unobserved parameters $\boldsymbol{\theta}$ and combines the a priori assessment of the modeler, $p(\boldsymbol{\theta})$, with the information obtained from the data $p(\mathbf{x}|\boldsymbol{\theta})$. Conceptually, Bayesian estimation boils down to a rather mechanical application of Bayes rule, once the full model $p(\mathbf{x}, \boldsymbol{\theta})$ is specified. Below, we will explain how this applies to different agent-based models and discuss, in particular, the role of prior choices in Bayesian modeling.

In practice, the normalization constant $p(\mathbf{x})$ of the posterior density is often intractable, involving an integral over the parameter space. Accordingly, many approximation methods have been proposed which either aim to approximate it with a tractable density or allow to draw posterior samples from its unnormalized density. Hamiltonian Monte-Carlo (HMC) sampling is an example of the latter approach. As a Markov chain Monte Carlo method, it produces a sequence of possibly correlated samples. A comprehensive and readable introduction to HMC and its properties can be found in Betancourt (2017). Here, a rather short overview of the method should suffice.

2.2 Markov chain Monte Carlo (MCMC)

Consider a target density $p^*(\boldsymbol{\theta})$, e.g., the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ from a Bayesian model. MCMC aims to construct a *transition density* $T(\boldsymbol{\theta}'|\boldsymbol{\theta})$ which leaves the target density invariant, i.e.,

$$p^*(\boldsymbol{\theta}') = \int T(\boldsymbol{\theta}'|\boldsymbol{\theta})p^*(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Such a transition density can then be utilized to draw a sequence of samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots$ with $p(\boldsymbol{\theta}_2, \dots|\boldsymbol{\theta}_1) = \prod_{i=1}^{\infty} T(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$. The Metropolis–Hastings algorithm uses two steps in order to compute a suitable transition starting at $\boldsymbol{\theta}_i = \boldsymbol{\theta}$:

1. Draw $\boldsymbol{\theta}'$ from a proposal density² $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$
2. Either retain the current sample, i.e., $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}$ or transition to $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}'$ with acceptance probability

$$a_{\boldsymbol{\theta}'|\boldsymbol{\theta}} = \min \left[1, \frac{p^*(\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{p^*(\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})} \right]. \quad (1)$$

This so-defined transition density not only leaves the target density invariant, but, under suitable conditions, also ensures that the chain converges to its unique invariant density starting from any initial condition $\boldsymbol{\theta}_1$ (Bishop 2011).

2.3 Hamiltonian Monte Carlo (HMC)

While, in theory, the Metropolis–Hastings algorithm can produce samples from the desired target density, especially in high dimensions it can suffer from slow conver-

² Choosing a suitable proposal density is a crucial step in the Metropolis–Hastings algorithm as it controls how effectively the resulting transitions can move across the sampling space.

gence. The underlying reason is that most of the probability mass is confined to a small subspace, the so-called typical set. If the proposal density is not well matched to the target density, many steps of the Markov chain are either rejected, e.g., when leaving the typical set, or slowly and randomly move along the typical set. HMC utilizes gradient information in order to generate long sweeps of the proposed states which are nevertheless accepted. To this end, HMC constructs a gradient flow which follows the typical set (Betancourt 2017).

Formally, HMC samples from an augmented state space (θ, \mathbf{m}) with density

$$\begin{aligned} p(\theta, \mathbf{m}) &= p(\theta)p(\mathbf{m}|\theta) \\ &= e^{\log p(\theta)+\log p(\mathbf{m}|\theta)} \\ &= e^{-H(\theta, \mathbf{m})}. \end{aligned}$$

The desired marginal density $p(\theta)$ can then be obtained by dropping the \mathbf{m} component of each joint sample. In analogy with physical systems, \mathbf{m} is considered as the momentum of a particle at position θ . Intuitively, \mathbf{m} controls the speed at which the position is moved along the typical set. The gradient flow following the typical set is obtained by integrating the *Hamiltonian* $H(\theta, \mathbf{m}) = -\log p(\theta) - \log p(\mathbf{m}|\theta)$ as

$$\begin{aligned} \dot{\theta} &= \frac{\partial}{\partial \mathbf{m}} H(\theta, \mathbf{m}) = -\frac{\partial}{\partial \mathbf{m}} \log p(\mathbf{m}|\theta) \\ \dot{\mathbf{m}} &= -\frac{\partial}{\partial \theta} H(\theta, \mathbf{m}) = \frac{\partial}{\partial \theta} \log p(\theta) + \frac{\partial}{\partial \theta} \log p(\mathbf{m}|\theta). \end{aligned}$$

Hamiltonian dynamics have several well-known properties. In particular, they are time-reversible and conserve volume and total probability $p(\theta, \mathbf{m})$. Thus, no further corrections are necessary when using the state (θ', \mathbf{m}') obtained by integrating the Hamiltonian for some time as a proposal. Due to time-symmetry and conservation of probability, the acceptance probability in Eq. (1) reduces to 1, i.e., the new state is always accepted.

HMC then proceeds in two steps. At each transition, a new momentum is sampled according to $p(\mathbf{m}|\theta)$ which is commonly taken as a Gaussian distribution independent of the current state θ , i.e., $\mathbf{m} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, the position is moved from θ with initial speed \mathbf{m} by following the gradient flow for some time. The final position (θ', \mathbf{m}') is then accepted as the next sample. By conservation of total probability, it holds that $p(\theta, \mathbf{m}) = p(\theta', \mathbf{m}')$ potentially leading to a long sweep across the typical set along a level contour of the probability density. Thus, at each step HMC first jumps to a new probability level and then follows the gradient flow at this level, thereby allowing for an efficient exploration of the typical set. In practice, the differential equation describing the Hamiltonian dynamics needs to be integrated numerically and care has to be taken that numerical errors do not accumulate. Fortunately, symplectic integrators can efficiently integrate Hamiltonian systems as numerical errors cancel and simulated trajectories closely approximate the theoretical dynamics. While time-symmetry and volume preservation are retained by symplectic integrators, the total probability is only

approximately conserved along numerical trajectories. Thus, in practice, HMC uses a Metropolis–Hastings step to either accept or reject the final position of a trajectory.

Especially in high-dimensional models, i.e., with many parameters, the use of gradient information to guide exploration is crucial to ensure efficient sampling. Note that HMC is restricted to continuous parameter spaces $\theta \in \mathbb{R}^K$, but could be combined with other methods when discrete parameters are desired. Often, it is advantageous to marginalize over discrete parameters as strong correlations between them can severely hinder efficient sampling.³ In theory, HMC is insensitive to strong correlations between parameters θ . In practice, the symplectic integrator uses a finite step size to numerically solve the Hamiltonian dynamics. In the case of posterior densities with high curvature, this can prevent the sampler to reach certain parts of the state space. Furthermore, it makes HMC sensitive to the scale of parameters as the step size would need to be adjusted accordingly.⁴ Fortunately, by reparameterizing the model it is often possible to simplify the geometry of the posterior density. Ongoing research explores the geometric aspects of HMC both from theoretical (Livingstone et al. 2019; Ma et al. 2015) and from practical (Betancourt and Girolami 2015) perspectives. Also, a range of novel convergence diagnostics, e.g., based on the stability of the numerical trajectory, have been developed (Betancourt 2017).

Next, we turn to agent-based models for financial markets and show how these can be expressed as statistical models. All models are then implemented and fitted models with the probabilistic programming language *Stan* (2017). Appendix A provides a short overview. The *Stan* code for all models is available in the online supplementary material. Implementing agent-based models in a well-tested Bayesian framework has several advantages. On the one hand, inference algorithms such as HMC supported by *Stan* are well optimized and tested. On the other hand, model specifications can be easily adapted and explored with small changes to their source code. Yet, *Stan* restricts the user to models with continuous random variables only. While many models, such as the ones presented below, can be approximated in this fashion in the limit of infinitely many agents, more detailed simulations involving many discrete agents are currently beyond the scope of existing toolboxes for probabilistic modeling. In this sense, our approach can be seen as a proof of concept which nevertheless covers several models of interest.

3 Market models

Here, we consider three models, in detail, namely by Vikram & Sinha (2011), Franke & Westerhoff (2012) and Alfarano, Lux & Wagner (2008). In particular, we explain how these models give rise to a latent state dynamics which can be simulated and estimated with Bayesian methods.

³ This especially applies to Gibbs sampling. Despite its popularity, Gibbs sampling is severely hindered by correlations in the posterior which can render it utterly useless in high-dimensional problems.

⁴ In contrast, Gibbs sampling is severely effected by strong dependencies but insensitive to the scale of parameters.

3.1 Model by Vikram & Sinha (VS)

The market in the VS model is populated by N traders. At each time step t , a trader i either buys ($S_i(t) = 1$), sells ($S_i(t) = -1$) or stays inactive ($S_i(t) = 0$). The normalized net demand from all traders is then given as $M_t = \frac{1}{N} \sum_{i=1}^N S_i(t)$, and the price adjusts as $p_{t+1} = \frac{1+M_t}{1-M_t} p_t$. An agent's decision to buy/sell or staying out depends on the perceived mispricing between the current price p_t and its running average $p_t^* = \langle p_t \rangle_\tau$ which is considered as a proxy for the fundamental price of the asset. The probability of an agent to trade is then given by

$$\mathbb{P}(|S_i(t)|) = \exp^{-\mu \left| \frac{p_t - p_t^*}{p_t^*} \right|}$$

and a trading agent buys $S_i(t) = 1$ or sells $S_i(t) = -1$ at random with equal probability.

In order to obtain a statistical model of volatility, in particular with a continuous latent state as required for HMC sampling, we have adapted the model as follows:

- For a large number of agents $N \rightarrow \infty$, the net demand M_t converges to a Gaussian distribution with mean zero (as $\mathbb{E}[S_i(t)] = 0$) and variance $\frac{\mathbb{P}(|S_i(t)|=1)}{\sqrt{N}}$.
- Here, we have used that agents trading decisions $S_i(t)$ are independent and

$$\begin{aligned} \mathbb{E}[S_i(t)] &= \frac{1}{2} \mathbb{P}(|S_i(t)| = 1) \cdot 1 + \frac{1}{2} \mathbb{P}(|S_i(t)| = 1) \cdot (-1) \\ &\quad + (1 - \mathbb{P}(|S_i(t)| = 1)) \cdot 0 = 0 \\ \text{Var}[S_i(t)] &= \mathbb{E}[S_i(t)^2] \\ &= \frac{1}{2} \mathbb{P}(|S_i(t)| = 1) \cdot 1^2 + \frac{1}{2} \mathbb{P}(|S_i(t)| = 1) \cdot (-1)^2 \\ &\quad + (1 - \mathbb{P}(|S_i(t)| = 1)) \cdot 0^2 \\ &= \mathbb{P}(|S_i(t)| = 1). \end{aligned}$$

- Next, considering the number of agents as unknown we introduce a scaling parameter σ_{\max}^2 for the variance and model the demand as $M_t \sim \mathcal{N}(0, \sigma_{\max}^2 \mathbb{P}(|S_i(t)| = 1))$.
- Finally, we approximate the log-return by linearizing the price impact⁵

$$\begin{aligned} r_{t+1} &= \log \frac{p_{t+1}}{p_t} \\ &= \log \frac{1 + M_t}{1 - M_t} \\ &\approx 2M_t \end{aligned}$$

⁵ We have also fitted the exact model, i.e., putting a normal distribution on the transformed returns $M_t = \frac{e^{r_{t+1}} - 1}{e^{r_{t+1}} + 1}$ without any noticeable difference.

where we have used that $\log(1+x) \approx x$ for $|x| \ll 1$.

Overall, we arrive at the following model dynamics⁶

$$\begin{aligned} \langle p_t \rangle_\tau &= (1 - \tau)p_t + \tau \langle p_{t-1} \rangle_\tau \\ \mathbb{P}(|S(t)| = 1) &= e^{-\mu \left| \log \frac{p_t}{\langle p_t \rangle_\tau} \right|} \\ r_{t+1} &\sim \mathcal{N}(0, \sigma_{\max}^2 \cdot 4\mathbb{P}(|S(t)| = 1)). \end{aligned} \quad (2)$$

Note that this is a state-space model with a continuous latent state driving the time-varying volatility $\sigma_{t+1} = \sqrt{\sigma_{\max}^2 \cdot 4\mathbb{P}(|S(t)| = 1)}$. Indeed, the famous GARCH(1, 1) (generalized auto-regressive conditional heteroscedastic) model (Bollerslev 1986) is of a similar form

$$\begin{aligned} \sigma_{t+1}^2 &= \alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma_t^2 \\ r_{t+1} &\sim \mathcal{N}(\mu, \sigma_{t+1}^2) \end{aligned} \quad (3)$$

The main difference between the VS (in our formulation) and the GARCH model is that the volatility is a function of past prices in the former and past returns in the latter model. Furthermore, due to being founded in an agent-based model all parameters of the VS model are readily interpretable as the sensitivity μ of the agents to mispricing and the weighting τ of the running price average. In contrast, parameters in the GARCH model are motivated purely from statistical grounds and cannot easily be related to agent behaviors.

From a Bayesian perspective, Eqs. (2) and (3) correspond to the likelihood $p(\mathbf{x}|\boldsymbol{\theta})$, i.e., the conditional probability of the observed data given the model parameters. To complete the model density $p(\mathbf{x}, \boldsymbol{\theta})$, we need to specify a prior distribution on the parameters. The choice of a prior distribution is often considered as subjective (whereas the likelihood has an aura of objectivism). Arguably, from the perspective of modeling the observed data this distinction is of limited relevance. Instead, note that fixing the prior implicitly fixes a distribution on the data space, i.e., obtained as $p(\mathbf{x}) = \int p(\mathbf{x}, \boldsymbol{\theta})d\boldsymbol{\theta}$ by marginalizing over the parameters. A model can be considered as misspecified when it assigns very low probability to the actual observed data. In contrast, a good model should be able to generate similar data with reasonable probability. This viewpoint is in line with Gelman et al. (2017) who argue that the prior can only be understood in the context of the likelihood. Indeed, prior and likelihood act together in shaping the model and expressing our expectations of plausible data.

Here, we propose the use of (weakly) informative priors which take into account our knowledge about the role played by the parameters when generating data from the likelihood model. As an example, consider the parameter $\tau \in (0, 1)$ of Eq. (2). While it might be natural to simply assign a uniform prior,⁷ τ controls the time constant of

⁶ Simulating this approximate model shows that it produces similar price series with strong volatility clustering as the original model.

⁷ Note that uniform priors, especially on unbounded spaces, should not be considered as uninformative. On the one hand, an improper uniform prior, i.e., when it cannot be normalized, expresses a strong belief about extreme parameter values by assigning infinite probability mass to values above any finite threshold. On the other hand, they are not invariant under model reparameterization as shown in the above example.

the running price average, i.e.,

$$\begin{aligned} \langle p_t \rangle_\tau &= (1 - \tau)p_t + \tau \langle p_{t-1} \rangle_\tau \\ &= (1 - \tau)p_t + \tau ((1 - \tau)p_{t-1} + \tau \langle p_{t-2} \rangle_\tau) \\ &= (1 - \tau) \sum_{k=0}^\infty \tau^k p_{t-k} \end{aligned}$$

as $\tau^k \rightarrow 0$ for $k \rightarrow \infty$. Comparing this to an exponentially weighted average in continuous time, i.e.,

$$\langle p_t \rangle_\rho = \int_0^\infty \rho e^{-\rho s} p_{t-s} ds$$

with time constant ρ^{-1} and exponential weighting kernel $k(s) = \rho e^{-\rho s}$ of unit weight, i.e., $\int_0^\infty k(s) ds = 1$, we match τ^k with $e^{-\rho k}$, thus interpreting $-\frac{1}{\log \tau}$ as the time constant l of the running average. A Uniform(0, 1) prior on τ then corresponds to an Inv-Gamma(1, 1) on $l = -\frac{1}{\log \tau}$ putting considerable probability mass on very short time constants below 1 day. In order to obtain a more reasonable distribution, we parameterize the model directly in terms of the time constant l and give it an Inv-Gamma(2, 1000) prior with a mean of 1000 and unbounded variance. Similarly, μ has been given a Gamma(3, 0.03) which assigns more than 95% of its probability mass to the interval [20, 250]. Together, these priors inform the VS model to stay away from the boundary $\mu \rightarrow 0$ or $\tau \rightarrow 0$ where it becomes trivial, i.e., $\mathbb{P}(|S_t| = 1) \equiv 1$ and thus $\sigma_t \equiv \sigma_{\max}$. Accordingly, it cannot be expected to generate data exhibiting pronounced volatility clustering in this case, and indeed, in the original reference Vikram and Sinha (2011), prices were averaged over 10^4 time steps, which are well covered by the chosen prior. Implementing both models is straightforward in Stan, and the full code is provided as supplementary material.

3.2 Model by Franke & Westerhoff (FW)

Franke & Westerhoff have developed a series of models and have estimated them by moment matching (Franke and Westerhoff 2012, 2011). Here, we follow their presentation and introduce the DCA-HPM model in their terminology.

In the FW model, the market is populated with two types of agents, namely fundamental and chartist traders. The fraction of fundamental traders at time step t is denoted by $n_t^f \in [0, 1]$. The corresponding fraction of chartist traders is then given by $n_t^c = 1 - n_t^f$. The log price, denoted by p_t , adjusts to the average demand from fundamental d^f and chartist d^c traders as

$$p_t = p_{t-1} + \mu \left(n_{t-1}^f d_{t-1}^f + n_{t-1}^c d_{t-1}^c \right). \tag{4}$$

The demand is composed of a deterministic and stochastic component. It is assumed that fundamental traders react to mispricing, i.e., the difference between p_t and the

(known) fundamental price p^* , whereas chartist traders react to past price movement, i.e., $p_t - p_{t-1}$. According to Franke and Westerhoff (2012), the demand dynamics is modeled as

$$\begin{aligned}d_t^f &= \phi(p^* - p_t) + \epsilon_t^f \quad \epsilon_t^f \sim \mathcal{N}(0, \sigma_f^2) \\d_t^c &= \xi(p_t - p_{t-1}) + \epsilon_t^c \quad \epsilon_t^c \sim \mathcal{N}(0, \sigma_c^2)\end{aligned}$$

with parameters $\phi, \xi > 0$ specifying the sensitivity to price differences for the fundamental and chartist traders. Note that these demands are unobserved as only their weighted sum effects the price. While such a dynamics could be modeled by means of a stochastic latent state, in the present case it is possible to marginalize out the demand. As the sum of two normally distributed random variables is again normal, the combined demand gives rise to a stochastic model for the log return $r_t = p_t - p_{t-1}$

$$\begin{aligned}r_t &\sim \mathcal{N}\left(\mu\left(n_{t-1}^f \phi(p^* - p_{t-1}) + n_{t-1}^c \xi(p_{t-1} - p_{t-2})\right),\right. \\&\quad \left.\mu^2\left(\left(n_{t-1}^f\right)^2 \sigma_f^2 + \left(n_{t-1}^c\right)^2 \sigma_c^2\right)\right).\end{aligned}\tag{5}$$

The volatility $\sigma_t = \mu\sqrt{\left(n_{t-1}^f\right)^2 \sigma_f^2 + \left(n_{t-1}^c\right)^2 \sigma_c^2}$ now depends on the fraction of chartist vs fundamental traders and changes over time. Franke & Westerhoff (2012) call this structured stochastic volatility, in analogy with structural models in economics, as the parameters of the agent-based model are grounded in behavioral terms and therefore economically meaningful.

The model is then completed by an update equation for the fraction of traders in each group. Here, we consider the DCA–HPM specification of Franke and Westerhoff (2012) which is given by

$$n_t^f = \frac{1}{1 + e^{-\beta a_t - 1}}\tag{6}$$

$$\begin{aligned}n_t^c &= 1 - n_t^f \\a_t &= \alpha_0 + \alpha_n(n_t^f - n_t^c) + \alpha_p(p^* - p_t)^2.\end{aligned}\tag{7}$$

The parameter a_t denotes the relative attractiveness of the fundamental over the chartist strategy. It includes a general predisposition α_0 and herding $\alpha_n > 0$ as well as mispricing $\alpha_p > 0$ effects. We chose this specification for two reasons:

1. The discrete choice approach (DCA) of Eq. (6) leads to a smoothly differentiable model density. This eases the exploration of the posterior when sampling with the HMC algorithm.
2. The herding + predisposition + misalignment (HPM) specification for the attractiveness Eq. (7) can be computed without access to the actual demands d_t^f and d_t^c . This is not true for the other specifications of Franke and Westerhoff (2012) where the agent's wealth depends on previous demands which, in turn, leads to a

stochastic volatility model where (one of) the demands have to be modeled as a stochastic latent variable. For simplicity, we have not considered this complication in the present paper.

Overall, the model dynamics is fully specified by Eqs. (5), (6) and (7). The parameters of the model are given by $\theta^{FW} = (\mu, \phi, \sigma_f, \xi, \sigma_c, \beta, \alpha_0, \alpha_n, \alpha_p, p^*)$. Note that β and μ are redundant as they simply control the scale of $\alpha_0, \alpha_n, \alpha_p$ and $\xi, \phi, \sigma_f, \sigma_c$, respectively. Thus, throughout we fix them at $\beta = 1$ and $\mu = 0.01$ as in the simulation exercise of Franke and Westerhoff (2012).

When estimating the model on real stock returns below, we do not know the fundamental price. When simulating from the model or matching moments as in Franke and Westerhoff (2012), the fundamental price can be considered as fixed. Yet, when fitting the model to actual return data more flexibility in order to estimate reasonable values for the unobserved fundamental price is needed. Here, we consider two specifications of the model: First, as in the VS model presented above, the fundamental price is derived as an average of past prices. Note that in order to be faithful and comparable to the VS model, we compute a running average of past prices and not log prices. Secondly, following Lux (2018), we assume that the log fundamental price is time varying as a Brownian motion

$$p_t^* \sim \mathcal{N}(p_{t-1}^*, \sigma_*^2).$$

This not only introduces another parameter σ_* but also turns the model into a stochastic volatility model, i.e., the volatility σ_t now includes a stochastic component. To see this, note that σ_t depends on a_{t-2} via n_{t-1}^f and the attractiveness in turn includes the stochastic fundamental log price p_{t-2}^* . Thus, the fundamental price plays a similar role as the log volatility h_t in a classical discrete time stochastic volatility (SV) model (Kim et al. 1998)

$$\begin{aligned} h_t &\sim \mathcal{N}(\mu + \phi(h_{t-1} - \mu), \sigma_h) \\ r_t &\sim \mathcal{N}(0, e^{h_t/2}) \end{aligned} \tag{8}$$

that we include as a benchmark alongside the GARCH(1, 1) model. Again, in contrast to the purely phenomenological dynamics in Eq. (8) the parameters of the FW model are interpretable in terms of behavioral trades of agents. Furthermore, the model specification combines aspects of local and stochastic volatility in that its volatility σ_t depends on the random fundamental price as well as past prices via Eqs. (6) and (7).

Nevertheless, implementing the model in *Stan* is readily possible. As before, the full code of the FW model—for both specifications—is provided as a supplementary material. Interestingly, along the same lines a variant of the VS model with a random walk specification for the fundamental price can be defined. For comparison, we have also implemented this model using both specifications. Note that the time-varying log fundamental prices p_t^* do not appear as a T -dimensional vector (where T denotes the number of observed time steps) in the parameter block. Instead, we have used a non-centered parameterization, i.e., `p_star` is computed from `epsilon_star` as

a transformed parameter. Formally, we can express this as follows:

$$p_t^* = p_{t-1}^* + \sigma_* \epsilon_t^* \quad \text{where} \quad \epsilon_t^* \sim \mathcal{N}(0, 1) \quad (9)$$

instead of

$$p_t^* \sim \mathcal{N}(p_{t-1}^*, \sigma_*^2).$$

This is a standard example of a reparameterization which does not change the model, but helps when HMC sampling as the innovation parameters ϵ_t^* all have unit scale and are a priori independent, no matter which variance σ_*^2 is currently sampled.

Again, we complete the model with weakly informative priors for all parameters. As few insights are available about the proper choice of the attractiveness parameters α_0, α_n and α_p , we assign weakly informative priors, e.g., $\alpha_0 \sim \text{student_t}(5, 0, 1)$, which restrict the scale of the parameter yet, being heavy-tailed, allow substantially larger values.⁸ In case of the standard deviation parameters σ_f, σ_c and σ_* , we impose stronger priors and resort to the observed data to set the proper scale. While not being purely Bayesian, this choice restricts the model to reasonable scales accounting for the fact that volatility could be measured in arbitrary units, e.g., percent per year. Figure 1 illustrates prior predictive checks, i.e., sample data generated according to the model with parameters drawn from the prior, for the FW model using a moving average specification for the fundamental price.⁹ While the scale of the data appears well matched, the model only occasionally produces volatility clustering as pronounced as in the real data. Nevertheless, we found these priors effective in simulation studies as well as when estimating the model on stock data. Furthermore, Appendix C contains additional robustness checks confirming the above rationality behind our choice of priors.

3.3 Model by Alfarano, Lux & Wagner (ALW)

Alfarano, Lux & Wagner model a financial market populated by N chartist traders as well as additional fundamental traders. The excess demand from the population of fundamental traders is given by

$$ED_f = T_f(p_f - p) \quad (10)$$

with total trading volume T_f and log fundamental price p_f . As in the FW model, prices are exclusively considered in logarithmic terms, i.e., to simplify notation the log price is denoted by p .

Chartists traders are either in an optimistic or pessimistic state. Optimists are assumed to buy a certain amount T_c at each time step, while pessimists sell amount T_c .

⁸ In contrast, a normal prior distribution would impose much more information as values larger than several tens of standard deviations are essentially ruled out.

⁹ Prior predictive samples using the random walk specification appear very similar (not shown).

Denoting the number of optimistic traders as n , the market sentiment $x = 2\frac{n}{N} - 1 \in [-1, 1]$ is defined and the excess demand from chartist traders given as

$$ED_c = NT_c x. \tag{11}$$

Then, assuming a Walrasian pricing mechanism log market prices are adjusted as

$$\frac{dp}{dt} = \beta (ED_f + ED_c) \tag{12}$$

$$= \beta (T_f(p_{f,t} - p_t) + NT_c x_t) \tag{13}$$

Note that the FW model assumed a very similar log price adjustment, albeit formulated in discrete terms in Eq. (4).

Lux (2018) now assumes instantaneous price adjustment, i.e., $\beta \rightarrow \infty$, and the market is cleared by matching fundamental and chartist demand. In this case, the market price is found to be

$$p_t = p_{f,t} + \frac{NT_c}{T_f} x_t \tag{14}$$

and the corresponding returns can be expressed as

$$r_t = p_{f,t} - p_{f,t-1} + \frac{NT_c}{T_f} (x_t - x_{t-1}) = \sigma_f \epsilon_{f,t} + \frac{NT_c}{T_f} (x_t - x_{t-1})$$

where the last line follows when assuming a Brownian motion for the log fundamental price, i.e., $p_{f,t} = p_{f,t-1} + \sigma_f \epsilon_{f,t}$ with $\epsilon_{f,t} \sim \mathcal{N}(0, 1)$. From a statistical perspective, market returns are distributed as

$$r_t \sim \mathcal{N}\left(\frac{NT_c}{T_f} (x_t - x_{t-1}), \sigma_f^2\right). \tag{15}$$

In contrast to the FW model, this model does not exhibit stochastic volatility. Furthermore, we do not need to model the fundamental price. Indeed, assuming instantaneous price adjustment market prices cannot deviate persistently from the fundamental price and it becomes essentially an observed quantity (compare Eq. 14).

Now, the market sentiment changes according to a herding process originally proposed by Kirman. Here, each agent randomly switches from pessimistic to optimistic or vice versa with transition rates $\pi^+ = a + bn$ and $\pi^- = a + b(N - n)$, respectively. Parameter a expresses a general tendency to switch state, while b models the herding with the switching probability increasing in the size of the contrarian population. As detailed in Alfarano et al. (2008), these transition rates lead to a sentiment dynamics with nonvanishing fluctuations even in the limit of an infinite population. The sentiment dynamics is then governed by the following Langevin equation:

$$dX_t = -2aX_t dt + \sqrt{2b(1 - X_t^2)} dW_t. \tag{16}$$

In our implementation, the continuous time equation (Eq. (16)) has been Euler-discretized with $\Delta t = 1$ day, i.e., $x_t = -2ax_{t-1} + \sqrt{2b(1-x_t^2)} \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, 1)$, and then coupled to the observed returns via Eq. (15). As in Lux (2018), we have fixed $\frac{NT_c}{T_f}$ as one and assume weakly informative truncated standard normal priors on the remaining parameters a , b and σ_f . As illustrated in Fig. 1, the model produces return time series similar to the FW model, albeit at a somewhat larger scale. It appears that a prior favoring smaller values for α and β would be beneficial. The prior chosen here provides a compromise between such an informative prior and the uniform one suggested in Lux (2018) which with high probability exhibits almost independent returns without clustering (not shown). Again, the full Stan code for the ALW model is provided as a supplementary material.

4 Results

Here, we present estimation results of the above models, both on simulated and on real price data.

4.1 Simulation studies

In order to check our model implementation, we simulated the FW model with parameters as given in table 1 of Franke and Westerhoff (2012), i.e., $\mu = 0.01$, $\beta = 1$, $\phi = 0.12$, $\xi = 1.50$, $\alpha_0 = -0.327$, $\alpha_n = 1.79$, $\alpha_p = 18.43$, $\sigma_f = 0.758$, $\sigma_c = 2.087$ and p^* being a moving average of past prices with a length scale of $l = 300$. Then, we re-estimated the model parameters on the simulated price series of $T = 2000$ time steps shown in Fig. 2.

The resulting samples from the posterior distribution are shown in Fig. 3. Overall, we have run four chains starting from independent random initial conditions and drawn 400 samples from each, after discarding an initial transient of another 400 samples as warm-up. Compared to other studies, the number of samples appears very low, but the high quality of the samples is clearly visible in the trace plots. The model appears to have converged after just about 50 samples, and all chains produce almost uncorrelated samples from the same distribution. This is also confirmed by standard convergence diagnostics such as Gelman & Rubin's \hat{R} , which compares the variance between and within chains, or the number of effective samples, which is based on the sample autocorrelation (not shown). If desired, more samples can easily be drawn as the shown estimation runs in a few minutes on a standard laptop.

Figure 4 shows the resulting posterior distributions together with the true parameters that generated the data. We found that about at least 1000 observed prices are necessary for the posterior to reliably cover the true parameters. Interestingly, preliminary runs on considerably longer time series of 5000 observations suggest that posterior uncertainty reduces only slightly. This suggests that the model is rather flexible with different parameter settings giving rise to similar return dynamics. Accordingly, the information that observed returns provide about the underlying parameters is limited. Appendix B provides further discussion of this point in the context of the GARCH model.

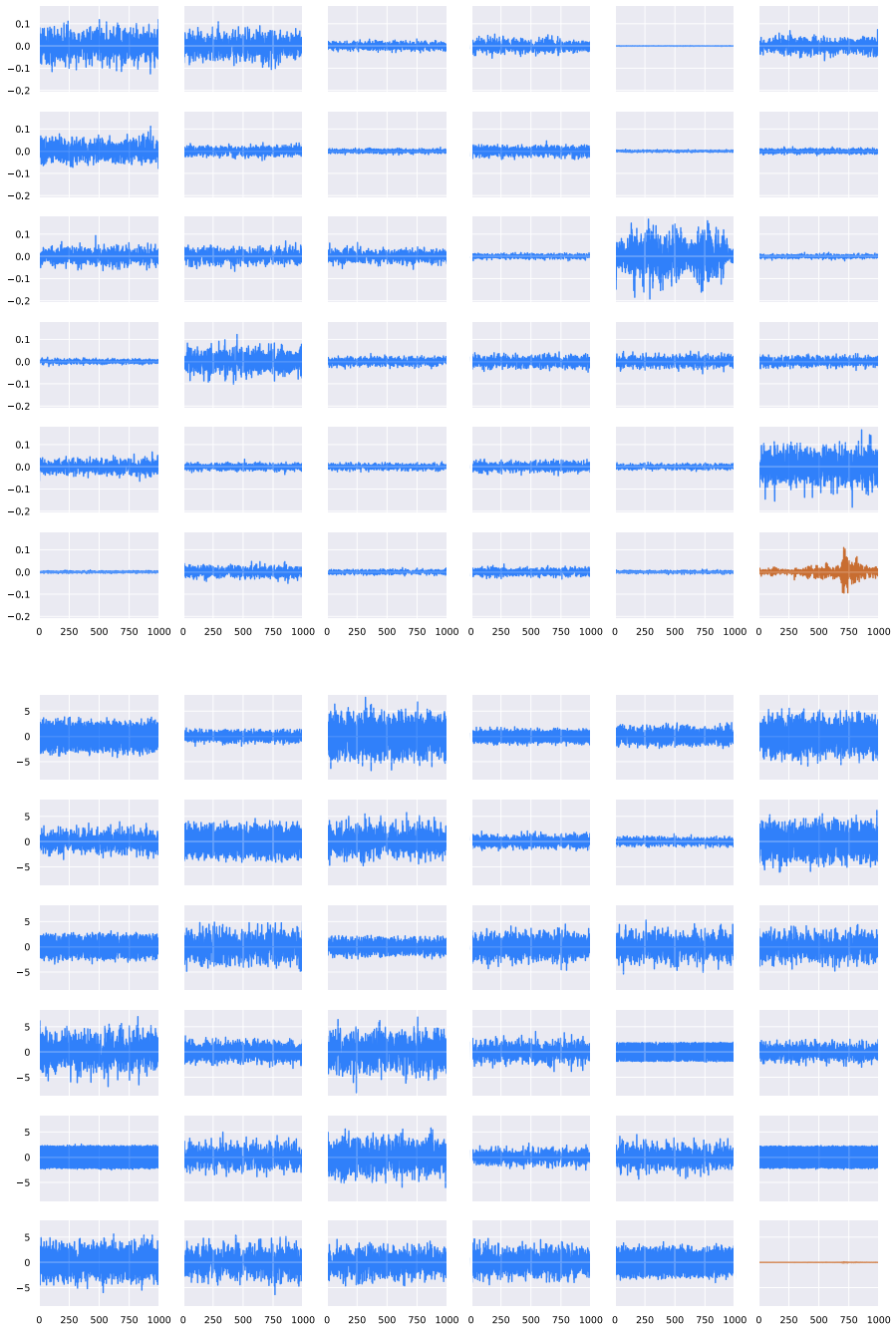


Fig. 1 Prior predictive checks, i.e., simulated model returns with parameters randomly drawn from the prior, for the FW (top) and ALW model (bottom). For comparison, an actual return time series is included in the lower right panel (note that the scale of returns generated by the ALW model is much larger than observed in actual data)

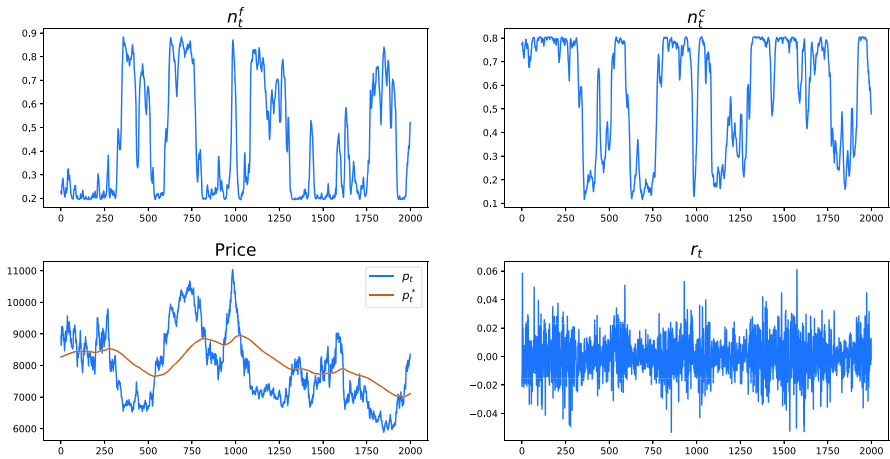


Fig. 2 Simulated price and return series of FW model. Note that a low fraction of fundamental traders n_f coincides with volatile market phases

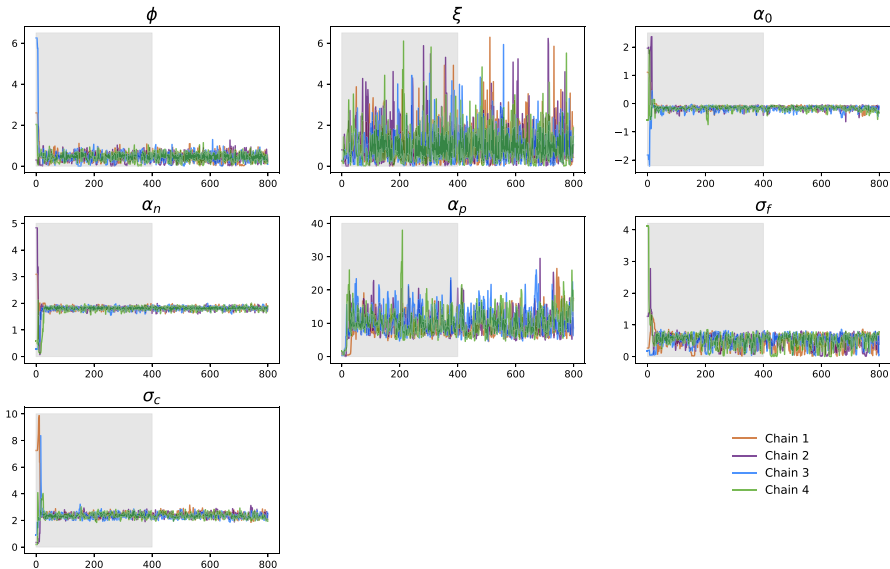


Fig. 3 Trace plot for model parameters ϕ , ξ , α_0 , α_n , α_p , σ_f and σ_c . Note that all chains appear to have converged to the same posterior distribution after just about 50 samples

Indeed, for the even more flexible models where the fundamental price is assumed to follow an (unobserved) random walk, some parameters could not be recovered at all from simulated data. Figure 5 shows the actual and inferred latent sentiment dynamics on data simulated from the ALW model. Estimates are shown as the posterior mean together with the 95% credibility bands around it. Here, only one chain has successfully recovered the actual sentiment time series, whereas another chain exhibits short excursions away from the actual sentiment. As the probability of both chains

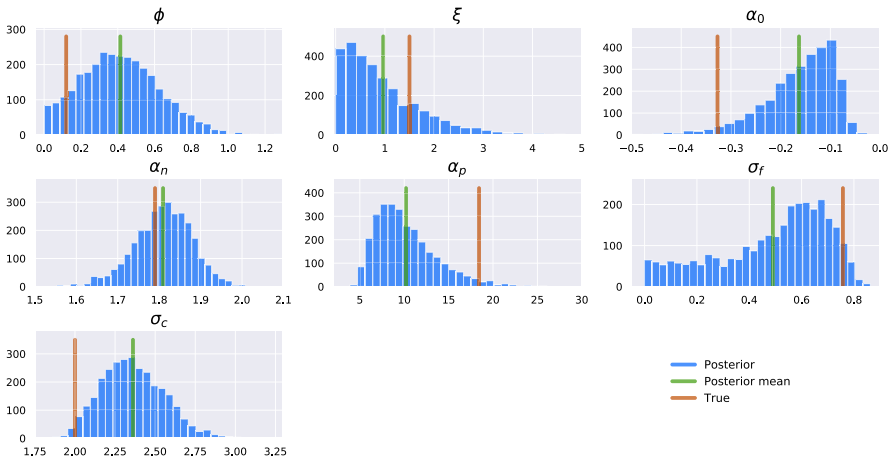


Fig. 4 Plot of posterior densities for parameters $\phi, \xi, \alpha_0, \alpha_n, \alpha_p, \sigma_f$ and σ_c . The true values are well covered by the posterior distributions

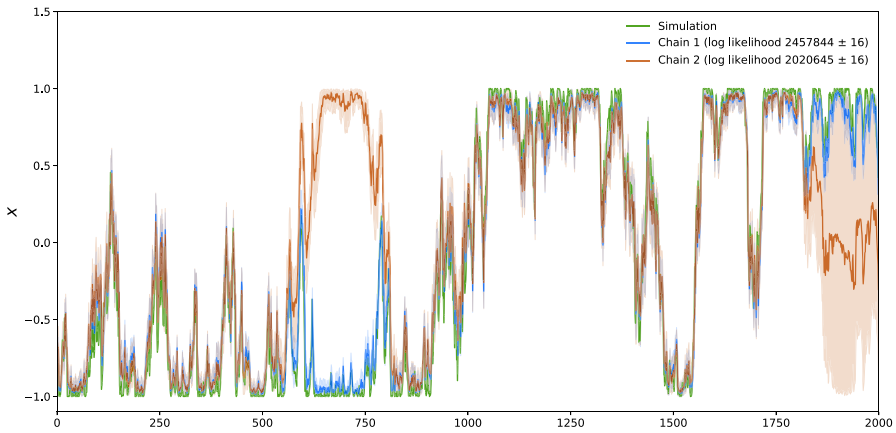


Fig. 5 Plot of market sentiment x_t over time. Chain 1 recovers the true sentiment dynamics of the simulated data, whereas chain 2 shows temporary deviations away from the true dynamics. Note that the likelihoods of these two posterior modes are markedly different

is vastly different, with the correct one being substantially higher, this appears to be a problem of the sampling algorithm which is getting stuck in a local mode of the posterior. It might well be that other algorithms, such as the sequential Monte Carlo methods employed by Lux (2018), are less susceptible to this problem. On the other hand, HMC is highly effective in sampling the global parameters of the model which is a major bottleneck for sequential Monte Carlo methods (Livingstone et al. 2019; Monnahan et al. 2017).

Furthermore, the FW model with the random walk specification for the fundamental price shows an even deeper non-identifiability. The top panel of Fig. 6 compares the estimated fraction of fundamental traders for two different chains of samples to the actual simulated time series. While one chain stays close to the actual values, the

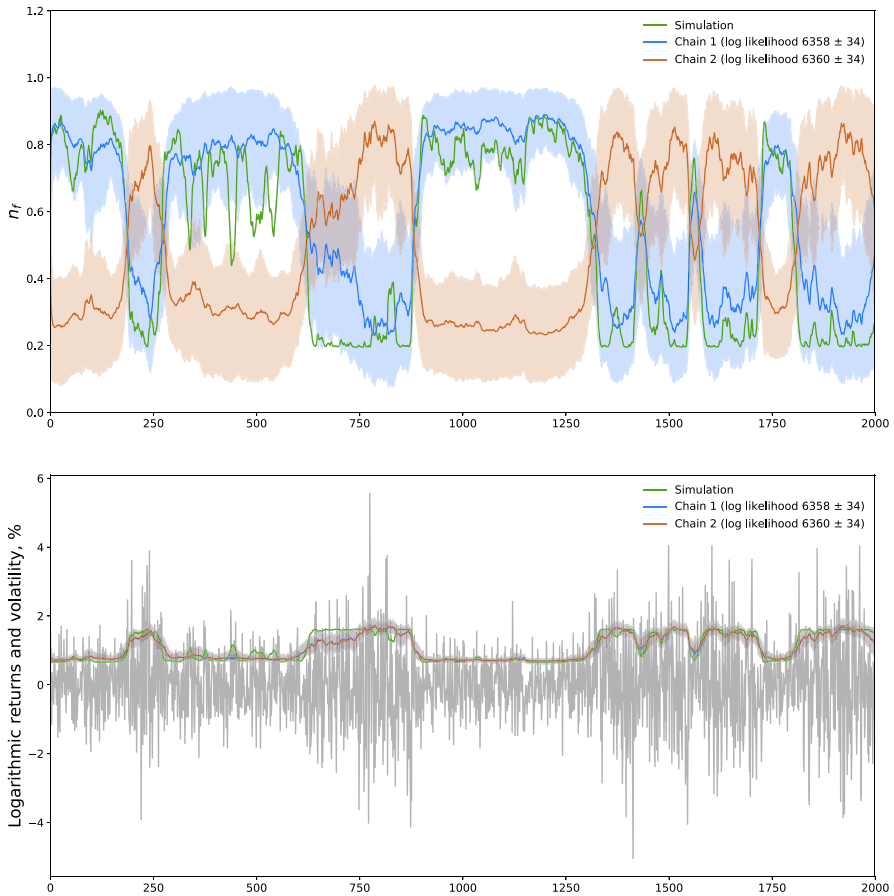


Fig. 6 Fraction of fundamental traders n_t^f (upper panel) and volatility estimates σ_t over time. Chain 1 recovers the true fraction of fundamental traders of the simulated data, whereas chain 2 shows a mirror image flipping the role of fundamental and chartist traders. Note that the likelihoods and estimated volatilities of these two posterior modes are essentially identical

other estimate appears to be a mirror image. Interestingly, in this case, the likelihood of both chains is almost identical. Indeed, the lower panel of Fig. 6 shows that the estimated volatilities are almost identical as well and correspondingly both estimates assign very similar probability to the observed returns. This suggests a multimodal posterior where each chain samples from a well-defined, yet different, mode of the distribution. Furthermore, in at least one of the modes we find that $\sigma_c < \sigma_f$!

Accordingly, the model offers two very different explanations for the observed volatility dynamics. In the first scenario, the number of chartists is usually low and rises sharply in volatile market phases (as intended by the model). In contrast, in the second scenario the number of chartists is usually high and drops in volatile market phases. Volatility is then driven by the high demand uncertainty of fundamental traders. Interestingly, both scenarios lead to very similar estimates and predictions for

the volatility σ_t . The model accomplishes this by assuming a very different trajectory for the unobserved Brownian motion of the fundamental price leading in turn to very different mispricings and demands of the fundamental traders. Thus, the seemingly innocuous assumption that the fundamental price follows a Brownian motion apparently introduces a symmetry into the model. This not only makes the unobserved latent states of the model unidentifiable, but also reveals that our understanding of agent-based models and their generated time series dynamics is far from complete. In particular, we cannot use the model in order to unambiguously uncover the fundamental price as in Majewski et al. (2018) where the Kalman filter computes a uni-modal posterior approximation. Furthermore, the herding parameters are hard to interpret as they act differently in different modes. Further work is needed in order to characterize and ideally remove this unidentifiability.

Thus, currently identification of some of the models discussed here is plagued by two major issues: on the one hand, of algorithmic nature with the sampler getting stuck in poor local optima of the posterior density; on the other hand, multimodality of the model likelihood leading to several posterior modes with vastly different interpretation for the model parameters. Here, we leave such investigations for future studies and instead focus on model predictions. In particular, a detailed comparison of the efficiency, e.g., in terms of effective sample size per wall clock time, and reliability of different algorithms for volatility models with complicated posterior geometries is beyond the scope of this paper. Instead, we focus on the models itself and continue with a principled model comparison in the next section.

4.2 Cross-validation and handling of missing data

To circumvent these identification issues, we now focus on predictions where identifiability is of no concern. In order to also circumvent the problem of the sampler getting stuck in a local minimum, we run several Markov chains for each model and only use the chain with the highest in-sample likelihood for predicting. Then, we compare models based on their probability assigned to held-out data. In the context of time series, models are commonly compared based on rolling look-ahead predictions, i.e., using the predictive distribution for future returns $r_{T+\tau}$ or volatilities $\sigma_{T+\tau}$ based on the T previous time points. Time is then rolled forward, and the model is evaluated again. In general, this approach requires refitting the model for each prediction.

Here, for computational reasons, we instead resort to leave-one-out (LOO) predictions, i.e., predicting current returns in the context of past and future returns, excluding the current one. With the method of Pareto smoothed importance sampling (PSIS), the corresponding predictive likelihoods

$$p(r_i | r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_T)$$

can be estimated from posterior samples (Vehtari et al. 2017). Note that in contrast to the full posterior, the LOO likelihood is conditioned on all but the i th data point. In order to estimate the LOO likelihood from posterior samples, the i th data points need to be effectively removed before evaluating the prediction. In general, this is far from

Fold	Data points									
1	<i>r</i> ₁	<i>r</i> ₂	<i>r</i> ₃	<i>r</i> ₄	r ₅	r ₆	<i>r</i> ₇	<i>r</i> ₈	<i>r</i> ₉	...
2	r ₁	<i>r</i> ₂	<i>r</i> ₃	<i>r</i> ₄	<i>r</i> ₅	r ₆	<i>r</i> ₇	<i>r</i> ₈	r ₉	...
3	r ₁	r ₂	<i>r</i> ₃	<i>r</i> ₄	<i>r</i> ₅	<i>r</i> ₆	r ₇	r ₈	<i>r</i> ₉	...

Fig. 7 Schematic of K -fold cross-validation for time series data ($K = 3$). Here, each model is fitted K times on the observed data points (bold) and evaluated in terms of predictions on the held-out set containing every K th data point (italic)

trivial and we refer the reader to Vehtari et al. (2017) for details about how PSIS estimates LOO likelihood. Here, we just note that the method smoothens the estimate by fitting a generalized Pareto distribution to the tail of observed log likelihoods. Thereby, a more stable estimate as well as a useful diagnostics is obtained. In particular, large tail exponents in the Pareto distribution suggest that estimates could have unbounded variance and should not be trusted. Indeed, this happens especially for influential data points and no reliable estimate can be obtained in this case.

At least for such data points, results need to be validated by direct estimation of the predictive likelihood, i.e., refitting the model on some of the data points and predicting the held-out ones. To this end, we also implemented a direct variant of K -fold cross-validation where the data are partitioned into K parts. LOO then corresponds to N -fold cross-validation which would require fitting the model N times on all but one data point and predict the left-out one. Here for computational reasons, we have chosen to hold out every 20th data point instead. The model is then fitted 20 times, on all but the remaining data points and asked to predict the unobserved ones. (Figure 7 shows the resulting schema for $K = 3$.) The choice $K = 20$, i.e., to hold out every 20th data point, provides a good compromise between computational performance (low K) and almost independent predictions (high K). Using information theory, it has been shown that volatility is only weakly dependent across a wide range of stochastic volatility models (Pfante and Bertschinger 2019). Thus, we can reasonably expect that volatilities are nearly independent after a couple of weeks. Numerical results in the next section confirm this assumption with estimates based on LOO and 20-fold cross-validation being statistically indistinguishable.

Accordingly, we have implemented all models in a way that allows for unobserved or missing returns. To this end, we pass in all returns r_1, \dots, r_T , observed and unobserved, together with a binary vector `miss_mask` indicating at each time whether the return is considered missing or not. From a Bayesian perspective, missing data points are simply unobserved latent variables that can be inferred together with other latent variables/parameters. Thus, for each missing return a corresponding parameter ϵ_{miss} is introduced and sampled alongside the other model parameters. To enhance sampling efficiency, the missing returns are represented in terms of innovations which are then transformed to actual returns via $r_t = \mu_t + \sigma_t \epsilon_{\text{miss}}$ akin to the non-centered parameterization explained in Eq. (9). As the model density is defined on r_t and not the underlying parameters ϵ_{miss} , we need to account for the resulting change of measure by multiplying the density with the absolute Jacobian $\frac{\partial r_t}{\partial \epsilon_{\text{miss}}}$ of the transformation.

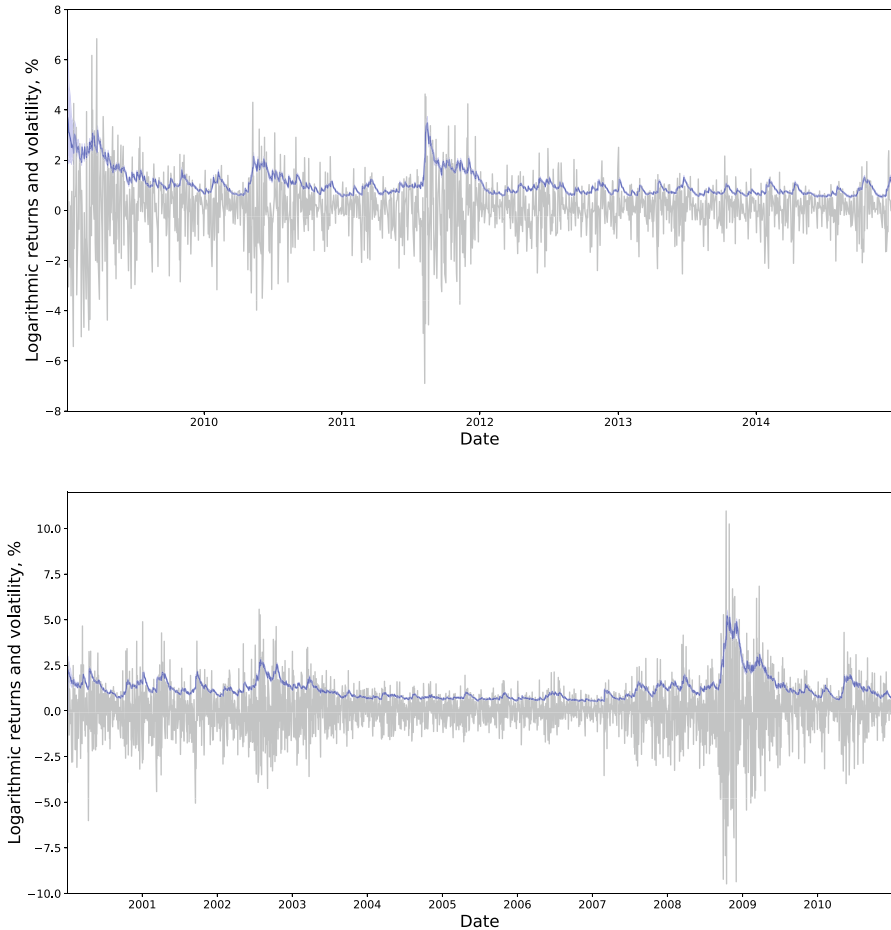


Fig. 8 GARCH model fit on the S&P 500

4.3 Fitting the S&P 500

Finally, we have fitted all models on price data from the S&P 500 stock market index. As a benchmark, a standard GARCH(1, 1) and SV model have been included for comparison. The corresponding fits from January 2009 to December 2014 and January 2000 to December 2010 are shown in Figs. 8, 9, 10, 11 and 12. The estimated model volatility is overlaid on the actual market returns.

Volatility estimates are shown as the posterior mean together with the 95% credibility bands around it. The posterior of the volatility σ_t at time step t is based on data points from returns r_t observed over all T data points, i.e., $p(\sigma_t|r_1, \dots, r_T)$ for $t = 1, \dots, T$. In the terminology of time series models, this is known as the smoothing distribution. Comparing the volatility estimates and predictions of the different models, a few remarks are in order:

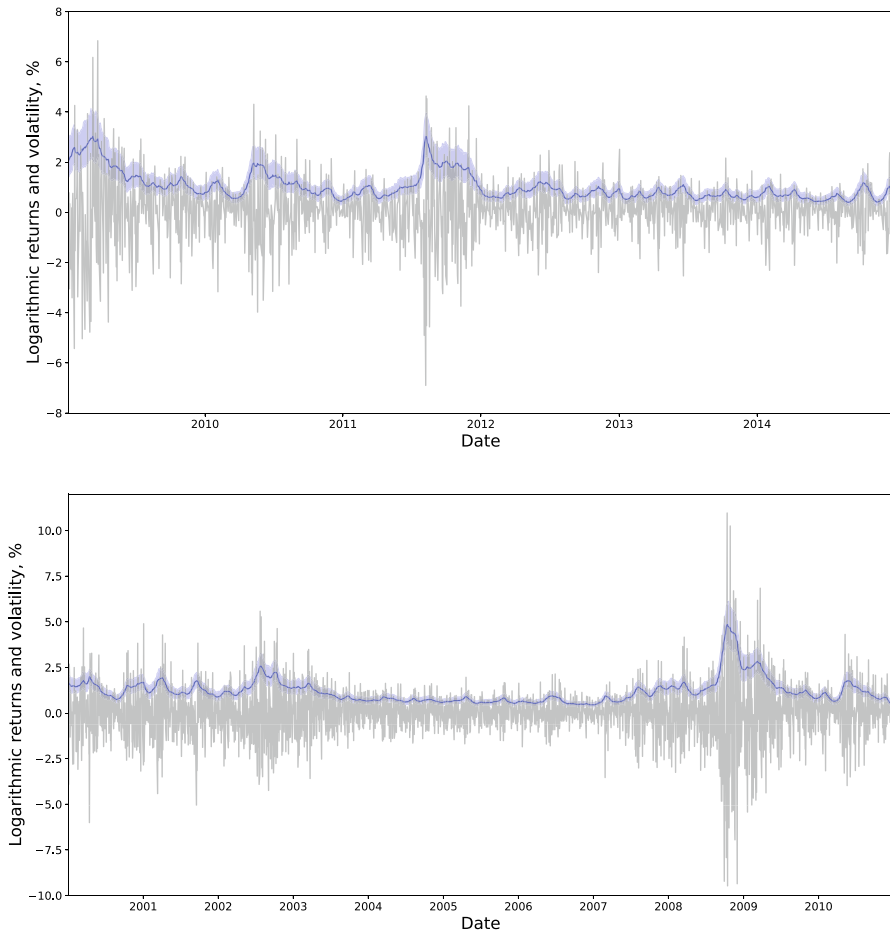
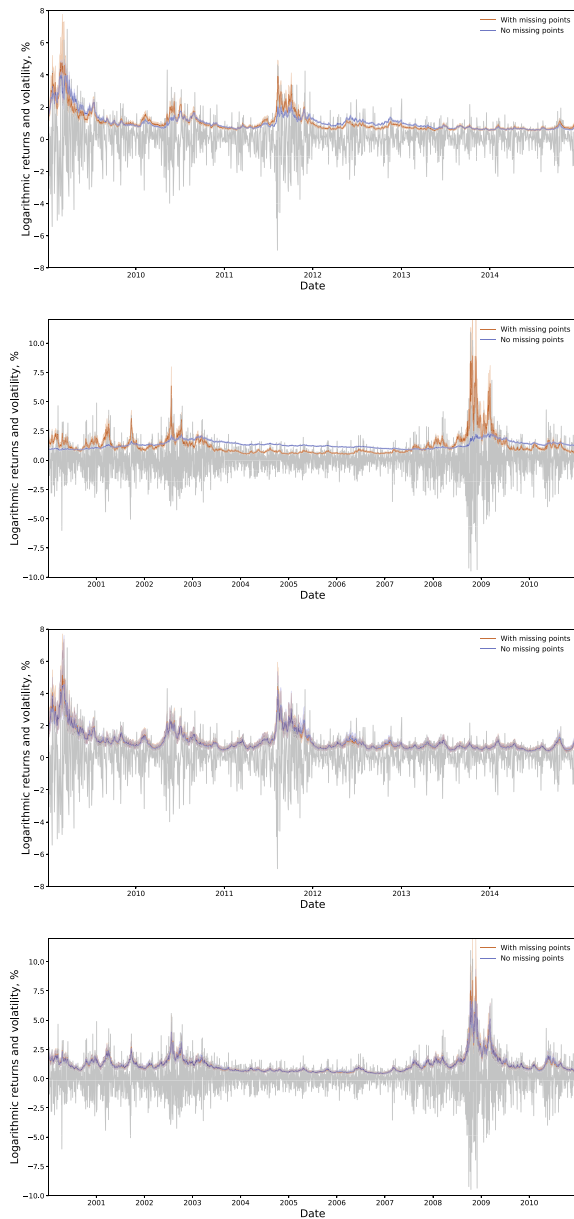


Fig. 9 SV model fit on the S&P 500

- The stochastic volatility models SV and VS, FW in random walk specification exhibit higher uncertainty in their volatility estimates as compared to models assuming a deterministic volatility dynamics, e.g., GARCH. Note that this does not imply that predictions are worse, but just reflects the intrinsic difficulty and imprecision of volatility estimation.
- The ALW model most closely matches the actual returns, exhibiting highly variable volatility estimates. Thereby, the model appears to overfit the actual return data.
- The fits of the VS and FW model with the moving average specification improves when some returns are unobserved, i.e., missing. We will discuss this issue in more detail below.

Having compared the models graphically, we proceed with a more principled assessment of model fit. In particular, Table 1 contains the predictive log likelihoods estimated using LOO and 20-fold cross-validation, respectively. As explained above,

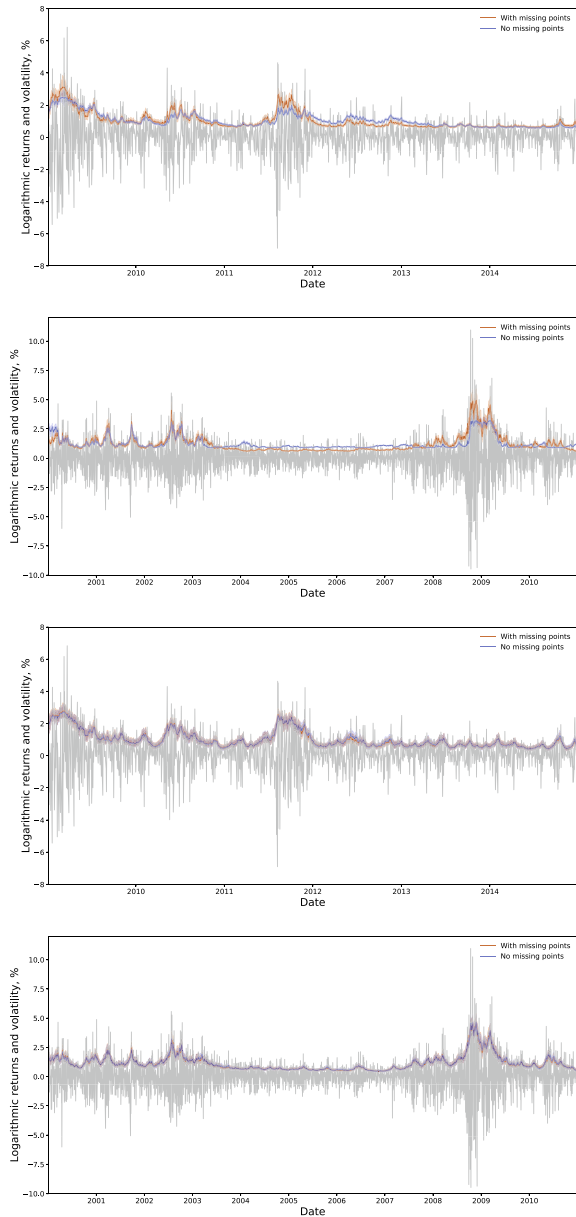
Fig. 10 VS model fit with moving average (top) and random walk (bottom) specification for the fundamental price on the S&P 500. Volatility estimates are shown with and without missing data. Note that the estimate with missing data is markedly better for the moving average specification for the fundamental price, whereas no difference is visible for the random walk specification



each model has been fitted using 8 randomly initialized chains and predictions are based on the best chain only. As expected, for the GARCH and SV model the estimates of predictive likelihood based on LOO and 20-fold CV are statistically indistinguishable. Similarly, the estimates for the VS and FW model using a random walk specification for the fundamental price are very similar.

For the other models, the situation is slightly different for several reasons:

Fig. 11 FW model fit with moving average (top) and random walk (bottom) specification for the fundamental price on the S&P 500. Volatility estimates are shown with and without missing data. Note that the estimate with missing data is markedly better for the moving average specification for the fundamental price, whereas no difference is visible for the random walk specification



- VS_ma, FW_ma: The models using the moving average specification for the fundamental price show better predictive log likelihoods when estimated with cross-validation as compared to LOO. This is somewhat surprising, as the LOO estimate is based on fitting all returns and then correcting the likelihood via PSIS which should lead to an overestimation, if anything.

The solution is already shown in Figs. 10 and 11 which show that the models with

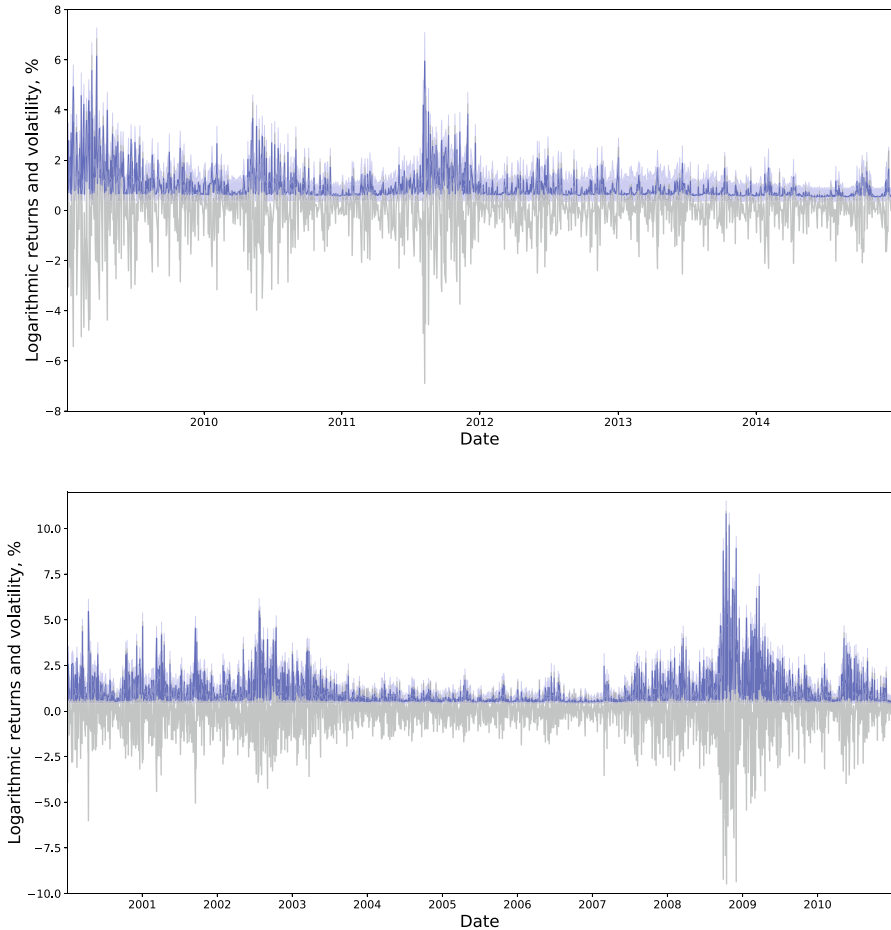


Fig. 12 ALW model fit on the S&P 500

Table 1 Model comparison based on leave-one-out (LOO) and cross-validation (CV) best chains predictive likelihoods

Model	Jan. 2009–Dec. 2014		Jan. 2000–Dec. 2010	
	LOO	CV	LOO	CV
GARCH	4867 ± 37	4871 ± 37	8573 ± 53	8577 ± 53
SV	4917 ± 36	4910 ± 37	8647 ± 48	8643 ± 48
VS_ma	4844 ± 40	4904 ± 35	8115 ± 69	8654 ± 46
VS_walk	4929 ± 37	4925 ± 36	8678 ± 48	8649 ± 48
FW_ma	4827 ± 40	4901 ± 35	8409 ± 56	8634 ± 47
FW_walk	4928 ± 36	4927 ± 36	8663 ± 48	8661 ± 48
ALW_walk	5561 ± 20	4930 ± 36	9936 ± 26	8067 ± 122

missing data are fitting the data better. Thus, it appears that the added flexibility when filling in missing data points is used in order to improve the model fits. From this perspective, the volatility dynamics of these models appears to be misspecified for the actual returns.

Note that no such difference is visible for the models with the random walk specification for the fundamental price.

- *ALW_walk*: In this case, the LOO estimates are not reliable and the PSIS method for LOO estimates warns about that accordingly. Thus, almost all data points are very influential as shown in Fig. 12, and the model fit would be substantially different if even a single data point were changed. Accordingly, the LOO estimates strongly overstate the actual cross-validation results and should not be trusted.

As the LOO estimates are unreliable in some case, we focus on the cross-validation results in the following. Tables 2 and 3 provide a more detailed pairwise comparisons between all models. Here, for each model combination the difference in total predictive log likelihood and the standard deviation of this difference is given. Note that mean differences are consistent with Table 1, but the standard deviations are computed specifically for each pairwise comparison. To this end, model predictions are compared pointwise for each data point predicted by both models to obtain mean and standard deviation of likelihood differences directly. For visual convenience, each column shows the advantage of the column model over the row model. The only models which are never significantly worse than any other model, i.e., with a likelihood difference of two standard deviations below zero, across both considered data periods are *FW_walk* and *VS_walk*. All other models are significantly outperformed by some model over some time period. In particular, the GARCH model is outperformed by every other model except for the ALW model on data including the crisis year of 2008. Thus, the best agent-based models beat this standard benchmark model and are on par with a simple SV model. This is encouraging as such models are routinely fitted to volatility data and shows that agent-based models provide viable alternatives with several advantages. First, in contrast to purely phenomenological SV models they have interpretable parameters related to trader behavior, allowing to uncover market sentiment or trader strategies over time. Secondly, they not only model the conditional volatility dynamics, but also provide non-trivial predictions for conditional returns. (Compare the return distributions of Eqs. (3) and (5) for instance.) We leave investigations to which extent conditional predictions of such models are able to identify trends and rallies in stock prices to future work.

To get an even better understanding of the relative performance between several models, Figs. 13 and 14 provide a running comparison between the best agent-based model, *FW_walk* and the SV as well as the ALW model, both within sample and on predicted data points. The top panel of Fig. 13 shows the cumulative advantage $\text{cum_adv}_t^{\text{FW/SV}}$ of the *FW_walk* model over the SV model within sample log likelihood, i.e.,

Table 2 Pairwise model comparison on 2009–2014 data: the advantage of the column model over the row model in cross-validation predictive likelihood with a standard error

	SV	VS_ma	VS_walk	FW_ma	FW_walk	ALW_walk
GARCH	39 ± 13	32 ± 12	53 ± 12	29 ± 12	56 ± 12	58 ± 21
SV		-6 ± 11	14 ± 3	-9 ± 12	16 ± 4	19 ± 17
VS_ma			20 ± 10	-3 ± 6	23 ± 9	25 ± 17
VS_walk				-23 ± 11	2 ± 4	4 ± 17
FW_ma					26 ± 9	28 ± 17
FW_walk						2 ± 16
ALW_walk						

Bold marks correspond to a significant advantage (difference more than two standard errors above zero) and italic marks to a significant disadvantage (difference more than two standard errors below zero)

Table 3 Pairwise model comparison on 2000–2010 data: the advantage of the column model over the row model in cross-validation predictive likelihood with a standard error

	SV	VS_ma	VS_walk	FW_ma	FW_walk	ALW_walk
GARCH	66 ± 17	76 ± 18	71 ± 18	56 ± 17	83 ± 16	-509 ± 107
SV		10 ± 13	5 ± 6	-9 ± 13	17 ± 5	-576 ± 106
VS_ma			-4 ± 12	-19 ± 7	7 ± 10	-586 ± 107
VS_walk				-14 ± 14	12 ± 7	-581 ± 106
FW_ma					26 ± 10	-566 ± 107
FW_walk						-593 ± 107
ALW_walk						

Bold marks correspond to a significant advantage (difference more than two standard errors above zero) and italic marks to a significant disadvantage (difference more than two standard errors below zero)

$$\text{cum_adv}_t^{FW/SV} = \sum_{i=1}^t \log p^{FW}(r_i|r_1, \dots, r_T) - \log p^{SV}(r_i|r_1, \dots, r_T).$$

The lower panel shows the corresponding cumulative log likelihood difference on predicted data points. The estimates of predictive log likelihood are obtained by combining data from all 20 folds, i.e., predictive likelihoods of time points 1, 21, ... are obtained from fold 1 which held out the corresponding returns, predictive likelihoods of time points 2, 22, ... are obtained from fold 2 and so on. As before, the mean difference is shown alongside with the 95% credibility band around it. The last value of the lower panel at time T corresponds to the total predictive log likelihood difference and

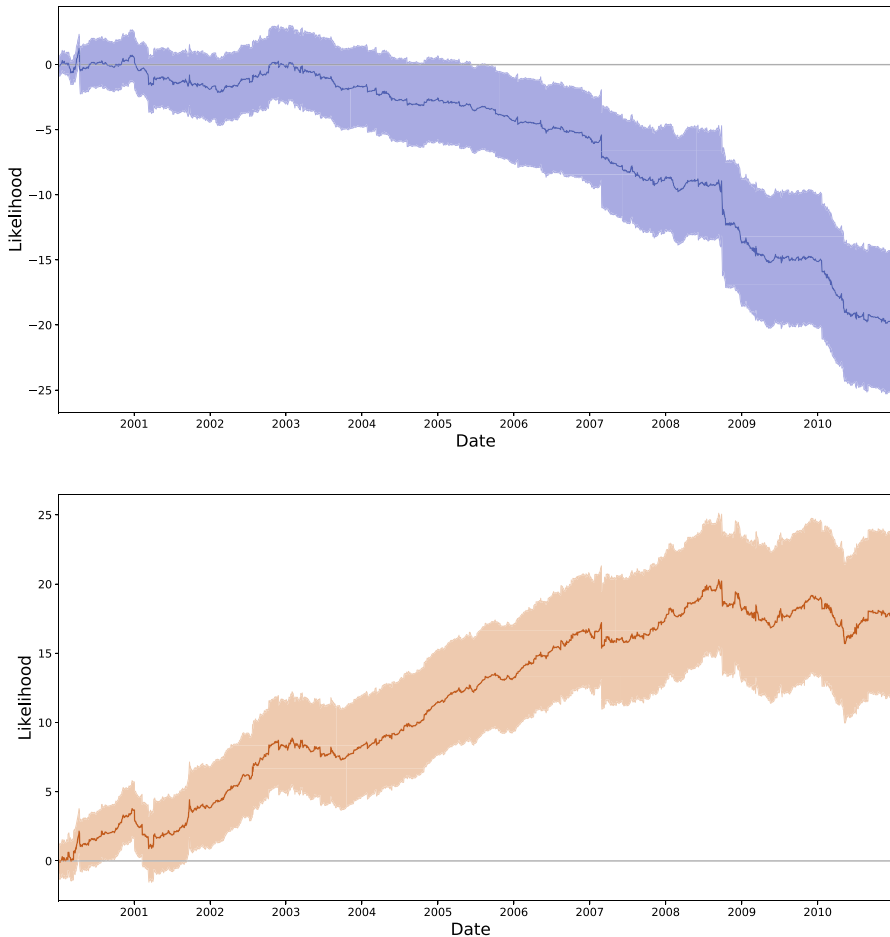


Fig. 13 Comparison between the FW model and the SV model in terms of the difference in cumulative log likelihood on observed (top) and held-out (bottom) returns. While the SV model is better in-sample, its out-of-sample predictions are actually worse than the FW model. Note that the final difference of 17 ± 5 in predictive log likelihood after all data points is reported in Table 3

is recorded in Table 3. Interestingly, the SV model is better in-sample with a negative cumulative advantage of the FW model of about -20 ± 5 . Out-of-sample, i.e., on the predicted data points, the situation is reversed with the FW model now being better for a total cumulative advantage of 17 ± 5 . Thus, the FW model is less flexible in fitting each data point in sample, yet providing better predictions. By Occam's razor, such a model should be preferred being both simpler—in statistical terms of effective parameters—yet more reasonable in terms of predicting the actual dynamics.

Figure 14 shows the corresponding comparison between the FW and the ALW model. Here, the strong overfitting of the ALW model is apparent with a huge advantage in-sample (upper panel), but much weaker predictions (lower panel). Interestingly, most of the predictive performance is lost during the year 2008 of the financial crisis.

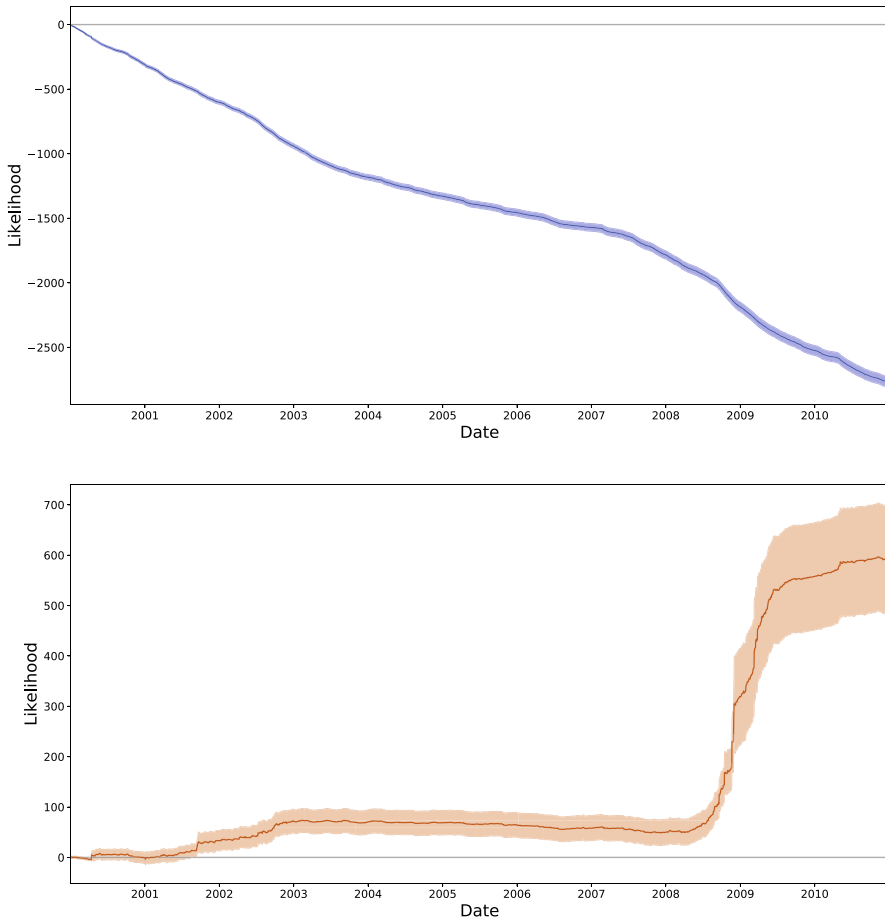


Fig. 14 Comparison between the FW model and the ALW model in terms of the difference in cumulative log likelihood on observed (top) and held-out (bottom) returns. The ALW models strongly overfits being much better in-sample and much worse out-of-sample. Particularly, during the crisis unfolding in 2008 the predictions of the FW model are substantially better. Again, the final difference of 593 ± 107 in predictive log likelihood is reported in Table 3

This suggests that the ALW model is unable to accurately reflect the unusual return dynamics unfolding during the financial crisis. Indeed, on the data sample from 2009 to 2014 excluding the critical year its predictive performance is on par with the other models as noted in the last column of Table 2. We take this as strong evidence that the assumption of instantaneous price adjustment is problematic and especially during the crisis persistent mispricing has to be assumed.

5 Concluding discussion

We have fitted a range of different agent-based models on actual return data from the S&P 500 stock index. To this end, we have formulated all models in statistical terms as continuous state-space time series models.¹⁰ Then, we have implemented several models in *Stan*, a modern probabilistic programming language for Bayesian modeling. The build in HMC sampling algorithms appears well suited to explore the posterior distributions arising in these agent-based models. Overall, HMC tends to mix rapidly and reveals much of the complicated posterior structure arising in some of the models, e.g., multimodality arising from non-identifiable latent state dynamics of the FW model with random walk specification for the fundamental price. Yet, further work on these identification issues is certainly required: on the one hand, to advance theoretical understanding of the approximate symmetry in the model likelihood leading to several modes with vastly different interpretation for the model parameters; on the other hand, to reliably detect and overcome algorithmic challenges when sampling from complicated multimodal posterior distributions.

Here, in order to circumvent these issues we focused on predictions instead. In particular, using cross-validation—LOO as well as 20-fold—we provided a systematic quantitative comparison of all models in terms of predictive log likelihoods. Our main findings are:

1. First, agent-based models are a viable alternative to more classical purely phenomenological econometric models. In particular, the best of the models tested clearly outperform a GARCH benchmark and are on par with a simple SV model.
2. Secondly, by comparing several models with different dynamics and assumptions we provide strong evidence that the fundamental price is more likely modeled as a random walk than a moving average. Furthermore, price adjustment needs to allow for persistent mispricing which is especially important in order to predict return dynamics during the latest financial crisis.
3. Model differences arising from different herding mechanisms appear minor compared to assumptions regarding the fundamental price dynamics and adjustment. In particular, we find no significant differences between predictions of the VS and FW model when using a random walk specification for the fundamental price. Thus, revealing the exact mechanism driving market sentiment or traders' expectations on empirical grounds alone appears difficult at best.¹¹

Compared to standard econometric models, agent-based models provide several advantages. They not only allow to estimate market sentiment or preferences in trading strategies alongside with fundamental prices, but also often include conditional price predictions as the FW and ALW model in this study. In the future, we plan to explore predictions from agent-based models in more detail, hoping that this provides the

¹⁰ Continuous state spaces are required for HMC sampling. Other algorithms, such as Gibbs sampling or particle filters, are not restricted in this respect. Yet, sampling from discrete spaces can mix prohibitively slow as commonly observed in mixture models.

¹¹ This is somewhat unfortunate but should be expected given the intrinsic uncertainty of volatility estimation (Pfante and Bertschinger 2019). Indeed, in econometric comparisons of stochastic volatility specifications, the best performing model often depends on the specific time period or stock ticker used for estimation.

information needed in order to infer the nature of the herding mechanism by empirical means.

Acknowledgements Open Access funding provided by Projekt DEAL. The authors thank Dr. H. C. Maucher for funding their positions. We also thank Damian Challet for helpful comments and suggesting the random walk variant of the VS model.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A The Stan language

The probabilistic programming language *Stan* (2017) allows the user to describe the joint probability $p(\mathbf{x}, \boldsymbol{\theta})$ of a model in a high-level programming language. In turn, the program is then compiled to C++ and several inference algorithms, including Hamiltonian Monte Carlo, are built-in. The required gradients are computed via a C++ library for automatic differentiation (Carpenter et al. 2015), thus freeing the user from manually implementing and debugging gradient calculations.

Stan comes with an extensive documentation which includes many example models and useful tricks for efficiently implementing them (Stan Development Team 2017). A minimal *Stan* program consists of three blocks

1. a **data** block which declares variables corresponding to observed quantities \mathbf{x} ,
2. a **parameters** block which declares variables corresponding to unobserved parameters $\boldsymbol{\theta}$ and
3. a **model** block which contains statements computing the log density of the model, i.e., $\log p(\mathbf{x}, \boldsymbol{\theta})$.

For instance, the following *Stan* program estimates the mean of Gaussian observations with a known standard deviation:

```

1 data {
2   int<lower=0> N; // number of data points
3   vector[N] x; // observed data points
4 }
5 parameters {
6   real mu; // unobserved mean
7 }
8 model {
9   mu ~ normal(0, 10); // computes log prior density log p(mu)
10  x ~ normal(mu, 1); // computes log likelihood log p(x | mu)
11 // total log density is log p(mu) + log p(x | mu)
12 }

```

The \sim statements in the model block relate a variable with a density and are shorthand notation for the more basic statements summing up log density contributions, e.g., `target += normal_lpdf(x | mu, 1)`.

As shown in the example, all variables are typed and need to be declared before use. A particularly convenient feature of *Stan* is that variables can be given constraint types, e.g., a standard deviation parameter could be declared as `real<lower=0> sigma`. Internally, the variable is then automatically transformed to an unbounded space and the log density is adjusted for the resulting change of measure. Due to this method, many different data types including vectors, matrices, but also constraint spaces such as simplices or covariance matrices are readily supported.

Stan supports several inference algorithms, namely gradient descent optimizers for maximum a posteriori estimation, HMC sampling and stochastic gradient variational Bayes. While HMC is the least efficient of these algorithms, it usually provides the closest approximation to the true posterior distribution. Furthermore, during warm-up, also known as burn-in, *Stan* adapts several parameters of the algorithm such that the algorithm appears essentially parameter-free to the user. This is especially effective for the No U-Turn Sampler (NUTS) which automatically adjusts the length of simulated trajectories. In a nutshell, NUTS integrates the Hamiltonian dynamics until it starts turning back toward itself which is locally decided based on the gradient direction. Care needs to be taken to ensure that the resulting transitions leave the target density invariant. To this end, trajectories are expanded in a treelike fashion by successively doubling their length forward and backward in time. The next state is then sampled uniformly from the resulting overall trajectory. For further details about the *Stan* programming language and the NUTS algorithm, we refer the interested reader to the *Stan* manual (Stan Development Team 2017) and Betancourt (2017), respectively.

B Parameter recovery in the GARCH model

Here, we provide additional recovery experiments for the GARCH model. We have chosen this model as it is well understood and posterior sampling poses no problems for *Stan*. Indeed, all chains quickly and reliably converge such that the samples drawn faithfully represent the true posterior distribution. Furthermore, as improper flat priors are used on α_0 , α_1 and β_1 their posterior just reflects the shape of the likelihood function. Furthermore, the standard normal priors on μ and σ_0 are only weakly informative considering the used parameter values.

For the experiment shown in Fig. 16, we have simulated the GARCH model with parameters $\mu = \frac{0.06}{250}$, $\alpha_0 = \frac{0.05^2}{250}$, $\alpha_1 = 0.1$, $\beta_1 = 0.9$ and $\sigma_0 = 0.1$ for 5000 time steps, i.e., trading days. We have then refitted the model on the first 2000 time steps as well as the last 2000 time points. Figure 15 shows the corresponding return series. In Fig. 16, the posterior densities together with the true parameters that generated the data and the posterior means are shown.

In all cases, the parameters are recovered in the sense that the true parameters fall well within a region of high posterior density. Yet, depending on the particular data set the posterior mean can substantially deviate from the true parameter. As the posterior

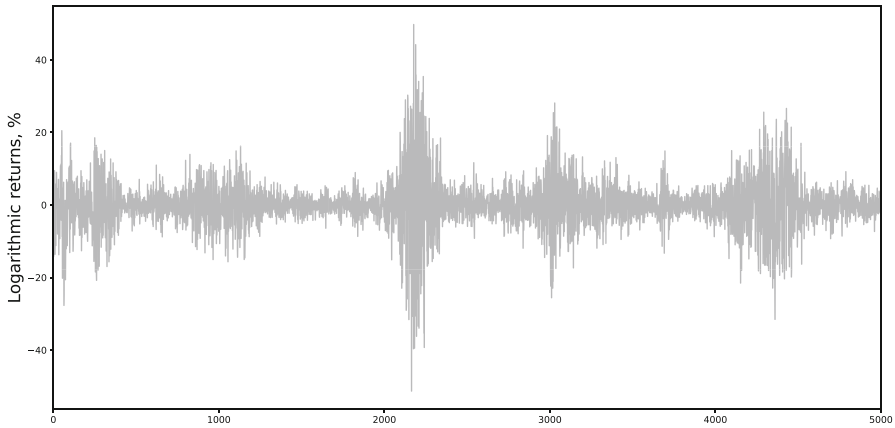


Fig. 15 Simulated returns from GARCH model with parameters $\mu = \frac{0.06}{250}$, $\alpha_0 = \frac{0.05^2}{250}$, $\alpha_1 = 0.1$, $\beta_1 = 0.9$ and $\sigma_0 = 0.1$. The simulated returns exhibit clear volatility clustering as well as realistic magnitudes

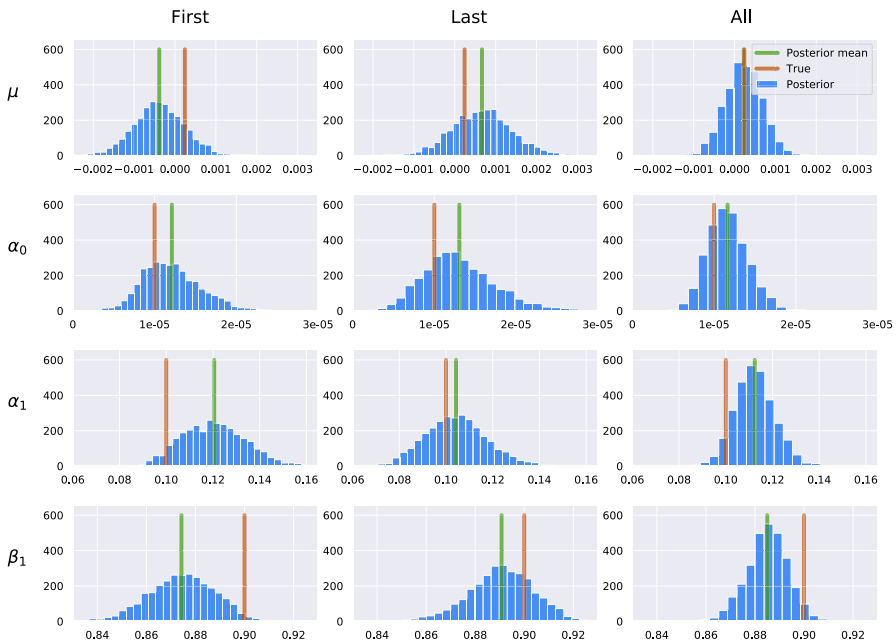


Fig. 16 Posterior distributions for the GARCH model fitted on the first 2000, last 2000 and all 5000 simulated returns of Fig. 15. Note that the true value is well covered by the posterior distribution even though the posterior mean can be substantially different

mean is known to minimize the expected squared distance from the true parameters, this implies that parameter estimates exhibit rather high variance in frequentist terms. It also shows that return data provide only limited amount about the underlying model parameters. For instance, especially the average return μ is known to exhibit high uncertainty and also in our example the interval $[-0.00065, 0.001]$ containing 95% of

Table 4 Model comparison for different priors based on leave-one-out (LOO) and cross-validation (CV) best chains predictive likelihoods

Model	LOO	CV
SV (original)	4917 ± 36	4910 ± 37
SV (flat)	4917 ± 36	4910 ± 37
FW_ma (original)	4827 ± 40	4901 ± 35
FW_ma (flat)	4761 ± 40	4830 ± 41
FW_ma (mixed)	4829 ± 39	4874 ± 36
FW_walk (original)	4928 ± 36	4927 ± 36
FW_walk (flat)	4981 ± 41	4885 ± 41

the posterior mass when estimated on all 5000 data points is not significantly different from zero. Furthermore, noting that the data were generated with μ corresponding to an average return of 6% per year, it corresponds to an estimated average yearly return between -16% and 25%.

C Prior robustness checks

Here, we investigate the robustness of model fits with respect to prior choices. As a benchmark, we refitted the SV model with improper uniform priors¹² on all parameters. Table 4 shows the LOO and CV predictive likelihoods on the S&P 500 returns between 2009 and 2014. There is no difference between the SV model with the weakly informative (original) and improper (flat) priors.

The corresponding results are also shown for the FW model, in the random walk and the moving average specification. In this case, the FW_walk model slightly overfits when using improper flat priors. This is also visible in Fig. 17 as the estimated volatility more closely tracks the return data under the flat prior. Table 4 confirms that the model predictions are still competitive with the SV model, but the LOO estimate becomes unreliable and deviates clearly from the CV estimate. Overall, the prior suggested in the main text has a weakly regularizing effect and smoothens the model estimates and predictions.

This effect is even more pronounced in the FW_ma model. In this case, the volatility estimate resulting from improper uniform priors is mostly constant and substantially worse (see Fig. 18 and Table 4). Indeed, we found that the informed priors on σ_f and σ_c as well as the boundary avoiding prior on the time constant l of the moving average are crucial. When these are combined with improper priors on all other parameters (mixed), volatility estimates and model predictions are similar to the results arising from priors (original) as suggested in the main text. When weakening any of the three

¹² Note that Stan transforms constrained parameters, e.g., $\sigma_h > 0$, to an unconstrained space. Thus, the improper uniform prior is imposed on $\log \sigma_h$ corresponding to $p(\sigma_h) \propto \frac{1}{\sigma_h}$ which is the Jeffreys prior for the standard deviation of a normal distribution with fixed mean. Furthermore, this prior is invariant with respect to multiplication with positive reals and thus imposes no information on the scale of σ_h .

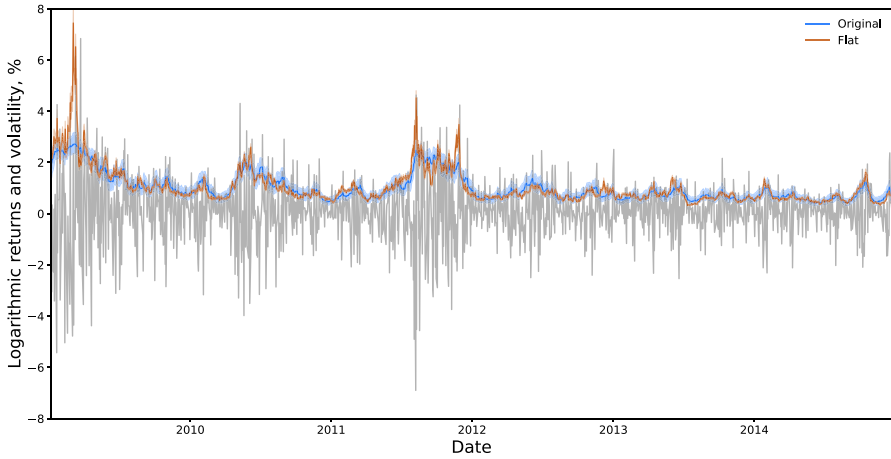


Fig. 17 FW_walk model fit on the S&P 500 (2009–2014) with different priors

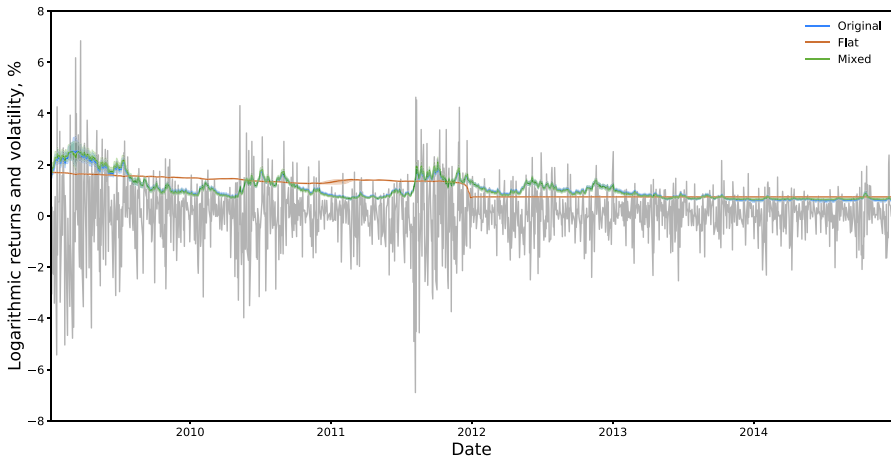


Fig. 18 FW_ma model fit on the S&P 500 (2009–2014) with different priors

priors to an improper one, the model essentially behaves as the one with flat priors on all parameters (not shown).

Overall, these experiments confirm the intuition that went into the choice of priors as explained in Sect. 3. In addition, a weakly regularizing effect is found in all cases and provides further justification for the use of Bayesian methods for model fitting, in particular leading to improved stability of volatility estimates and model predictions.

D Identifiability of the FW model

In Sect. 4.1, we observed two modes in the posterior of the FW model which lead to very similar predictions. Here, we show that under certain conditions this is indeed approximately the case.

In particular, we construct two sets of parameters $\sigma_f, \sigma_c, \alpha_0, \alpha_n, \alpha_p$ and $\tilde{\sigma}_f, \tilde{\sigma}_c, \tilde{\alpha}_0, \tilde{\alpha}_n, \tilde{\alpha}_p$ such that $\tilde{\sigma}_f = \sigma_c, \tilde{\sigma}_c = \sigma_f$ and the model volatilities are exactly equal, i.e.,

$$(n_{t-1}^f)^2 \sigma_f^2 + (n_{t-1}^c)^2 \sigma_c^2 = (\tilde{n}_{t-1}^f)^2 \tilde{\sigma}_f^2 + (\tilde{n}_{t-1}^c)^2 \tilde{\sigma}_c^2.$$

By $\tilde{\sigma}_f = \sigma_c, \tilde{\sigma}_c = \sigma_f$ this implies that $\tilde{n}_{t-1}^f = n_{t-1}^c, \tilde{n}_{t-1}^c = n_{t-1}^f$ corresponding to the observed flip in the fraction of chartist and fundamental traders.

From Eq. (6), we have

$$\begin{aligned} n_t^f &= \frac{1}{1 + e^{-\beta a_{t-1}}} \\ &= \tilde{n}_t^c \\ &= 1 - \tilde{n}_t^f \\ &= \frac{e^{-\beta \tilde{a}_{t-1}}}{1 + e^{-\beta \tilde{a}_{t-1}}} \\ &= \frac{1}{1 + e^{\beta \tilde{a}_{t-1}}}. \end{aligned}$$

Thus, we need to choose parameters such that $a_{t-1} = -\tilde{a}_{t-1}$.

Assuming that the fundamental price is always correct,¹³ i.e., $p_t^* = p_t$, we have

$$\begin{aligned} a_t &= \alpha_0 + \alpha_n(n_t^f - n_t^c) + \alpha_p(p_t^* - p_t)^2 \\ &= \alpha_0 + \alpha_n(\tilde{n}_t^c - \tilde{n}_t^f) \\ &= \alpha_0 - \alpha_n(\tilde{n}_t^f - \tilde{n}_t^c) \\ &= -\left(-\alpha_0 + \alpha_n(\tilde{n}_t^f - \tilde{n}_t^c) + \alpha_p(p_t^* - p_t)^2\right) \\ &= -\tilde{a}_t \end{aligned}$$

when choosing $\tilde{\alpha}_0 = -\alpha_0, \tilde{\alpha}_n = \alpha_n$ and $\tilde{\alpha}_p = \alpha_p$.

Note that the mean return

$$n_{t-1}^f \phi(p_t^* - p_{t-1}) + n_{t-1}^c \xi(p_{t-1} - p_{t-2})$$

can usually not be matched exactly in this fashion except when ϕ and ξ vanish. On the other hand, in stochastic volatility models the likelihood is commonly dominated

¹³ The actual symmetry observed in Sect. 4.1 is more complex with substantially different fundamental prices between the two modes. Our theoretical understanding is at best incomplete at this point.

by the volatility with only a weak effect of the mean return. Indeed, assuming a mean return of zero usually leads to almost identical fits and predictions of volatility compared with more elaborate models for the mean return. Numerically, we observe a similar result in the two posterior modes of the FW model, where the likelihood is almost identical, despite slightly different dynamics of the mean return in the two modes.

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
- An S, Schorfheide F (2007) Bayesian analysis of DSGE models. *Econom Rev* 26:113–172
- Barde S (2016) Direct comparison of agent-based models of herding in financial markets. *J Econom Dyn Control* 73:326–353
- Barde S (2017) A practical, accurate, information criterion for nth order Markov processes. *Comput Econ* 50:281–324
- Betancourt M (2017) A conceptual introduction to Hamiltonian Monte Carlo. ArXiv e-prints
- Betancourt M, Girolami M (2015) Hamiltonian Monte Carlo for hierarchical models. In: Upadhyay SK, Singh U, Dey DK, Loganathan A (eds) *Current trends in Bayesian methodology with applications*, 1st edn. Chapman and Hall/CRC, Boca Raton
- Bishop CM (2011) *Pattern recognition and machine learning*. Information science and statistics. Springer, Berlin
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31(3):307–327
- Carpenter B, Hoffman MD, Brubaker M, Lee D, Li P, Betancourt M (2015) The Stan Math Library: reverse-mode automatic differentiation in C++. *J Nano* 9:9
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: a probabilistic programming language. *J Stat Softw* 76(1):1–32
- Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quant Finance* 1(2):223–236
- Franke R, Westerhoff F (2011) Estimation of a structural stochastic volatility model of asset pricing. *Comput Econ* 38(1):53–83
- Franke R, Westerhoff F (2012) Structural stochastic volatility in asset pricing dynamics: estimation and model contest. *J Econ Dyn Control* 36(8):1193–1211
- Gelman A, Simpson D, Betancourt M (2017) The prior can often only be understood in the context of the likelihood. *Entropy* 19(10):555
- Ghoshgadze 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 Empir Finance* 37:1–19
- Grazzini J, Richiardi M, Tsionas M (2017) Bayesian estimation of agent-based models. *J Econ Dyn Control* 77:26–47
- Guerini M, Moneta A (2017) A method for agent-based models validation. *J Econ Dyn Control* 82:125–141
- Kim S, Shephard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with arch models. *Rev Econ Stud* 65(3):361–393
- Kukacka J, Barunik J (2017) Estimation of financial agent-based models with simulated maximum likelihood. *J Econ Dyn Control* 85:21–45
- Lamperti F (2018) An information theoretic criterion for empirical validation of simulation models. *Econom Stat* 5:83–106
- LeBaron B (2000) Agent-based computational finance: suggested readings and early research. *J Econ Dyn Control* 24(5):679–702
- Livingstone S, Betancourt M, Byrne S, Girolami M (2019) On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli* 25(4A):3109–3138
- Lux T (2009) Stochastic behavioral asset-pricing models and the stylized facts. In: Hens T, Schenk-Hoppé KR (eds) *Handbook of financial markets: dynamics and evolution*, handbooks in finance. North-Holland, San Diego, pp 161–215

- Lux T (2018) Estimation of agent-based models using sequential Monte Carlo methods. *J Econ Dyn Control* 91:391–408
- Ma YA, Chen T, Fox E (2015) A complete recipe for stochastic gradient MCMC. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (eds) *Advances in neural information processing systems*, vol 28. Curran Associates, Inc., Red Hook, pp 2917–2925
- Majewski A, Ciliberti S, Bouchaud JP (2018) Co-existence of trend and value in financial markets: estimating an extended Chiarella model. arXiv e-prints [arXiv:1807.11751](https://arxiv.org/abs/1807.11751)
- Monahan CC, Thorson JT, Branch TA (2017) Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods Ecol Evol* 8(3):339–348. <https://doi.org/10.1111/2041-210X.12681>
- Pfante O, Bertschinger N (2019) Volatility inference and return dependencies in stochastic volatility models. *Int J Theor Appl Finance* 22(03):1950013
- Samanidou E, Zschischang E, Stauffer D, Lux T (2007) Agent-based models of financial markets. *Rep Prog Phys* 70:409–450
- Shiller RJ (1980) Do stock prices move too much to be justified by subsequent changes in dividends? Working Paper 456, National Bureau of Economic Research
- Stan Development Team (2017) Stan modeling language users guide and reference manual. <http://mc-stan.org>. Version 2.17.0
- Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27(5):1413–1432
- Vikram SV, Sinha S (2011) Emergence of universal scaling in financial markets from mean-field dynamics. *Phys Rev E* 83:016101

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.