Journal of Economic Behavior and Organization 212 (2023) 366–391

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Moment set selection for the SMM using simple machine learning

Eric Zila [a,b], Jiri Kukacka [a,b,*]

[a] Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, 182 00 Prague 8, Czechia
[b] Charles University, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00 Prague 1, Czechia

ARTICLE INFO

Article history:
Received 14 April 2022
Revised 17 May 2023
Accepted 26 May 2023

JEL classification:
C13
C15
C22
C51
C58
G40

Keywords:
Agent-based model
Machine learning
Simulated method of moments
Stepwise selection

ABSTRACT

This paper addresses the moment selection issue of the simulated method of moments, an estimation technique commonly applied to intractable agent-based models. We develop a simple machine learning extension reducing arbitrariness and automating the moment choice. Two algorithms are proposed: backward stepwise moment elimination and forward stepwise moment selection. The methodology is tested using simulations on a Markov-switching multifractal framework and two popular financial agent-based models with increasing complexity. We find that both algorithms can identify multiple moment sets that outperform all benchmark sets. Moreover, we achieve considerable in-sample estimation precision gains of up to 66 percent for agent-based models. Finally, an out-of-sample empirical exercise with S&P 500 data strongly supports the practical applicability of our methodology as the estimated models pass the validity test of overidentifying restrictions.

1. Introduction

Since the 1960s, researchers have become increasingly aware of the stylized facts in empirical financial data. The absence of autocorrelation in returns, clustered volatility, heavy tails, and the emergence of speculative bubbles are among the essential patterns identified in data across all types of financial assets (Cont, 2001; Lux, 2009). In the search for an explanation of these empirical regularities, the field of finance turned to commonly used asset pricing models. However, the traditional approaches have not succeeded in this respect (Hong and Stein, 1999).

Beginning with Day and Huang (1990), the field of financial agent-based modeling emerged to face this challenge. Agent-based models in which boundedly rational agents interact and respond to their environment can replicate the stylized empirical facts and explain them endogenously (Cont, 2007; Chen et al., 2012). These models have attracted considerable attention from many economists in reaction to the global financial crisis between late 2007 and 2009 as mainstream approaches that assume perfect rationality and a single representative agent could not predict the looming collapse of the financial system.

* Corresponding author at: Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, 182 00 Prague 8, Czechia.

E-mail addresses: 10026254@fsv.cuni.cz (E. Zila), jiri.kukacka@fsv.cuni.cz (J. Kukacka).

https://doi.org/10.1016/j.jebo.2023.05.040
0167-2681/

The main shortcoming of the field is the empirical estimation (Lux and Zwinkels, 2018; Fagiolo et al., 2019). Due to the complexity of models and latency of some variables, standard analytical methods, such as the generalized method of moments (GMM) or maximum likelihood, often cease to be viable options. As closed-form solutions for more complex models often do not exist, computationally demanding simulation-based approaches are required. The most prominent alternative is the simulated method of moments (SMM) originally proposed by McFadden (1989) and Pakes and Pollard (1989). Its main advantages are a straightforward and transparent application without restrictive theoretical assumptions, customizability to various models, and well-defined mathematical properties.

In essence, the SMM repeatedly simulates the estimated model and calculates a set of carefully chosen characteristics of its output, such as the sample mean and variance or the first-order autocorrelation of returns. These characteristics can be associated with moments of the data-generating processes. The necessity of specifying the moments and the potential arbitrariness of this choice are among the most concerning problems of the method because the moment selection can markedly affect its estimation performance (Franke, 2009; Platt, 2020). The moment set is typically 'hand-crafted' intuitively based on expert knowledge of the model, following a deep study of its dynamics. However, no guarantee of the optimality of a moment set constructed in this manner can be provided. Some crucial moments may be overlooked due to limited knowledge of the model dynamics or arbitrary decisions of a researcher. Conversely, an overabundance of moments might lead to difficulties in the estimation process.

To reduce the arbitrariness of the moment choice, we propose a novel automated approach to the moment set selection. It is a straightforward machine learning extension applied on top of the SMM based on the stepwise selection framework. Although stepwise regression is outdated as a model specification tool, this simple technique addresses the problem very well. Our method performs moment selection similarly, in essence, to Lasso, a regularization approach used in statistical modeling and regression analysis, which performs a kind of continuous subset selection (Tibshirani, 1996; Hastie et al., 2009). We thus introduce elementary machine learning feature selection principles to agent-based econometrics, our paper's main methodological contribution. Subsequently, the identified moment sets can be used for empirical exercises and model validation.

After introducing our method, we study its performance using three models with increasing complexity and numbers of estimated parameters. These include the Markov-switching multifractal (MSM) model by Calvet and Fisher (2004), the popular financial agent-based model of herding by Alfarano et al. (2008), and the Franke and Westerhoff (2012) model, which is a more complex financial agent-based model for studying interactions of fundamental and speculative traders in an asset market. The two agent-based models are standardly estimated using the SMM in practice. Therefore, our methodology also has a direct practical implication, especially for the Franke and Westerhoff (2012) model, for which no closed-form solution exists, and standard estimation approaches are thus infeasible.

The first simplified application of the SMM to an agent-based model of financial markets can be traced to Gilli and Winker (2003) applying it to the Kirman (1993) model. Complete implementation of the method is then presented in Winker et al. (2007) and Franke (2009), who applies an alternative version of the method to the Manzan and Westerhoff (2007) model. Many other estimation attempts have followed these initial contributions. For instance, the SMM finds its use in a series of papers starting with the initial model presentation in Franke and Westerhoff (2011), its expansion and generalization in Franke and Westerhoff (2012), and detailed exploration in Franke and Westerhoff (2016). Schmitt and Westerhoff (2017a,b) further advance the guidance on an appropriate moment matching. Other examples of papers studying the application of the SMM to financial agent-based models are Fabretti (2013), who applies this method to the Farmer and Joshi (2002) model, Grazzini and Richiardi (2015); Chen and Lux (2018); Tubbenhauer et al. (2021), and Schwartz and Kirstein (2022), who study ergodicity of moment conditions; we refer to recent excellent surveys on financial agent-based modeling and econometrics by Dieci and He (2018); Lux and Zwinkels (2018) for an overview.

Recently, following a seminal paper by Salle and Yıldızoğlu (2014) on the kriging interpolation method, Lamperti et al. (2018); Bargigli et al. (2020); Zegadło (2021); Zhang et al. (2023); Chen and Desiderio (2022) are among the first to introduce machine learning methods to agent-based modeling and validation. However, they primarily aim at simplifying complex models, a process called meta-modeling or surrogate modeling. These approaches are based on intelligent/efficient sampling, e.g., for parameter space exploration and calibration purposes, not for methodological advancements of existing estimation methods. Platt (2022) then extends the approach of Grazzini et al. (2017) by applying deep neural networks to approximate the likelihood within a Bayesian estimation of the Brock and Hommes (1998) and Franke and Westerhoff (2012) models.

The paper proceeds as follows. The next Section 2 introduces the SMM and implementation details. Section 3 then presents the proposed machine learning extension, the selection criterion, and the full moment set. In Section 4, the three testbed models, together with their parameterizations, are described, and Section 5 summarizes our numerical setup and necessary pre-estimation analyses. Next, Section 6 presents a Monte Carlo simulation analysis and quantifies performance improvement gains. The main results are directly applied in the empirical Section 7. In Section 8, we summarize the most robust findings and properties observed across all testbed models. Finally, Section 9 concludes the paper. Additional results are relegated to the Appendix.

2. Simulated method of moments

The simulated method of moments (SMM, also referred to as the method of simulated moments) was initially introduced by McFadden (1989) and Pakes and Pollard (1989) as a modification of the GMM. Further, it was expanded by Lee and Ingram (1991) and Duffie and Singleton (1993) by describing conditions for asymptotic normality and consistency and utilizing the method for estimating asset pricing models. The SMM generally enables the estimation of analytically intractable time series models, such as complex agent-based models often containing latent variables. Its objective is to match selected moments of the simulated data as closely as possible to their empirical counterparts by optimizing model parameters. In principle, the method yields parameter estimates by minimizing the weighted sum of distances between the calculated statistics. Confidence intervals for single parameters are then obtained by approximating the estimator's distribution via Monte Carlo simulations according to Franke and Westerhoff (2011, 2016).

2.1. Formal definition

Consider an empirical time series $\{y_t\}_{t=1}^{T_{emp}}$ stored in the form of a column vector $\mathbf{y}^{emp}$. We can define a function $m(\mathbf{y})$, which computes the sample counterpart of a moment m given a time-series vector $\mathbf{y}$. Assume we choose a total of D moments of interest and denote their sample counterpart functions as $m_d(\mathbf{y}),\ d = 1, \ldots, D$. Then, we can form an empirical moment vector $\mathbf{m}^{emp} = [m_1(\mathbf{y}^{emp}), \ldots, m_D(\mathbf{y}^{emp})]^T$.

Suppose we have a fully specified, stochastic model $f(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters of the model. Assume we can use the model to generate a simulated time series $\mathbf{y}^{sim}(\boldsymbol{\theta})$ of length $T_{sim} \geq T_{emp}$. To moderate the effect of randomness when calculating the moment vector of simulated data, we use multiple random draws of the stochastic part of the model, each resulting in one simulated time series vector. We can organize these vectors into a matrix $\mathbf{Y}^{sim}(\boldsymbol{\theta}) = \left[\mathbf{y}_1^{sim}(\boldsymbol{\theta}), \ldots, \mathbf{y}_N^{sim}(\boldsymbol{\theta})\right]$, where N is the total number of simulated time series and $\mathbf{y}_n^{sim}(\boldsymbol{\theta})$ is the nth simulated time series vector. $\mathbf{Y}^{sim}(\boldsymbol{\theta})$ is then a $T_{sim} \times N$ matrix. Subsequently, we can define a function for computing the simulated sample counterpart of a moment for each of the D moments of interest as follows:[1]

$$m_d^{sim}(\mathbf{Y}^{sim}(\boldsymbol{\theta})) = \frac{1}{N}\sum_{n=1}^{N} m_d(\mathbf{y}_n^{sim}(\boldsymbol{\theta})). \tag{1}$$

Then, we can form a simulated moment vector $\mathbf{m}^{sim}(\boldsymbol{\theta}) = \left[m_1^{sim}(\mathbf{Y}^{sim}(\boldsymbol{\theta})), \ldots, m_D^{sim}(\mathbf{Y}^{sim}(\boldsymbol{\theta}))\right]^T$.

The process of matching moments is based on minimizing the distance between $\mathbf{m}^{emp}$ and $\mathbf{m}^{sim}(\boldsymbol{\theta})$. We can specify the objective function as follows:

$$J(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\theta})^T \mathbf{W} \mathbf{h}(\boldsymbol{\theta}), \tag{2}$$

where $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{m}^{emp} - \mathbf{m}^{sim}(\boldsymbol{\theta})$ is the difference between empirical and simulated moment vectors and $\mathbf{W} \in \mathbb{R}^{D \times D}$ is a positive definite weighting matrix. Then, the SMM estimator becomes:

$$\hat{\boldsymbol{\theta}} = arg\min_{\boldsymbol{\theta}} J\left(\boldsymbol{\theta}\right), \tag{3}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta}$ is the parameter space of the search. Finally, the J-test of overidentifying restrictions (Hansen and Singleton, 1982; Kocherlakota, 1990; Mao, 1990; Lee and Ingram, 1991) is typically utilized to statistically rigorously evaluate the joint compatibility of all D moments in situations when the SMM utilizes more moment conditions than there are K estimated parameters (Franke, 2009; Jang and Sacht, 2016):

$$\bar{J} = J(\hat{\boldsymbol{\theta}}) \xrightarrow{T \to \infty} \chi^2_{D-K}. \tag{4}$$

Under $H_0 : \bar{J} = 0$, assuming the model to be a valid approximation of the data-generating process concerning the given set of moments, the test statistic is asymptotically χ^2_{D-K} distributed. Rejection of H_0 indicates an invalid model that fails to replicate the distributional characteristics of empirical data in at least one dimension.

2.2. Weighting matrix

Many approaches exist to calculate the weighting matrix $\mathbf{W}$. In all cases, however, the aim is to assign greater importance to stable moments across realizations of the data-generating process. For a moment with high sampling variability, a higher difference between the empirical and simulated values of the moment should be considered less critical than if such a difference occurs for a moment with low variability. Additionally, the information about possible correlations between

[1] To compare the empirical moments to their simulated counterparts, we average the simulated moments out over all generated samples. Ideally, one could estimate each moment's distribution instead to evaluate the probability that the observed empirical moment values were generated by the corresponding distributions of the simulated moments. However, the employed simplification of the problem is standardly taken in the literature summarized in Section 1.

individual moments should also be considered. In consequence, one traditionally constructs the weighting matrix from the covariance matrix $\boldsymbol{\Sigma}$ of moments such that $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$.

Asymptotically, the Newey–West approach is an efficient and heteroskedasticity and autocorrelation consistent estimator of the long-run covariance matrix of the moments yielding the most efficient SMM estimator (Franke, 2009; Chen and Lux, 2018). The SMM is also asymptotically consistent provided that general regularity conditions hold (Lee and Ingram, 1991; Duffie and Singleton, 1993). However, Altonji and Segal (1996) show that this optimality does not carry over to small samples resulting in a biased SMM estimator. Even the identity matrix sometimes presents a better alternative to the Newey–West matrix, and the bias gets even worse for heavy-tailed-distributed data. Moreover, it does not decline with the increasing dimension of moments or the number of parameters. Instead, it can even worsen, resembling the 'curse of dimensionality.' In specific situations, the weighting matrix derived as an inverse of the covariance matrix might be even ill-behaved (Boivin and Giannoni, 2006). This phenomenon originates from the theory of the GMM for large samples (Hansen, 1982), implying that the variance of the optimal weighting matrix estimator increases in small samples with the number of moment conditions. One should thus try to conserve the number of moments used in the estimation to obtain desirable small sample properties. For larger dimensions of the weighting matrix, its identification becomes problematic, and its estimation is inefficient.

To avoid this problem, we follow Franke and Westerhoff (2012) and estimate the covariance matrix using a bootstrap approach that can better account for the finite-sample properties. Specifically, we bootstrap blocks of the empirical time series to construct new series of empirical returns from which the covariances of individual moments can be derived. Block bootstrap is preferred to individual observations because many moments traditionally used for the SMM estimation of financial models are long-range dependent. According to the authors, sample blocks of 250 observations are used for short-memory moments, such as mean, variance, and kth-order autocorrelations, $k \in \{1, 2, 3, 5\}$, and blocks of 750 observations for long-memory moments, namely lth-order autocorrelations, $l \in \{10, 15, 20, 25, 50, 100\}$. Blocks of higher length are utilized to reduce the joint-point problem of the block bootstrap. It corresponds to the situation where autocorrelations are, in part, calculated for points with an unrelated history caused by creating artificial links between randomly sampled blocks of observations. When generating new series, we randomly draw blocks from the empirical sample such that a new series of length T_{emp} is constructed.

As a slight improvement of the Franke and Westerhoff (2012) setup, we use overlapping blocks to reduce the small sample bias within the block bootstrap procedure. Under the Franke and Westerhoff (2012) bootstrap setup, the sample unit is a block (of either 250 or 750 observations), resulting in a small number of non-overlapping blocks prone to a similar small samples bias phenomenon as explained above regarding the GMM theory. The role of the estimation bias then grows considerably with the number of moments which perhaps does not cause complications to the authors, as they use only nine moments to estimate seven parameters. In our case, however, where we employ up to 22 moments in our machine learning procedure, the overlapping block bootstrap significantly increases the randomness that limits the impact of the small sample bias. For $T_{emp} = 6750$, we effectively sample from 6501 (6001) overlapping blocks of size 250 (750) instead of 27 (9) non-overlapping blocks allowing us to approximate the target distribution of the data generating process well. Moreover, in practice, we are not forced to cut off the last observations of the actual empirical time series that exceed a multiple of 250 (750).

Let us assume we bootstrap a total of B samples. Using each sample, we generate a total of B moment vectors $\mathbf{m}^b$, $b = 1, \ldots, B$. Then, we compute an estimate of the covariance matrix as follows:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{B}\sum_{b=1}^{B}(\mathbf{m}^b - \overline{\mathbf{m}})(\mathbf{m}^b - \overline{\mathbf{m}})', \tag{5}$$

where $\overline{\mathbf{m}} = \frac{1}{B}\sum_{b=1}^{B}\mathbf{m}^b$. As outlined previously, we then compute the weighting matrix $\mathbf{W}$ by taking the inverse of $\widehat{\boldsymbol{\Sigma}}$. Further in the text, we refer to the described weighting matrix setup as the *overlapping block bootstrap* weighting matrix.

3. Moment selection using simple machine learning

We now provide a complete specification of our machine learning-based approach to the moment set selection. Traditionally, researchers address the problem intuitively based on their expert knowledge of the model and the data-generating process it represents. This approach, however, raises questions regarding the optimality of moments arbitrarily chosen in this manner and the application of the SMM itself (Platt, 2020). Jalali et al. (2015) note that the number of moments should be at least as large as the number of estimated parameters, similar to the standard identification requirement in the GMM framework. Furthermore, the authors suggest that any other moments above this lower boundary should allow for better identification; thus, their presence should improve the estimation precision. This situation corresponds to the overidentification case in the GMM. According to Hansen (1982), overidentification allows for the construction of more efficient GMM estimators and the application of additional specification tests, which similarly holds for the SMM. However, the findings of Chen and Lux (2018) suggest that this rule might not always hold completely for the simulated version of the method under finite samples. The efficiency gains can be modest, and more extensive sets of moments in specific cases may result in even slightly worse estimation precision than their subsets for specific parameters. The authors further argue that correlation between the moments can lead to deviations of the $\bar{J}$ statistic from its asymptotic distribution, which affects the power of the test of overidentifying restrictions, especially for large sets of moments. The efficient method of moments might provide a partial solution to this issue; however, it more or less instead shifts the problem from the moment set selection to a choice of a necessarily misspecified auxiliary structural model based on which the moment conditions are determined (Carrasco and Florens, 2002; Franke, 2009).

Stepwise regression is one of the primary algorithms used for model selection. Furthermore, it represents one of the earliest and most straightforward uses of machine learning techniques in econometrics. Two basic approaches lie at its core: backward stepwise elimination and forward stepwise selection. The former successively excludes variables from a model that initially contains all variables at its disposal. The latter relies on an iterative addition of variables to a regression model composed of only the intercept. Here, we propose applying stepwise techniques to the moment set choice for the SMM estimation.

3.1. Backward stepwise moment elimination

The first algorithm is the backward stepwise moment elimination (BSME). In this case, the subject of iterative reduction is the moment set. We begin with a set containing all the available moments. In each step, we, one by one, remove every moment that is currently available in the set. The reduction yielding the most precise estimate given a pseudo-empirical time series is performed. The logic behind this choice is that the moment removed adds the smallest amount of information to the estimation method.

The algorithm written in pseudo-code can be found in Algorithm 1. A total of four structures must be present in the environment where the algorithm is executed. First, a function $f(\boldsymbol{\theta})$ for generating a time series using the model and accepting its parameters as an input must be at its disposal (line 1). Second, we have to choose pseudo-true parameter values $\boldsymbol{\theta}^{true}$ to generate pseudo-empirical data (line 2). Parameterizations from the literature are preferred, ensuring that the model behaves appropriately and generates desired stylized facts. Third, the pseudo-empirical dataset $\mathbf{y} = f(\boldsymbol{\theta}^{true})$ must be generated before engaging in the procedure (line 3). Last, a set of D moment functions $M_F = \{m_1, \ldots, m_D\}$ utilised in the procedure is required (line 4). Furthermore, we must be able to perform the SMM estimation process using any given set of moment functions.

Algorithm 1 Backward stepwise moment elimination.

1: $f(\boldsymbol{\theta}) \sim$ estimated model
2: $\boldsymbol{\theta}^{true} \sim$ pseudo-true parameters selected for the estimated model
3: $\mathbf{y} \sim$ pseudo-empirical time series generated by applying parameters $\boldsymbol{\theta}^{true}$ to model $f(\boldsymbol{\theta})$
4: $M_F \sim$ full set $\{m_1, \ldots, m_D\}$ of all moment functions
5:
6: **procedure** BSME
7: $\quad M_S \leftarrow M_F$
8: $\quad$ **while** $\textsc{size}(M_S) > 1$ **do**
9: $\quad\quad$ **for all moments** m **in** M_S **do**
10: $\quad\quad\quad M_T \leftarrow M_S \setminus \{m\}$
11: $\quad\quad\quad$ conduct SMM for time series $\mathbf{y}$ using model $f(\boldsymbol{\theta})$ and moment functions in M_T
12: $\quad\quad\quad$ compute quality measure using results of SMM and parameters $\boldsymbol{\theta}^{true}$
13: $\quad\quad$ **end for**
14: $\quad\quad M_S \leftarrow M_T$ such that quality measure is optimized across all **for loop** cycles
15: $\quad$ **end while**
16: **end procedure**

At the beginning of the procedure, we initiate M_S to contain all the moment functions (line 7). Subsequently, we shrink the set by one moment function in each elimination round of the algorithm. In essence, we consider an elimination round to be each round of the `while` cycle representing the foundation of the algorithm (line 8). The `while` cycle continues until the size of the set M_S is smaller than or equal to one. In other words, the algorithm continues for as long as there is at least one moment to be left out of the set M_S. Within each elimination round, the procedure iterates over all moments in the set M_S using a `for` cycle (line 9). In each iteration, a temporary set M_T is created by removing the moment that the `for` cycle currently iterates over from the set M_S (line 10). Next, the parameters of the model $f(\boldsymbol{\theta})$ are estimated using the SMM applied to the pseudo-empirical time series $\mathbf{y}$ while utilizing moment functions contained in M_T (line 11). Based on the estimation results and the true parameter values $\boldsymbol{\theta}^{true}$, a measure of the quality of estimation is then computed (line 12). Finally, the moment function set M_T generating the best results in estimation quality is selected and used to replace the set M_S for the subsequent elimination round (line 14).

3.2. Forward stepwise moment selection

The second algorithm proposed is the forward stepwise moment selection (FSMS). Overall, the algorithm works in the same manner as BSME; however, in this case, the initial moment set is empty. In each iteration, one moment is added to the moment set such that the fit achieved for the pseudo-empirical time series is the best possible. Thus, the procedure should reveal moments with the highest value for estimation.

Pseudo-code of FSMS is presented in Algorithm 2. Only subtle differences exist between the two algorithms. First, the set

Algorithm 2 Forward stepwise moment selection.

1: $f(\boldsymbol{\theta}) \sim$ estimated model
2: $\boldsymbol{\theta}^{true} \sim$ pseudo-true parameters selected for the estimated model
3: $\mathbf{y} \sim$ pseudo-empirical time series generated by applying parameters $\boldsymbol{\theta}^{true}$ to model $f(\boldsymbol{\theta})$
4: $M_F \sim$ full set $\{m_1, \dots, m_D\}$ of all moment functions
5:
6: **procedure** FSMS
7: $\quad M_S \leftarrow$ empty set $\varnothing$ of moment functions
8: $\quad$ **while** $M_S \neq M_F$ **do**
9: $\quad\quad$ **for all moments** m **in** $M_F \setminus M_S$ **do**
10: $\quad\quad\quad M_T \leftarrow M_S \cup \{m\}$
11: $\quad\quad\quad$ conduct SMM for time series $\mathbf{y}$ using model $f(\boldsymbol{\theta})$ and moment functions in M_T
12: $\quad\quad\quad$ compute quality measure using results of SMM and parameters $\boldsymbol{\theta}^{true}$
13: $\quad\quad$ **end for**
14: $\quad\quad M_S \leftarrow M_T$ such that quality measure is optimized across all **for loop** cycles
15: $\quad$ **end while**
16: **end procedure**

M_S is initiated as an empty set (line 7). Second, the `while` cycle representing selection rounds of the algorithm cycles for as long as the set M_S does not contain every single moment from the full moment set M_F (line 8). In other words, the `while` cycle does not stop until the set M_S contains D elements. Additionally, the `for` cycle of each selection round iterates over all moment functions from the full set M_F that are not already contained in the set M_S (line 9). Doing so establishes the impact of each moment function addition on the overall estimation performance. Finally, the temporary set M_T is created by adding the moment that the `for` cycle currently iterates over to the set M_S (line 10). Apart from these few differences, the algorithm remains the same.

3.3. Selection criterion

While describing the algorithms, we do not elaborate on the quality metric used to differentiate between the results of estimations employing different moment sets. However, it has a crucial impact on the selection process. We use the widely accepted root mean square error (RMSE). Traditionally, this metric is calculated in the context of a single parameter. Therefore, we make a slight adjustment to the RMSE that accommodates the necessity of involving multiple parameters simultaneously.

Assume we have a model with a total of K estimated parameters. We may thus generate a simulated time series using a specific parameterization. We organize these known parameters into the form of a vector $\boldsymbol{\theta}^{true} = \left[\theta_1^{true}, \dots, \theta_K^{true}\right]$. Subsequently, we apply the SMM to the simulated time series. Depending on the number of Monte Carlo replications, we obtain a certain number of point estimates for each parameter and compute the reported parameter estimate as their mean values across all replications. All parameter estimates form a vector $\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1, \dots, \hat{\theta}_K\right]$. Additionally, we can use the Monte Carlo point estimates to obtain the standard error and confidence interval for each parameter. The single parameter normalized RMSE$_{norm}$ is then calculated as follows:

$$RMSE_{norm} := RMSE_{norm}(\hat{\theta}_i) = \sqrt{\frac{(\hat{\theta}_i - \theta_i^{true})^2 + \hat{\sigma}_i^2}{|\theta_i^{true}|}}, \tag{6}$$

where $\hat{\sigma}_i{}^2$ is the estimated variance of the SMM estimator of the parameter θ_i. Traditionally, only the nominator containing the bias term $(\hat{\theta}_i - \theta_i^{true})^2$ and the variance term $\hat{\sigma}_i^2$ is used as the metric. Division of the metric by the true parameter in absolute value represents a normalization step. As a result, it ensures that parameters of different magnitudes can be compared and aggregated into a single value. Subsequently, we can compute the metric for multiple parameters as follows:

$$RMSE_{\hat{\boldsymbol{\theta}}} := \frac{1}{K}\sum_{k=1}^{K} RMSE_{norm}(\hat{\theta}_k). \tag{7}$$

3.4. Effective moment set selection

The BSME and FSMS algorithms identify numerous best moment sets in terms of the $RMSE_{\hat{\boldsymbol{\theta}}}$ metric specified in (7), one for each elimination/selection round, and compare their suitability for the SMM estimation. Among these *best moment sets*, one would generally choose the set minimizing the $RMSE_{\hat{\theta}}$ metric that we label the *overall best moment set*. However, we argue that some performance within a reasonable range can be sacrificed to obtain an *effective moment set*, a less complex parsimonious alternative to the overall best moment set, lessening the computing power requirements. In this regard, we follow an example by Hastie et al. (2009, pg. 62) applying it under very similar circumstances within the general stepwise feature selection framework.

We consider all the best moment sets for effective moment set selection. Assuming that we estimate K parameters and complete L independent replications to approximate the SMM estimator distribution, we end up with $K \cdot L$ point estimates for each moment set. Given a moment set and the corresponding set of point estimates $\hat{\theta}_{kl}$, $k \in \{1, \dots, K\}$, $l \in \{1, \dots L\}$, we compute the mean normalized bias $b(\hat{\boldsymbol{\theta}})$ as follows:

$$b(\hat{\boldsymbol{\theta}}) = \frac{1}{K \cdot L} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{|\hat{\theta}_{kl} - \theta_k^{true}|}{|\theta_k^{true}|}, \tag{8}$$

where $\hat{\theta}_{kl}$ is the point estimate of parameter k for replication l and θ_k is the true value of parameter k. The effective moment set is then the smallest best moment set whose $b(\hat{\boldsymbol{\theta}})$ lies within the 95% confidence band of $b(\hat{\boldsymbol{\theta}})$ of the overall best moment set. Following standard theoretical requirements for parameter identification in the GMM and SMM, we further employ an additional constraint for the effective moment set: it contains at least as many moments as there are estimated parameters.

3.5. Full moment set and benchmarks

The presented approaches must be supplied with an initial full set of moments used in the elimination/selection process. Theoretically, there is an immense number of moments that can be used. However, the initial set should have a reasonable size due to the computational complexity of the SMM. Assume that one wants to use D_F initial moments. Then, for FSMS, there are a total of $\frac{D_F(D_F+1)}{2}$ moment sets for which the estimation is conducted. For BSME, the total number is one less because the initial full set with D_F moments is not evaluated. Thus, the computation of results for a vast number of moment sets is avoided, which is in stark contrast to the naive procedure of 'best-subset selection,' which considers all possible combinations of the moment set. This is because 'greedy' stepwise procedures can be expected to choose only among the most promising alternatives. Given D_F moments, a total of $2^{D_F} - 1$ different moment sets would be computed within the best-subset selection algorithm. For $D_F = 15$ moments, the FSMS procedure computes the results for only 120 out of 32,767 total available moment sets. The comparison becomes even starker as the number of moments increases. For $D_F = 50$, the results for only 1275 moment sets must be calculated out of $1.1 \cdot 10^{15}$ possible combinations.

We thus turn to financial agent-based models literature to choose the moments included in the initial full set. In Franke and Westerhoff (2012), the SMM is applied to a structural stochastic volatility asset pricing model. In Chen and Lux (2018), the estimation procedure is studied in connection to a simple financial agent-based model introduced by Alfarano et al. (2008). Both models replicate the main stylized facts of financial data, specifically the absence of autocorrelation of raw returns, heavy tails of the returns' distribution, clustered volatility, and potentially long memory patterns. In both papers, the moment sets are constructed to capture the dynamics behind these stylized facts. However, even though the same stylized empirical patterns are supposed to be captured, the moment sets differ markedly between the papers. Three sets, one from Franke and Westerhoff (2012) and two from Chen and Lux (2018), are thus selected as benchmarks for the proposed SMM extension. Most recently, identical sets of moments were applied for the approximate Bayesian computation method to the Alfarano et al. (2008) model in Lux (2022a). The full set provided to the algorithms is then based on those three benchmarks.

Moreover, following Franke (2009); Schmitt and Westerhoff (2017b); Schwartz and Kirstein (2022), we also include a second Hill estimator for the top 2.5% of absolute returns to capture the tail behavior of data from real financial markets in more detail. We thus avoid insufficient pseudo-fat tails potentially generated by financial agent-based models and a potential mismatch between the relative magnitudes of extreme simulated and empirical data values. Finally, according to Schmitt and Westerhoff (2017a), we add the autocorrelation coefficients of raw returns for lags 2 and 3 to adequately capture the crucial random walk characteristic of real financial markets.

Table 1 defines the resulting complete set of moment functions m_i, $i = 1, \dots, 22$. The moment selection includes the unconditional second and fourth moments of raw returns, ith-order autocorrelations, $i \in \{1, 2, 3\}$, of raw returns, the unconditional first moment of absolute returns, the Hill estimator of the tail index of absolute returns based on 2.5% and 5% of extreme observations, jth-order autocorrelations, $j \in \{1, 5, 10, 15, 20, 25, 50, 100\}$, of absolute returns, and kth-order autocorrelations, $k \in \{1, 5, 10, 15, 20, 25\}$, of squared returns. Following the example of Franke and Westerhoff (2011, 2012, 2016), we diminish the impact of outliers by smoothing out the calculated autocorrelations over their neighborhoods. In other words, we compute the kth-order autocorrelation as the mean of lth-order autocorrelations, $l \in \{k-1, k, k+1\}$. For the 1-st-order autocorrelations, we use only the two orders available. We apply this technique to all autocorrelation moments computed for absolute and squared returns.

4. Testbed models

This section presents three models on which we test the FSMS and BSME algorithms. To provide evidence of the functionality of the proposed methodology and to observe its general properties, models with an increasing level of complexity (Lee et al., 2015; Mandes and Winker, 2017; Kukacka and Kristoufek, 2020; 2021) and an increasing number of estimated parameters are selected. The first testbed framework is the MSM model by Calvet and Fisher (2004) with two estimated parameters. As our methodology should mainly find its use in estimating intractable agent-based models, a popular simple financial model of herding by Alfarano et al. (2008) with three estimated parameters and a more complex structural stochastic volatility model by Franke and Westerhoff (2012) with seven estimated parameters are also considered.

4.1. Markov-switching multifractal model

The baseline model used to test the fundamental functionality of our approach is a univariate, discrete-time MSM model described by Calvet and Fisher (2004). It is a special case of the regime-switching modeling framework introduced to the asset-pricing literature by Hamilton (1989) following its original application in statistical physics by Mandelbrot (1974). The model successfully replicates all essential financial stylized facts, including volatility clustering and long memory patterns (Lux, 2008). As explained by Lux (2022c), the model can be estimated in full analytically using maximum likelihood; however, numerical methods such as SMM may become preferred when more complex settings of the model are assumed, causing sharp increases in computational demands of the analytical approach.

The log-returns r_t at time t of a single asset are modeled by taking the product of a finite number of latent volatility components with varying persistence, as follows:

$$r_t = \sigma \left(\prod_{i=1}^{k} M_{i,t} \right)^{1/2} \varepsilon_t, \tag{9}$$

where σ is a constant scale factor representing unconditional volatility, $\varepsilon_t \sim \mathcal{N}(0,1)$ is a noise term, k is the total number of volatility components, and $M_{i,t}$ is the value of volatility component i. In combination, the k volatility components generate a Markov chain with 2^k possible states, as each component can assume one of two values. At each step, the value of component i can either be renewed with probability γ_i or persist with probability $(1-\gamma_i)$. Following Lux (2022c), we set the renewal probabilities to:

$$\gamma_i = 2^{-(k-i)}. \tag{10}$$

A discrete and continuous distribution may be used when renewing a volatility component's value. We choose the simplest option and draw from the Binomial distribution with values $\{m_0, 2-m_0\}$, i.e., we select from the set of values with an equal probability assigned to each element whenever a volatility component's value is renewed.

The parametrization and optimization constraints follow Lux (2022c). The true values of the estimated parameters are $\sigma = 1$ and $m_0 = 1.2$. Also, we set $k = 8$, meaning the estimated model contains 256 different states.

4.2. Alfarano et al. (2008) model

Originally developed and applied through a series of papers (Alfarano et al., 2005; 2006; 2007), the second framework is the model of herding by Alfarano et al. (2008) built in the 'ant dynamics' tradition introduced by Kirman (1991, 1993). It was selected following the recent estimation efforts by Ghonghadze and Lux (2016); Chen and Lux (2018); Lux (2022a, 2018, 2022b), where the model standardly serves as a testbed for the performance of the SMM, GMM, approximate Bayesian computation, and new advanced methods based on the Markov chain Monte Carlo approach, such as sequential Monte Carlo based on a particle filter or the adaptive particle MCMC.

The model assumes two subpopulations of market participants: N_f fundamentalists and N_c noise traders, and a log fundamental value F_t of an underlying asset. Fundamentalists make their investment decision based on the deviation from the fundamental value, i.e., they buy/sell when the asset is under/overvalued. The average trading volume of fundamentalists is V_f, and their excess demand thus follows $D^f = N_f V_f (F_t - p_t)$, where p_t is the equilibrium market log-price at time t. The behavior of noise traders is determined by an opinion dynamics process, where an optimistic/pessimistic state of a given agent implies buying/selling V_c units of the asset at time t. A population sentiment index x_t is defined as equal to zero for balanced sentiment, while it reaches positive/negative values when noise traders are dominantly optimistic/pessimistic:

$$x_t = \frac{2n_{o,t}}{N_c} - 1 \in \langle -1, 1 \rangle, \tag{11}$$

where $n_{o,t}$ is the number of optimistic noise traders at time t. The value of the population sentiment index directly translates into the excess demand of noise traders that follows $D^c = N_c V_c x_t$.

The opinion dynamics process is associated with the subpopulation of noise traders who enter pairwise communication and switch between the two states. The Poisson intensity of autonomous switches $a \geq 0$ and the herding rate $b \geq 0$ drive the dynamics of the model:

$$\pi_{x,t}(- \to +) = \frac{N_c - n_{o,t}}{N_c}\left(a + b\frac{n_{o,t}}{N_c}\right) = (1 - x_t)\left[\frac{2a}{N_c} + b(1 + x_t)\right]N_c^2, \tag{12}$$

where $\pi_{x,t}(- \to +)$ denotes the probability that a pessimistic noise trader switches to an optimistic mood; the opposite direction works similarly.

Setting $b > a$ leads to a bimodal distribution of the sentiment index x_t, i.e., one observes a strong and persistent majority opinion with an episodic mean-reverting tendency among noise traders over time, as pairwise communication governs sentiment dynamics. Conversely, setting $a > b$ leads to a unimodal unconditional distribution of x_t centered around 0. In this case, autonomous opinion switches dominantly govern the sentiment with a stronger mean-reverting tendency. The model further assumes a Walrasian price adjustment mechanism with instantaneous market clearing under which the price change is derived from the overall excess demand: $\frac{dp}{dt} = D^f + D^c = N_f V_f (F_t - p_t) + N_c V_c x_t$. The continuous-time version of the model is summarized as follows:

$$\mathrm{d}F_t = \sigma_f \mathrm{d}B_{1,t}, \tag{13}$$

$$\mathrm{d}x_t = -2ax_t\mathrm{d}t + \sqrt{2b(1 - x_t^2) + \frac{4a}{n_c}}\mathrm{d}B_{2,t}, \tag{14}$$

$$p_t = F_t + \frac{N_c V_c}{N_f V_f} x_t \quad \Rightarrow \quad r_t = p_t - p_{t-1} \equiv \sigma_f e_t + \frac{N_c V_c}{N_f V_f}(x_t - x_{t-1}), \tag{15}$$

where F_t follows standard Brownian motion $B_{1,t}$ without drift, $\sigma_f \geq 0$ represents the standard deviation of its innovations, $B_{2,t}$ is another standard Brownian motion processes, r_t is the log-return at time t, and $e_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ due to the fundamental value for a unit time change $F_{t+1} \sim$ i.i.d. $\mathcal{N}(F_t, \sigma_f)$.

The parametrization follows Ghonghadze and Lux (2016); Chen and Lux (2018); Lux (2018, 2022b, 2022a) and the Langevin equation approximation of the model. The optimization constraints follow Lux (2022a). The true values of the three estimated parameters are $a = 0.0003$, $b = 0.0014$, and $\sigma_f = 0.03$. The other parameters of the model are set such that $N_c = 100$ and $\frac{N_c V_c}{N_f V_f} = 1$. Naturally, log-returns are used as the model output in our estimation exercises.

4.3. Franke and Westerhoff (2012) model

Finally, we consider a structural stochastic volatility model by Franke and Westerhoff (2011) and further developed in Franke and Westerhoff (2012). It describes a market where two population fractions of agents, chartists and fundamentalists, interact and respond to their dynamic environment by switching between the two groups. Since its adoption, the model has been used in many estimation exercises (Franke and Westerhoff, 2016; Lux, 2022b; Platt, 2022, to name few).

The price of an asset traded on the market is governed by demand. Specifically, a market maker changes the log-price p_t from one period to the next by observing the excess demand for the asset and proportionately adjusting the price such that $p_t = p_{t-1} + \mu(n_{t-1}^f d_{t-1}^f + n_{t-1}^c d_{t-1}^c)$, where d_t^f and d_t^c describe the excess demands of fundamentalists and chartists, respectively, n_t^f and $n_t^c = 1 - n_t^f$ represent the corresponding population share of each group, and μ stands for the adjustment rate of the asset price by the market maker. Depending on the current trading approach of a trader, their excess demand is governed by one of the following equations:

$$d_t^f = \phi(p^* - p_t) + \varepsilon_t^f \qquad \varepsilon_t^f \sim \mathcal{N}(0, \sigma_f^2), \tag{16}$$

$$d_t^c = \chi(p_t - p_{t-1}) + \varepsilon_t^c \qquad \varepsilon_t^c \sim \mathcal{N}(0, \sigma_c^2), \tag{17}$$

where p^* is the fundamental log-value of the asset, ε_t^f and ε_t^c represent the noise terms for fundamentalists and chartists, respectively, σ_f and σ_c determine the corresponding volatilities of the noise terms, and ϕ and χ are the respective adjustment parameters of the demands. The excess demands are governed by simplistic rules used throughout the literature. The excess demand of fundamentalists is driven by the deviation of the current asset price from its fundamental value. In contrast, the excess demand of chartists recognizes the asset's price change from the previous period to the current period. Adjustments of population shares are governed by the index a_t, indicating the propensity to switch such that:

$$a_t = \alpha_0 + \alpha_n(n_t^f - n_t^c) + \alpha_p(p_t - p^*)^2, \tag{18}$$

$$n_t^f = \frac{1}{1 + \exp(-\beta a_{t-1})}, \tag{19}$$

where α_0 is a 'predisposition' parameter representing a general trading preference towards one of the strategies, $\alpha_n \geq 0$ represents a 'herding' parameter, and $\alpha_p \geq 0$ is a price 'misalignment' parameter measuring the impact of mispricing from the fundamental log-value. Finally, β is the intensity of choice within the popular binomial logistic discrete choice approach.

Franke and Westerhoff (2012) present an entire family of models varying in the approach to population fraction switching and its determinants. Regarding the former, the transition probability approach (TPA) of Franke and Westerhoff (2011) is complemented by the discrete choice approach (Brock and Hommes, 1998, DCA) presented in (19). As for the determinants of the switching index, accumulated past profits of alternative trading strategies called wealth (W) are included among herding (H), predisposition (P), and misalignment (M) terms that we use in (18). Consequently, various subsets of the four determinants can be utilized to create multiple models. The DCA-HPM setup combination is found to be superior; therefore, it is also used in our paper.

The parameterization of the seven estimated coefficients, which follows Franke and Westerhoff (2012), is also used in Platt (2022), whom we also follow in terms of the optimization constraints: $\phi = 0.12$, $\chi = 1.5$, $\sigma_f = 0.758$, $\sigma_c = 2.087$, $\alpha_0 = -0.327$, $\alpha_n = 1.79$, $\alpha_p = 18.43$. The remaining parameters are set such that $\mu = 0.01$, $p^* = 0$, and $\beta = 1$. Again, log-returns are used as the model output for the estimation routine.

5. Monte Carlo simulation study

Although all three models are well known and verified in the recent literature, we first perform their basic time-series analysis. Next, we conduct a detailed preliminary analysis to find a suitable simulation setup that optimizes computational time without compromising the estimation precision of the SMM. This is evaluated via $\text{RMSE}_{\hat{\theta}}$ for the most challenging estimation exercise with the Franke and Westerhoff (2012) model.

5.1. Analysis of the warm-up period

First and foremost, we test the length of the warm-up period to ensure sufficient discarding of the initial observations to eliminate any influence of the initial conditions. We follow Welch's method (Welch, 1983) recently discussed in Vandin et al. (2022). Based on 1000 independent realizations of the process, a warm-up period of 100 observations appears sufficient to provide all three testbed models enough time to stabilize their dynamics. We thus implement a warm-up period of 200 initial observations for cautionary reasons in the subsequent analysis.

5.2. General simulation setup for all models

The analysis is executed in `Julia 1.6.1`, and we take advantage of the `Distributed.jl` package for parallel computing. The length of the analyzed series is always $T_{sim} = T_{emp} = 6750$ after discarding initial warm-up observations. Interestingly, while for the double length $T_{sim} = 2 \cdot T_{emp}$, the computing time doubles, the precision increases by only approx. 3%.

5.3. Descriptive statistics and the analysis of stationarity and ergodicity

Typical time series outputs of the testbed models are depicted in Fig. A.3 in Appendix A. Table A.6 further displays aggregate descriptive statistics, their 95% sample confidence intervals, and rejection rates of six essential time series statistical tests based on 1000 random runs and the general simulation setup. As expected, all three models violate normality, demonstrated through excess kurtosis and confirmed statistically by 100% rejection rates for the Jarque–Bera test. For all three models, stationarity is supported by the Augmented Dickey–Fuller test. At the same time, stationarity is not rejected by the Kwiatkowski–Phillips–Schmidt–Shin test. Finally, we conduct two ergodicity tests according to Grazzini (2012) and Delli Gatti and Grazzini (2020), strongly supporting the ergodic characteristics for all three models. The general regularity conditions for the SMM estimation (Lee and Ingram, 1991; Duffie and Singleton, 1993) are thus verified and met.

5.4. General Monte Carlo and SMM estimation setup

We evaluate each Monte Carlo experiment based on $L = 96$ independent replications. While the parallel server capacities at our disposal provide largely sufficient computation budget for most of the standard tasks, our simulation setup is dictated by the computationally costly Monte Carlo executions of the FSMS and BSME machine learning algorithms.[2] Nevertheless, while for 192 runs the computational time naturally doubles, the SMM precision increases by less than 1%.

[2] Most of the tasks were performed on a server with an Intel(R) Xeon(R) Gold 6126 CPU @ 2.60 GHz processor with 48 cores and 768 GB RAM. Parallelizing independent replications to all available cores, one complete FSMS (BSME) procedure requiring 253 (252) individual SMM estimations takes approximately 133 h for the most complex Franke and Westerhoff (2012) model. For the whole experiment with three models and both algorithms summarized in Fig. 1, we need less than six times as much since computations for simpler models are shorter but remain the same order of magnitude of the computational complexity. However, after the machine learning exploration of the moment sets is done for a given model, subsequent practical applications require the execution of few single estimations that take only few hours under the same setup.

Table 1
Moments sets: the full set and benchmarks.

Moment function	*Label*	*CHL4*	*FW9*	*CHL15*
$m_1 = E(r_t^2)$	RAW-VAR	✓		✓
$m_2 = E(r_t^4)$	RAW-KURT	✓		✓
$m_3 = E(r_t r_{t-1})$	RAW-AC1	✓	✓	✓
$m_4 = E(r_t r_{t-2})$	RAW-AC2			
$m_5 = E(r_t r_{t-3})$	RAW-AC3			
$m_6 = E(\|r_t\|)$	ABS-MEAN		✓	
$m_7 = Hill(\|r_t\|, 2.5)$	ABS-HILL2.5			
$m_8 = Hill(\|r_t\|, 5)$	ABS-HILL5		✓	
$m_9 = E(\|r_t\|\|r_{t-1}\|)$	ABS-AC1		✓	✓
$m_{10} = E(\|r_t\|\|r_{t-5}\|)$	ABS-AC5		✓	✓
$m_{11} = E(\|r_t\|\|r_{t-10}\|)$	ABS-AC10		✓	✓
$m_{12} = E(\|r_t\|\|r_{t-15}\|)$	ABS-AC15			✓
$m_{13} = E(\|r_t\|\|r_{t-20}\|)$	ABS-AC20			✓
$m_{14} = E(\|r_t\|\|r_{t-25}\|)$	ABS-AC25		✓	✓
$m_{15} = E(\|r_t\|\|r_{t-50}\|)$	ABS-AC50		✓	
$m_{16} = E(\|r_t\|\|r_{t-100}\|)$	ABS-AC100		✓	
$m_{17} = E(r_t^2 r_{t-1}^2)$	SQR-AC1	✓		✓
$m_{18} = E(r_t^2 r_{t-5}^2)$	SQR-AC5			✓
$m_{19} = E(r_t^2 r_{t-10}^2)$	SQR-AC10			✓
$m_{20} = E(r_t^2 r_{t-15}^2)$	SQR-AC15			✓
$m_{21} = E(r_t^2 r_{t-20}^2)$	SQR-AC20			✓
$m_{22} = E(r_t^2 r_{t-25}^2)$	SQR-AC25			✓

Note: Benchmark naming is based on the underlying papers: Chen and Lux (2018) for *CHL4* and *CHL15*; Franke and Westerhoff (2012) for *FW9*.

The SMM is implemented with the overlapping block bootstrap weighting matrix with bootstrap size $B = 5000$ as described in Section 2.2 (Franke and Westerhoff, 2011; 2012). For the SMM optimizations, simulated moments are computed as averages over 100 independent simulations. Interestingly, while for more simulations, the computational time increases linearly, for 250 simulations, the estimation precision increases by not even 3%. Finally, a default recommended differential evolution optimizer from the `BlackBoxOptim.jl` package is used. The optimization algorithm was tested against competing optimizers available in other packages, such as the standard `Optim.jl`, and it delivers the best performance for the given optimization problem. Technically, we constrain the optimization by 4000 functional evaluations (2000 for the MSM model as its relative simplicity ensures faster convergence). Based on our preliminary analysis, allowing for more evaluations linearly increases the computational time without delivering better estimation precision.

6. Analysis of results

Figure 1 presents the outputs of the two machine learning procedures, BSME and FSMS, for all three testbed models. We apply the proposed methodology first to the MSM model as a challenge to its basic functionality, having important practical implications due to the computational difficulty of analytical estimation approaches for more complex specifications of the model (Lux, 2022c). For the simple financial agent-based model of herding by Alfarano et al. (2008), several moment conditions can be derived analytically, and one can potentially apply standard GMM estimation (Ghonghadze and Lux, 2016). Still, the complexity of the model calls for a more flexible SMM approach. Finally, for the Franke and Westerhoff (2012) model, various subsets of its seven parameters are typically estimated. However, our ultimate goal is to estimate the complete parameter set simultaneously. Thus, the results obtained for these two financial agent-based models have important implications, as one has to address the moment set selection issue in empirical practice.

6.1. Two machine learning processes

In each round of the BSME algorithm, one moment is eliminated from the initially full moment set such that the set achieving the lowest $\mathrm{RMSE}_{\hat{\theta}}$ in the given round determines the elimination. This principle is reversed for FSMS, i.e., the moments are gradually added to an initially empty set. Therefore, the number of moments included coincides with the selection round for FSMS. Conversely, the first round of BSME procedure considers the largest sets of 21 moments, and each subsequent round lowers the number of moments by one.

For the MSM model with two estimated parameters, both algorithms effectively detect the most important moments as the sets of size two already provide strong estimation performance, which remains stable for all larger sets. However, there is a sizeable difference in the quality of individual estimates for the other two models. Especially for FSMS, the increasing complexity of the models with a larger number of estimated parameters requires more crucial moments to be included to sufficiently describe their dynamics. This typically results in an initial steep decline in $\mathrm{RMSE}_{\hat{\theta}}$ for FSMS in panels (d) and (f) as more moments gradually come into play. The variation of the estimation performance for the two agent-based models is

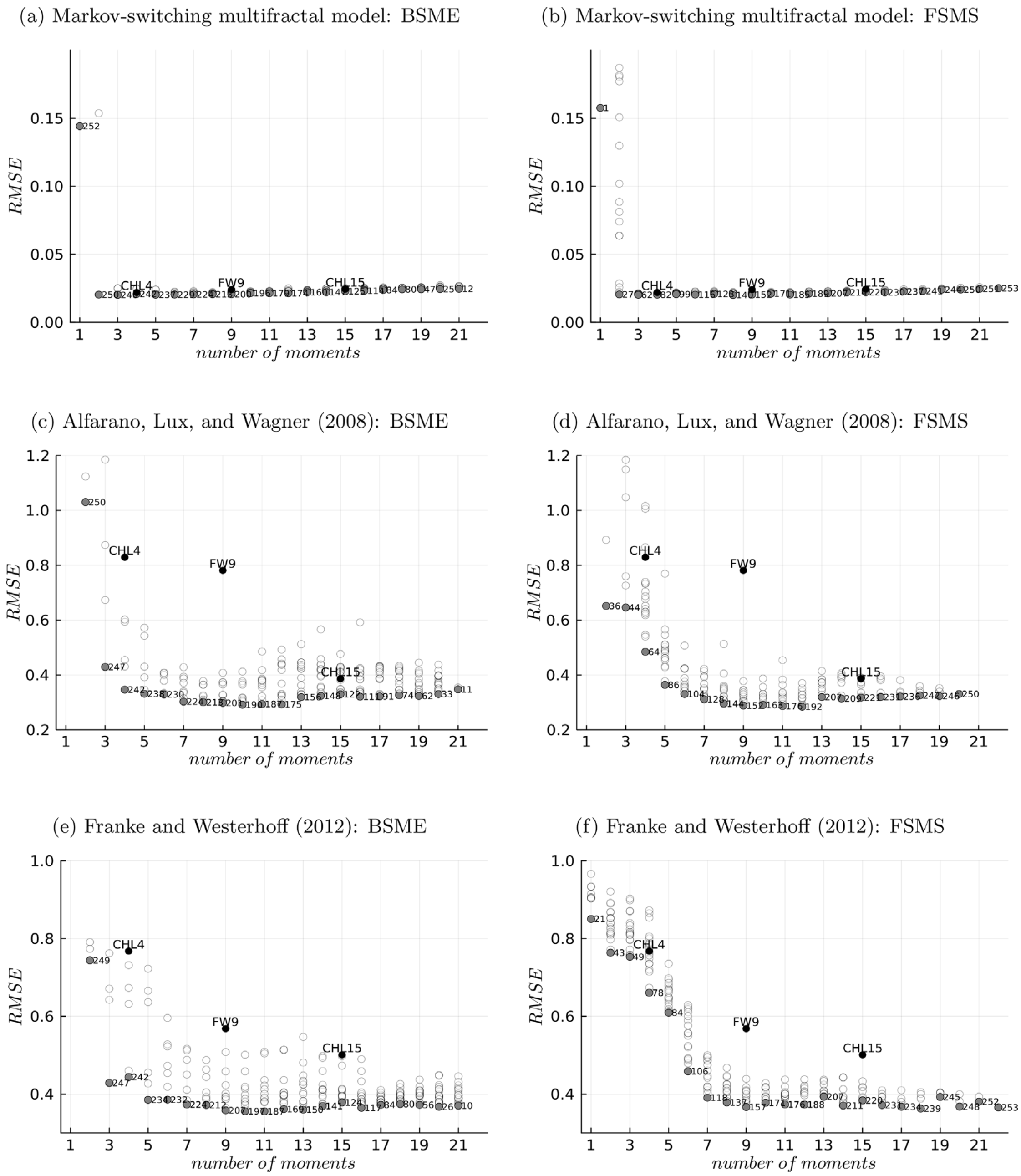


Fig. 1. Results for BSME and FSMS.
Note: Transparent points depict results for all moment sets involved in the selection process, numbered gray points depict results for the best performing moment sets of each selection round, and named black points depict results for the benchmark moment sets according to Table 1. For visualization purposes, moment sets with high $RMSE_{\hat{\theta}}$ compared to the benchmark sets are excluded. The number of moments included in a set coincides with the selection round for FSMS; for BSME, this relationship is reversed: the results of the first BSME round are associated with 21 moments, the second round with 20 moments, etc.

larger than the random walk model and relatively stable across most rounds except for the smallest sets, where it naturally increases.

6.2. Three benchmark moment sets

Immediately, one can observe that the performance of all three benchmark sets taken from Franke and Westerhoff (2012) and Chen and Lux (2018) is not nearly as good as that of most moment sets identified by both algorithms for the two agent-based models. This contrasts with the straightforward case of the MSM model, panels (a) and (b), where the benchmark sets provide practically the same estimation performance as the best sets of size two and larger. For the other two agent-based frameworks, the performance of benchmarks is positively related to the number of parameters they contain. Especially for the Alfarano et al. (2008) model, while the *CHL4* benchmark appears strongly inferior, the *CHL15* set achieves a slightly worse but comparable performance to that of the best moment sets identified by our method. This is not surprising as it has been suggested and thoroughly analyzed by one of the model's authors. The performance of the *FW9* benchmark lies between these two but, by far, falls short of the performance of the best moment sets. Finally, we observe a disproportional difference between the performance of the *CHL4* benchmark and the best moment sets identified by the two procedures. While for BSME, all sets of length four or more strongly outperform *CHL4*, FSMS performs worse in this respect. The explanation lies in the inherent differences between the two procedures and the potential interplay between different moments. We elaborate on this phenomenon further in Sections 6.3, 8.1, and 8.2. Unlike for the previous model, for Franke and Westerhoff (2012) and the comparison *FW9* vs. *CHL15* benchmarks, the improvement is much smaller, primarily because the *FW9* benchmark was explicitly designed for the given model.

6.3. BSME, FSMS specifics for individual models

Looking at Fig. 1, panel (b), one can notice a large variance at the moment set performance in the second round of FSMS, with set #27 of size two strongly outperforming most of the other moment sets. The selected moment is ABS-MEAN supplementing RAW-VAR chosen in the first round. Furthermore, the estimation precision measured by $\mathrm{RMSE}_{\hat{\theta}}$ increases markedly. The dispersion then practically disappears in the next rounds. This follows the notion that the two different measures of volatility capture the dynamics of the MSM model almost perfectly, while other moments provide no added value to the estimation. Similarly, the results in panel (a) for BSME exhibit practically no variation under the elimination process up to the moment size of two.

Focusing on panel (c) for the BSME process, instead of only two moments crucial to the SMM performance for the previous model, interactions of multiple moments are necessary for the proper estimation of the Alfarano et al. (2008) model. This claim is further supported by the benchmark moment set *CHL15* performance, which almost reaches the estimation performance of the best moment sets identified by the two machine learning algorithms. Panel (d) for FSMS suggests that slightly larger sets lead to more efficient estimation than sets of only three or four moments. While adding the fifth and sixth moments produces a considerable performance enhancement, the improvement curve achieves relative flatness in the subsequent steps.

Panel (f) depicting the FSMS process for the Franke and Westerhoff (2012) model reveals an apparent steep decrease in $\mathrm{RMSE}_{\hat{\theta}}$ with additional moments up to round no. 7. This means that multiple moment functions gradually add to the explanatory power of the expanding set up to some saturation level for a complex model with many estimated parameters. Interestingly, in panel (e) displaying the BSME process, even the best set consisting of only five moments demonstrates estimation performance similar to all larger best sets. This is surprising because, in theory, the such set is too small to identify the estimated parameters. Part of the reason may be that the moments excluded in the corresponding elimination rounds are among the least impactful across the whole procedure. Also, the BSME algorithm tends to preserve valuable moment interactions that may provide enough information to identify additional parameters.

6.4. Effective moment set selection

Figure 2 depicts the effective moment set selection as described in Section 3.4. The full vertical line corresponds to the choice of the effective moment set lying within the 95% confidence band of the overall best set. Effective sets represent less complex parsimonious alternatives to the overall best moment sets and are, by definition, always their subsets. As such, they contain only the key moments necessary for adequate estimation performance. The specific composition of the effective moment sets and the overall best sets is further described in detail in Table 2, and their estimation performance is summarized in Table 3.

Results in Figs. 1 and 2 suggest that especially for the MSM and Franke and Westerhoff (2012) models, the effective moment sets are also the minimal possible sets for which the model parameters are just-identified. This corresponds to the standard econometric method of moments. The other moments, on the use of which the SMM is based, deliver almost no improvement to the estimation performance for the best sets of individual rounds but tend to decrease the dispersion of estimation performance of other sets of the same size for the Franke and Westerhoff (2012) model. Thus, one must select the correct effective moment set because otherwise, some larger sets that can take advantage of the SMM-related over-identification of parameters will likely be more efficient. While our selection algorithms ensure that the most appropriate effective moment set is indeed selected, if someone were to experiment intuitively/manually with the moment selection, the SMM approach represented by more than the minimal number of moments would undoubtedly be helpful, since, due to the decreasing dispersion across different sets, it reduces the risk of a suboptimal pick.

Table 2
Moments sets: composition of the effective and overall best sets.

		MSM		ALW (2008)		FW (2012)	
		BSME	FSMS	BSME	FSMS	BSME	FSMS
Effective set		#250	#27	#238	#128	#224	#118
Overall best set		#246	#140	#190	#192	#187	#239
Moment function	*Label*						
$m_1 = E(r_t^2)$	RAW-VAR	**1**	**1**	**1**	**3**	**2**	**7**
$m_2 = E(r_t^4)$	RAW-KURT			**2**	**4**		**5**
$m_3 = E(r_t r_{t-1})$	RAW-AC1				10		11
$m_4 = E(r_t r_{t-2})$	RAW-AC2			10	11		12
$m_5 = E(r_t r_{t-3})$	RAW-AC3					**6**	18
$m_6 = E(\|r_t\|)$	ABS-MEAN	**2**	**2**	**3**	**5**	**3**	**3**
$m_7 = Hill(\|r_t\|, 5)$	ABS-HILL2.5			**5**	**6**	9	8
$m_8 = Hill(\|r_t\|, 5)$	ABS-HILL5		7	9	9	**1**	**6**
$m_9 = E(\|r_t\|\|r_{t-1}\|)$	ABS-AC1						17
$m_10 = E(\|r_t\|\|r_{t-5}\|)$	ABS-AC5		8				14
$m_11 = E(\|r_t\|\|r_{t-10}\|)$	ABS-AC10					**5**	15
$m_12 = E(\|r_t\|\|r_{t-15}\|)$	ABS-AC15				12		
$m_{13} = E(\|r_t\|\|r_{t-20}\|)$	ABS-AC20						
$m_{14} = E(\|r_t\|\|r_{t-25}\|)$	ABS-AC25			7	**2**	**4**	9
$m_{15} = E(\|r_t\|\|r_{t-50}\|)$	ABS-AC50						10
$m_{16} = E(\|r_t\|\|r_{t-100}\|)$	ABS-AC100			**4**	**1**	8	**4**
$m_{17} = E(r_t^2 r_{t-1}^2)$	SQR-AC1			8	**7**	10	
$m_{18} = E(r_t^2 r_{t-5}^2)$	SQR-AC5		6	6	8	12	16
$m_{19} = E(r_t^2 r_{t-10}^2)$	SQR-AC10		5			**7**	13
$m_{20} = E(r_t^2 r_{t-15}^2)$	SQR-AC15						
$m_{21} = E(r_t^2 r_{t-20}^2)$	SQR-AC20		3				**1**
$m_{22} = E(r_t^2 r_{t-25}^2)$	SQR-AC25	3	4				**2**

Note: 'MSM' is the Markov-switching multifractal model, 'ALW (2008)' is the Alfarano et al. (2008) model, and 'FW (2012)' is the Franke and Westerhoff (2012) model. The moments belonging to effective (sub-)sets are in bold, and other moments supplementing effective sets into overall best sets are in footnote size. The depicted number corresponds to the importance of the given moment. That means, for BSME, the most important moment eliminated last is marked by '1', the second-last by '2', etc., while for FSMS, it directly corresponds to the round in which the respective moment was selected as presented in Fig. 1.

Table 3
Estimation improvements: overall best and effective sets.

		CHL4	*FW9*	*CHL15*
Markov-switching multifractal model—2 parameters				
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.0220	0.0241	0.0244
FSMS—Best set #140 (8)	0.0201	8.6%	16.6%	17.6%
FSMS—Efficient set #27 (2)	0.0205	6.8%	14.9%	16.0%
BSME—Best set #246 (3)	0.0202	8.2%	16.2%	17.2%
BSME—Efficient set #250 (2)	0.0205	6.8%	14.9%	16.0%
Alfarano et al.(2008)—3 parameters				
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.829	0.781	0.386
BSME—Overall best set #190 (10)	0.292	64.8%	62.6%	24.4%
BSME—Effective set #238 (5)	0.331	60.1%	57.6%	14.2%
FSMS—Overall best set #192 (12)	0.285	65.6%	63.5%	26.2%
FSMS—Effective set #128 (7)	0.311	62.5%	60.2%	19.4%
Franke and Westerhoff (2012)—7 parameters				
Benchmark	$\text{RMSE}_{\hat{\theta}}$		0.568	0.501
BSME—Overall best set #187 (11)	0.355		37.5%	29.1%
BSME—Effective set #224 (7)	0.373		34.3%	25.5%
FSMS—Overall best set #239 (18)	0.363		36.1%	27.5%
FSMS—Effective set #118 (7)	0.390		31.3%	22.2%

Note: Numbers in parentheses after set numbers indicate the set size. Estimation improvements compared to the benchmark moment sets are reported in %.

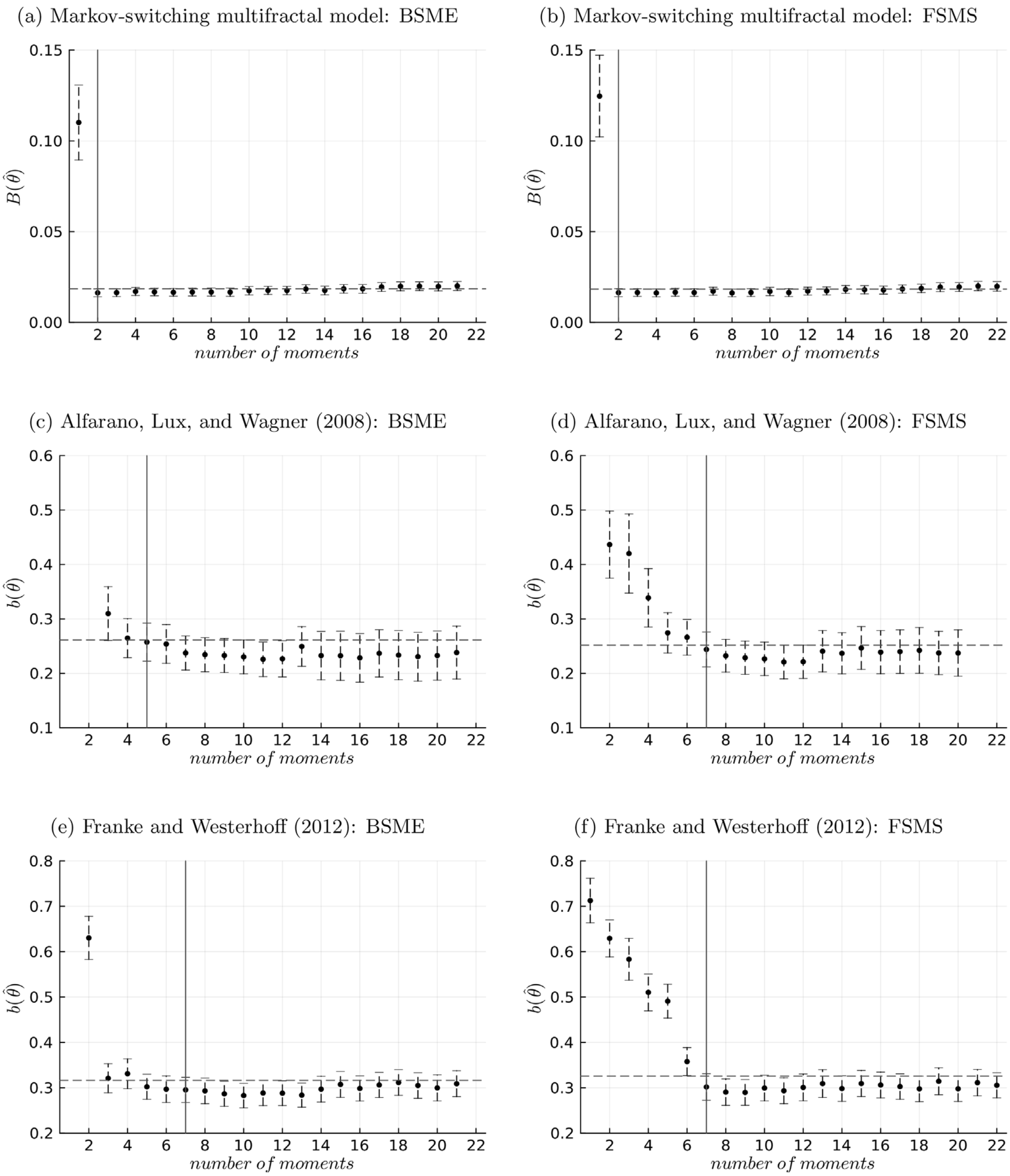


Fig. 2. Effective moment sets.
Note: Black points depict the mean normalized bias $b(\hat{\boldsymbol{\theta}})$ (8) for each of the best moment sets of each round presented in Fig. 1 and its corresponding 95% confidence band. The dashed horizontal line represents the upper 95% confidence bound for the overall best moment set. The full vertical line marks the effective moment set. For visualization purposes, moment sets with high $b(\hat{\boldsymbol{\theta}})$ are excluded.

6.5. Composition of the effective and overall best sets

The composition of the resulting sets is presented in Table 2.[3] The moments belonging to effective (sub-)sets are in bold, and other moments supplementing effective sets into overall best sets are in footnote size. One can thus study essential similarities and differences between the two pairwise algorithms. For the MSM model, RAW-VAR survives until the very end of the BSME elimination process, and it is selected first for FSMS. Next, ABS-MEAN also reveals its crucial importance.

[3] The remaining machine learning evolution of the moments in each round is presented in Appendix B for completeness.

This moment is the second last elimination and the second inclusion, additionally reflecting the impact of the calibrated unconditional volatility σ. The two identical effective sets selected in both directions thus indicate a considerable degree of stability of our method. Finally, some autocorrelations of squared returns, and a measure of heavy tails, ABS-HILL5, also appear potentially useful for capturing the model dynamics.

The two procedures lead to very similar outcomes for the Alfarano et al. (2008) model. Concretely, the seven moments eliminated last by BSME overlap almost perfectly with those seven selected first by FSMS. The only insignificant difference is the exchange of SQR-AC1 and SQR-AC5. For FSMS, two high-order autocorrelations of absolute returns, ABS-AC100 and ABS-AC25, are selected first, thereby reflecting a more complex nature of the model output and the importance of volatility clustering and potentially long memory for model dynamics. The BSME elimination procedure retains to the very end, on the other hand, ABS-MEAN, RAW-KURT, and RAW-VAR. Interestingly, two different measures of volatility, RAW-VAR and ABS-MEAN, and two measures of excess kurtosis and heavy tails, RAW-KURT and ABS-HILL2.5, make it to the parsimonious effective sets.

The comparison of the two processes for the Franke and Westerhoff (2012) model does not tell such a consistent story as for the two simpler models above. This is, however, expected for such a complex model with many estimated parameters. Still, crucial measures of volatility, RAW-VAR and ABS-MEAN, important measures of heavy tails, RAW-KURT and ABS-HILL5, and a composition of several high-order autocorrelations of returns highlighting the impact of volatility clustering and potentially long memory make it to both effective sets. Intriguingly, the first-order autocorrelation of raw returns (RAW-AC1) capturing the unpredictability of price changes, which is the only moments shared by all three benchmarks (see Table 1), is not part of any of the effective sets for any model.

6.6. Estimation improvements for the overall best and effective sets

Table 3 compares the performance of all important moment sets. Based on the previous interpretation, all twelve identified sets outperform all benchmark sets. Very similar $\text{RMSE}_{\hat{\boldsymbol{\theta}}}$ values for the MSM model stem from ‘sufficiently perfect’ estimation of the two model parameters based on already two moments identified by both stepwise approaches. The most significant improvements of more than 60% can be observed for the Alfarano et al. (2008) model against *CHL4* and *FW9* benchmarks, regardless of whether the overall best or effective sets are being compared. Overall, the four detected moment sets reveal comparable estimation performance. Still, FSMS identifies slightly better-performing sets. Only smaller improvements between 14 and 26% are achieved against *CHL15*. The essential comparison for the Franke and Westerhoff (2012) model is against *FW9*. Here, the estimation improvements achieve between 31 and 38% while the improvements against *CHL15* reach between 22 and 29%. In this case, BSME identifies slightly better-performing sets.

Interestingly, for the MSM model with the identical model setup, Lux (2022c, his Tab. 1) achieves the best estimation performance with $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.048$ and $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.052$ recomputed according to (7), using the maximum likelihood and simulated maximum likelihood, respectively. Moreover, for the Alfarano et al. (2008) model, Lux (2018, his Tab. 1) achieves the best performance with $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.229$ using sequential Monte Carlo based on a particle filter, while Lux (2022a, his Tab. 1, Method 3 and Method 2) achieves the best performance with $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.188$ and $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.260$ using the approximate Bayesian computation method. Furthermore, for the Franke and Westerhoff (2012) model, Platt (2022, his Tab. 7) achieves the best performance with $\text{RMSE}_{\hat{\boldsymbol{\theta}}} = 0.404$ using the likelihood-based Bayesian method utilizing a neural network.

6.7. Convergence of the parameter estimates

Finally, Fig. C.4 in Appendix C displays how the estimation performance in terms of graphically distinguished bias and variance evolves across elimination/selection rounds for each parameter separately. It allows for analyzing the convergence towards the true values and the reduction of variance with increasing the set of moments. Regarding true values, we emphasize that the model parameterizations and optimization constraints are always based either on the original paper introducing the model, as for Franke and Westerhoff (2012), or a most relevant follow-up econometric research of the same models, specifically Lux (2022c,a) and Platt (2022). All parameterizations are specified in detail in Section 4.

For clarity, we only comment on the most apparent patterns for each model. Panels (a) and (b) reveal a relative similarity for the estimation of the two parameters for the MSM model. We observe no bias and minimal variance of the estimator; only the m_0 parameter defining the interval of the Binomial distribution for the renewed volatility components requires two moments to be identified, while the unconditional volatility σ is already estimated almost perfectly by RAW-VAR. In panels (c) and (d), the often observed challenge in the estimation of the intensity of autonomous switches a (Chen and Lux, 2018; Lux, 2022a) is confirmed for the Alfarano et al. (2008) model. The SMM estimator naturally needs at least three parameters to stabilize its bias and variance. However, it then reveals a very stable behavior for all three parameters, very slightly converging in the number of moments, except for FSMS and very large sets for a and the smallest sets for the standard deviation of fundamental innovations σ_f. Similarly to the previous case, panels (e) and (f) also support the theoretical necessity of at least seven moments to gain stabilized behavior of the estimator in the case of the Franke and Westerhoff (2012) model. Otherwise, we face extensive volatility and serious biases for most model parameters. Estimation of the two adjustment coefficients of the fundamentalists’ and chartists’ demands, ϕ and χ, turns out to be the most complicated as we observe a stable upward bias and large variance for the latter and a potential for statistical insignificance at the 5% level for both.

6.8. Alternative weighting matrices

While the utilized weighting matrix **W** can be any positive-definite matrix of given dimensions, estimation efficiency is highly dependent on its selection. At the same time, only a very limited number of studies compare various approaches to its construction. We thus quantify the sensitivity of the SMM estimation performance concerning different weighting matrices for our most complex testbed Franke and Westerhoff (2012) model. Specifically, we analyze the performance of the two effective moment sets identified by our algorithms vs. the two natural benchmarks.

The four alternative approaches comprise the trivial *identity* matrix, which essentially neglects the importance of moments' sampling variability; and the simplified *diagonal* overlapping block bootstrap matrix in which the off-diagonal components are ignored (Boivin and Giannoni, 2006; Franke, 2009; Franke et al., 2015; Franke, 2018; Jang and Sacht, 2016; 2021, provide discussions on practical and economic rationale of this simplification). Next, the *non-overlapping block bootstrap* weighting matrix used in Franke and Westerhoff (2012) for which the empirical series is divided into non-overlapping blocks resulting in less variability in the bootstrapped series; and the *history sampling bootstrap* matrix utilized by Franke and Westerhoff (2016) where points and their corresponding history are sampled individually to avoid the joint-point problem of block bootstrap.

Table D.7 in Appendix D is to be compared to the bottom part of Table 3 with our standard overlapping block bootstrap setup. One directly observes that the trivialized setup with the identity matrix delivers largely inferior results. The $\text{RMSE}_{\hat{\boldsymbol{\theta}}}$ for the effective sets are roughly twice as large, and for the benchmarks, roughly half as large compared to our initial results. Importantly, very small and even negative improvements suggest that benchmarks deliver comparable or even better estimation performance. Similarly, the diagonal matrix is associated with much worse performance for our effective sets, but it shows only slightly worse $\text{RMSE}_{\hat{\boldsymbol{\theta}}}$ for the benchmarks.

On the other hand, while generally delivering slightly worse performances and estimation improvements, the non-overlapping block bootstrap and the history sampling bootstrap are comparable to the overlapping block bootstrap matrix. Conversely, it is worth commenting that while the $\text{RMSE}_{\hat{\boldsymbol{\theta}}}$ values for the target *FW9* benchmark are very similar among all three approaches, the history sampling bootstrap matrix seems worth strong recommending for the *CHL15* moment set for which it delivers markedly better estimation performance. In contrast, the non-overlapping block bootstrap largely fails for this benchmark. Nevertheless, the relative estimation improvements are positive in all but one case.

7. Empirical application

We now examine how the resulting moment sets identified by our approach can cope with real-world financial data. This empirical exercise also serves as an important out-of-sample verification tool as our extension on top of the SMM is essentially based on an in-sample simulation-based analysis. We again estimate our most complex testbed model by Franke and Westerhoff (2012) using the four moment sets suggested by our algorithms. We also directly compare these results to the performance of the two natural benchmark sets: *FW9* by the same authors and *CHL15* by Chen and Lux (2018). Regarding empirical data, we follow the best research practice and apply the SMM estimation routine to the log-returns of the S&P 500. The sample covers the period from 1980-01-02 to 2022-09-08, with 10745 observations [database accessed on 2022-09-12]. We kindly refer the reader to the GitHub repository for downloading the dataset. For the moment sets larger than the number of estimated parameters, we apply the *J*-test of overidentifying restrictions (4) to assess the model's validity and replicate the characteristics of empirical data.

7.1. Estimation setup

The empirical estimation procedure adopts the setup of the Monte Carlo simulation study with two minor alterations. First, as an individual estimation task can take full advantage of our computing budget, we foster the statistical validity by increasing the number of independent replications to 500 to face noisy empirical data better. Second, based on a preliminary exploration with the specified dataset and taking specific findings of Section 6.7 into account, we slightly modify constraints of the parameter space to support optimization search and to customize it for all six moment sets. While the numerical study adopts intervals used in Platt (2022), empirical optimization constraints for the recent dataset are summarized in Table 4. Most notably, we shrink the interval for the fundamentalists' adjustment parameters ϕ as the empirical estimates always stick near the lower boundary. On the other hand, we significantly extend the intervals for the herding and price misalignment parameters, α_n and α_p, to avoid hitting the upper boundary.

7.2. Empirical results

The results summarized in Table 4 strongly support our simple machine learning methodology and three of the four resulting moments sets. First and foremost, models originating from both benchmarks are rejected by the *J*-test at the 5% significance level as the true models (*CHL15* very strongly and *FW9* weakly with the *p*-value of 4.2%). Conversely, the parameterization implied by the overall best set #187, which is identified by BSME and includes 11 moments, is not rejected as being a good approximation of the data generation process with a high probability as the corresponding *p*-value exceeds 54%. Unfortunately, the model specification test is not applicable to the two effective sets as they consist of only seven

Table 4
Franke and Westerhoff (2012): results of the empirical study for S&P 500 data.

	Benchmarks		Resulting moment sets (see Table 3)			
Par./No. (D)	*FW9* 9	*CHL15* 15	BSME best #187 (11)	BSME eff. #224 (7)	FSMS best #239 (18)	FSMS eff. #118 (7)
$\hat{\phi}$	0.010	0.009	0.025	0.020	0.011	0.016
$\langle 0, 0.1\rangle$	(0.005, 0.012)	(0.008, 0.010)	(0.013, 0.061)	(0.013, 0.055)	(0.008, 0.017)	(0.012, 0.020)
$\hat{\chi}$	0.767	0.991	2.04	2.01	0.694	2.03
$\langle 0, 4\rangle$	(0.018, 2.39)	(0.034, 2.92)	(0.117, 3.86)	(0.095, 3.93)	(0.011, 2.30)	(0.132, 3.92)
$\hat{\sigma}_f$	0.870	0.841	0.821	0.835	0.850	0.858
$\langle 0.5, 1\rangle$	(0.853, 0.882)	(0.819, 0.862)	(0.770, 0.851)	(0.760, 0.868)	(0.818, 0.872)	(0.828, 0.877)
$\hat{\sigma}_c$	4.21	4.57	3.92	4.02	4.36	3.78
$\langle 2, 6\rangle$	(3.27, 5.30)	(4.17, 5.23)	(3.39, 4.90)	(3.03, 5.29)	(3.27, 5.69)	(3.28, 4.67)
$\hat{\alpha}_0$	0.022	−0.142	−0.239	−0.182	−0.133	−0.206
$\langle -0.5, 0.5\rangle$	(−0.218, 0.276)	(−0.223, −0.089)	(−0.459, −0.131)	(−0.408, 0.162)	(−0.276, 0.090)	(−0.294, −0.100)
$\hat{\alpha}_n$	1.74	1.98	2.39	2.14	2.13	1.95
$\langle 1, 4\rangle$	(1.31, 2.44)	(1.74, 2.33)	(1.95, 2.97)	(1.41, 2.88)	(1.64, 2.52)	(1.66, 2.28)
$\hat{\alpha}_p$	24.32	28.56	13.32	13.12	22.91	17.41
$\langle 5, 35\rangle$	(13.33, 29.35)	(24.79, 32.58)	(10.18, 19.21)	(9.24, 19.16)	(15.74, 29.63)	(12.12, 24.69)
$\bar{J} = J(\hat{\boldsymbol{\theta}})$	6.32	29.38	3.10	1.68	42.30	2.39
	(5.88, 6.84)	(28.26, 30.62)	(2.37, 3.72)	(1.37, 1.94)	(38.60, 44.64)	(1.80, 3.01)
χ^2_{D-7} (5%)	5.99	15.51	9.49	–	19.68	–
p-value	0.042	0.000	0.541	–	0.000	–
J-test (5%)	reject H_0	reject H_0	not reject H_0	–	reject H_0	–

Note: The constraints for optimization are given in ⟨⟩ brackets. The sample means based on 500 random runs are reported, while the 95% confidence intervals of the sample estimates are reported in () parentheses. The figures are rounded to 2 or 3 decimal places.

moments while the same number of model parameters is estimated. Still, we observe two very small values of optimized $J(\hat{\boldsymbol{\theta}})$, 1.68 and 2.39, which at least suggest a very good fit to the empirical dataset, although not statistically evaluable. Importantly, however, due to the inevitable misspecifications of any model, the credibility of $J(\hat{\boldsymbol{\theta}})$ statistics derived from alternative moments should not be considered inferior to those based on the efficient moments. Finally, the only set strongly rejected by the J-test is the overall best set #239 identified by FSMS that includes 18 moments. The empirical results for this set highlight the potential issue of the large moment sets that can perform superbly in-sample, especially in simulations, but might ultimately deliver poor out-of-sample estimation performance.

Focusing on the means and 95% confidence intervals of the individual estimates, we generally observe smaller, compared to the past literature (Franke and Westerhoff, 2012; 2016), but statistically significant fundamentalists' adjustment parameters ϕ. The difference is not surprising as our empirical dataset also covers the Global Financial Crisis of 2007–2008, the resulting Great Recession period, the COVID-19 recession, and the ongoing Russian invasion of Ukraine. Next, the models generally rejected as a good approximation of the data generation process are associated with nearly insignificant chartists' adjustment parameters χ, whose challenging estimation nature has been revealed already in Section 6.7. The volatility of the fundamentalists' noise term, σ_f, keeps the magnitude similar to estimates reported in the past. However, the other volatility associated with chartists, σ_c, roughly doubles, contributing to the consistent story of a less stable stock market in the last two decades. Finally, while α_n and α_p representing the herding and price misalignment parameters, respectively, are again estimated comparably to the past literature, the predisposition constant α_0 is only statistically significantly negative for the *CHL15* benchmark, the overall best set #187 (BSME), and the effective set #118 (FSMS).

8. Discussion

In the final discussion, we first consider the relationship between the complexity of the models and the size of the identified moment sets and compare and contrast the specifics of the two algorithms. Next, we study particular moments prioritized by both procedures across all models, and finally, we summarize our recommendations for practical SMM applications.

8.1. Sizes and overlaps of the identified moment sets

The main analysis focuses on the connection between the number of moments and the resulting SMM estimation performance. However, an important link also arises between the number of parameters and the size of the suggested moments sets. Table 5 displays sizes of the overall best and effective moment sets. One immediately notices a generally positive relationship. Moreover, for all three models, the sets chosen by BSME are smaller/not larger than those identified by FSMS. The difference stems from the very nature of the two algorithms and a greater ability of BSME to take advantage of the interplay between specific moments. While in the first rounds of the selection process, FSMS is forced to choose the moments according to their solo performance, for BSME, such a complex interplay is considered from the very beginning of the process of

Table 5
Moment sets: sizes, intersections, unions.

	BSME	FSMS	$\bigcap$	$\bigcup$
Markov-switching multifractal model—2 parameters				
Overall best set	3	8	3	8
Effective set	2	2	2	2
Alfarano et al.(2008)—3 parameters				
Overall best set	10	12	10	12
Effective set	5	7	5	7
Franke and Westerhoff (2012)—7 parameters				
Overall best set	11	18	10	19
Effective set	7	7	3	11

Note: BSME and FSMS columns correspond to the size of the moment set identified by the algorithm, $\bigcap$ and $\bigcup$ to the number of moments of the intersection and union of the effective moment sets. Numbers in parentheses indicate the $\text{RMSE}_{\hat{\theta}}$ of the given intersection/union.

gradual elimination of the full moment set. In other words, parameters identified as crucial in the first rounds of the FSMS procedure can be found to be less important or even redundant from the perspective of the BSME search if a combination of some already included moments largely compensates for their absence.

The same argumentation explains the different patterns in the two columns of Fig. 1 for the two agent-based models. For BSME, the step-by-step elimination process results in a greater estimating efficiency until the last rounds, followed by large jumps in the $\text{RMSE}_{\hat{\theta}}$ for the smallest moment sets. In contrast, the FSMS expanding set leads to a hyperbolic-like shape of the imaginary efficiency curve, which gradually decreases and reaches relative flatness only in the later stages of the selection process.

The numbers of intersecting moments indicate that both algorithms prioritize very similar moment conditions. We even observe that for the MSM and Alfarano et al. (2008) models, the sets identified by BSME are exact subsets of those suggested by FSMS. The two overall best sets almost perfectly overlap also for the Franke and Westerhoff (2012) model and share a total of 10 moments. On the contrary, both effective sets, including seven elements, only share three crucial moments, the indicators of the overall volatility and a measure of heavy tails (ABS-MEAN, RAW-VAR, and ABS-HILL5), while the composition of autocorrelations differs. A potential absence of some specific moment, thus, does not seem to matter much for this model due to the interplay between other moments that can fully compensate for it.

8.2. Moment composition across models

The previous section concludes that the suggested moment sets share many elements for each model, even though their sizes vary. However, we have not considered the particular composition of the moment sets across all models from an aggregate perspective. For clarity, we only focus on the effective moment sets that are, by definition, always the subsets of the overall best sets and, as such, contain only the key moments. The evolution of their compositions is summarized in bold in Table 2 that allows for a straightforward analysis of the moment preference stability across the two algorithms and the overall moment importance across the three modeling approaches.

At first glance, a densely populated upper part of Table 2 shows that key measures of heavy tails of the returns' distribution, together with standard 'scale' moments capturing overall volatility (RAW-VAR, ABS-MEAN), are crucial for the SMM performance. Importantly, the benchmark sets summarized in Table 1 always include only one of the possible measures for a given phenomenon. In contrast, three effective sets for agent-based models include both RAW-KURT and one of the Hill estimators, and all six effective sets include both RAW-VAR and ABS-MEAN. Next, focusing on the measures of clustered volatility and potentially long-memory patterns in the middle and bottom parts, autocorrelations of absolute returns are generally more critical than those of squared returns for financial agent-based models. Also, while some high-order autocorrelations are essential for capturing the phenomenon of volatility clustering, their actual composition does not matter crucially due to mutual interactions between the moments that compensate for the impact of one particular.

An additional seemingly surprising observation valid across all three models is that the measures of one of the most pronounced stylized facts, the unpredictability of raw returns, are generally not crucial for the performance of SMM. Autocorrelations of raw returns up to the third lag (RAW-AC1, RAW-AC2, and RAW-AC3) make it to the composition of an effective set only in one of six cases, and even then, it is selected as its last element. One could, however, already have expected this behavior based on the respective descriptive statistics regarding the AR(1) dynamics in Table A.6, where autocorrelation coefficients on the first lag are negligible. Thus, the values of RAW-AC1 to RAW-AC3 contribute negligibly also to the overall criterion distance function $J(\boldsymbol{\theta})$ in (2). We hypothesize that these moments would gain markedly more importance if the absence of autocorrelation in raw returns is violated in some other financial model. However, our testbed ones do replicate this essential empirical stylized fact well.

8.3. Recommendations for practical application

Overall, while the algorithms follow their specific paths, our extension identifies very similar key moment conditions for the simpler of the two testbed financial agent-based models using both procedures. For the more complex model, the two algorithms lead to the same core of the moment sets while the compositions of additional autocorrelations differ. Importantly, BSME generally prefers parsimonious sets that better avoid the problem of a superb in-sample performance potentially followed by a poor out-of-sample empirical fit, as observed in our empirical application.

Considering the results of the simulation analysis enriched by the findings of the resulting out-of-sample empirical application, our general recommendation for the potential users of our methodology is to follow the structure of Section 7. That means considering both overall best and effective sets identified by the two algorithms, provided their sizes are not too large. Next, contrast their empirical estimation performance to few moment sets used in the literature. The *J*-test of overidentifying restrictions provides a rigorous model comparison and validation guidance. When a suggested set seems unnecessarily large, typically produced by FSMS, it is recommended to prefer a smaller companion set, typically suggested by BSME.

Effective sets represent less complex parsimonious subset alternatives to the overall best moment sets, demonstrating comparable performance while decreasing the necessary computing budget. However, for models with many estimated parameters, they tend to be of the same size. This, unfortunately, inhibits rigorous statistical testing of the estimated model fit to the moment characteristics of the empirical dataset. Thus, if an effective set is at least one element larger, this is generally the most preferred option for a single empirical application.

9. Conclusion

This paper introduces machine learning feature selection methods to agent-based econometrics. We propose a machine learning extension of the SMM to expand the estimation methodology for analytically intractable financial agent-based models. The complexity or latency of some variables often prevents the application of standard econometric tools, such as the GMM or maximum likelihood. Therefore, simulated counterparts are traditionally employed instead. However, the most prominent alternative is the SMM, which faces several difficulties.

Specifying the moments used for the estimation and the potential arbitrariness of this choice are among the most concerning problems of the method. The moment sets are traditionally hand-crafted intuitively concerning the stylized facts of financial data. However, an insufficient number of moments may be provided to the estimation method, with vital parts of the model dynamics not captured. On the other hand, in practical applications to small samples, too many moments can complicate the estimation of the covariance matrix of the moment conditions. The proposed extension thus automates this choice. The main aim is to identify preferable moment sets for a given model yielding superior estimation performance.

The intuition behind our methodology follows one of the most straightforward and earliest applications of machine learning in econometrics. Specifically, it revolves around the techniques defined within the stepwise regression framework. Instead of model selection, we apply them to select the set of moments used by the SMM. We propose two algorithms, backward stepwise moment elimination (BSME) and forward stepwise moment selection (FSMS). Subsequently, we test them on three models with increasing complexity. The testbed models include the discrete-time regime-switching MSM and two popular financial agent-based frameworks. The results are compared with estimates generated using three benchmark moment sets standardly used in the literature and constructed specifically for the given models.

We find that the two algorithms consistently detect the most important moments for every model. The moment sets identified by our methodology achieve a substantial performance gain over the benchmarks. Consequently, we apply these preferable sets in an out-of-sample empirical exercise to successfully estimate the Franke and Westerhoff (2012) model using S&P 500 data.

As a follow-up to this work, if the full moment set was appropriately expanded, then an application study of our methodology to small-scale macroeconomic agent-based models could be conducted as well. Moreover, other specifications of the selection techniques might be used to boost its performance. For instance, stepwise selection of moment couples could be applied; however, the considerably increased computational burden must be taken into account.

Availability of data and materials

The empirical dataset collected and analyzed during the current study and a sample Julia code for an illustrative replication of the results are available in the GitHub repository: github.com/jirikukacka/Zila_Kukacka_Moment_Selection [created 2022-10-11].

Authors' contributions

Both authors contributed equally to this manuscript.

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Eric Zila: Methodology, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jiri Kukacka:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Resources, Supervision, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgments

This paper benefited greatly from Reiner Franke's detailed comments and research discussions. We are also indebted to the participants of the EAEPE Conference 2020, the ERFIN Workshop 2021, the ICNTEF Conference 2020 and 2021, the EcoSta Conference 2021, the FMND Workshop 2022, the Workshop on Computational and Experimental Economics 2022, the Behavioral Macroeconomics Workshop 2022, and the Model Evaluation and Causal Search Workshop 2022 for their many useful suggestions, as well as to Sergey Bolshakov and Jan Hanzal for their excellent research assistantship. Finally, we thank the three anonymous reviewers for their inspiring comments and detailed recommendations.

Jiri Kukacka gratefully acknowledges the financial support from the Czech Science Foundation under the project 'Linking financial and economic agent-based models: An econometric approach' [Grant number 20-14817S] and from the Charles University UNCE program [Grant number UNCE/HUM/035]. This work was supported by the Cooperatio Program at Charles University, research area Economics.

Appendix A. Model outputs and descriptive statistics

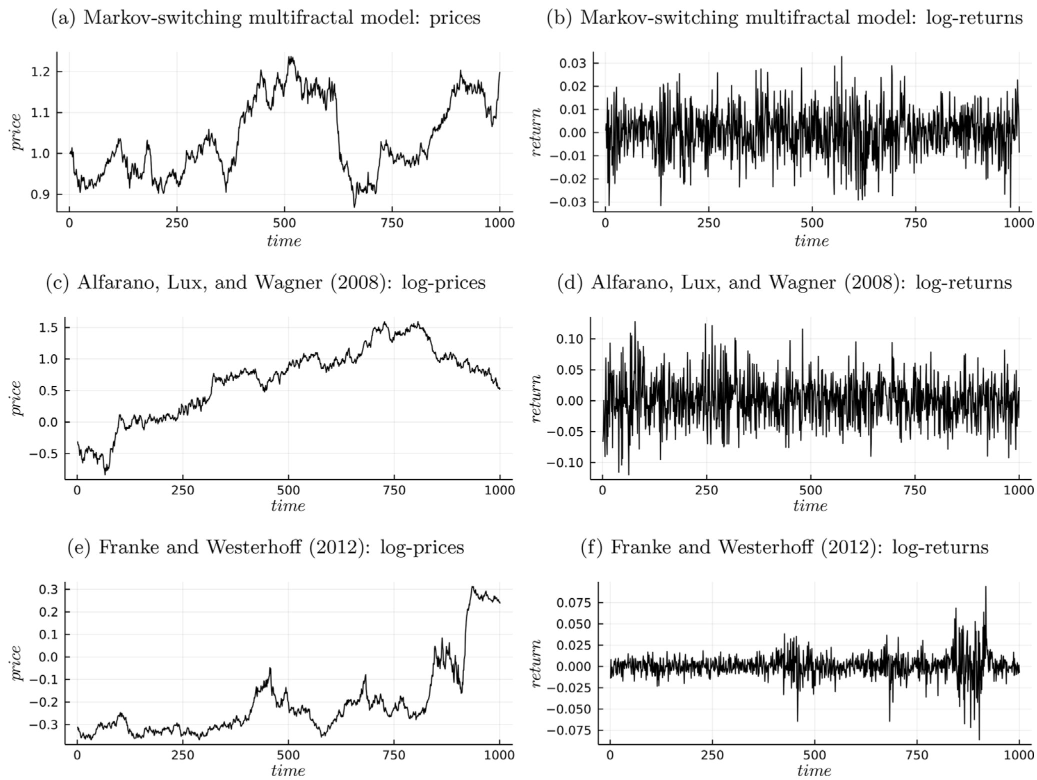


Fig. A.3. Model outputs.
Note: The figure displays typical time series outputs of the testbed models under the general simulation setup. For visual clarity, only a period of length 1000 is used for generation.

Table A.6
Descriptive statistics.

	MSM	ALW (2008)	FW (2012)
Mean	0.00	0.00	0.00
	(−0.02;0.02)	(0.00;0.00)	(0.00;0.00)
SD	1.00	0.04	0.01
	(0.94;1.05)	(0.04;0.05)	(0.01;0.01)
Skew	0.00	0.00	0.00
	(−0.10;0.10)	(−0.10;0.11)	(−0.15;0.16)
Ex Kurt	1.08	0.88	1.79
	(0.76;1.49)	(0.51;1.24)	(1.29;2.31)
$\widehat{AR}(1)$	0.00	0.00	0.01
	(−0.03;0.03)	(−0.03;0.03)	(−0.02;0.04)
L-B	8.7%	9.7%	15.6%
J-B	100.0%	100.0%	100.0%
ADF	100.0%	100.0%	100.0%
KPSS	4.8%	2.2%	0.0%
G2012	6.1;6.6;6.0;4.8%	6.3;5.4;5.5;4.0%	4.7;4.6;5.2;5.6%
DGG2020	5.0%	4.6%	6.2%

Note: 'MSM' is the Markov-switching multifractal model, 'ALW(2008)' is the Alfarano et al. (2008) model, and 'FW(2012)' is the Franke and Westerhoff (2012) model. The averaged results are based on 1000 random runs and the general simulation setup. The resulting 95% sample confidence intervals are reported in parentheses. $\widehat{AR}(1)$ reports the estimated AR(1) coefficient based on an ARIMA (1,0,0) model. Rejection rates based on a 5% significance level are reported for the statistical tests: H_0 for the Ljung-Box test (L-B) is 'no autocorrelation on the first lag,' H_0 for the Jarque-Bera test (J-B) is 'normality,' H_0 for the Augmented Dickey-Fuller test (ADF) with constant is 'unit root presence/covariance nonstationarity,' H_0 for the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) is 'level stationarity,' H_0 for the Grazzini (2012) test (G2012) is 'ergodicity' based on the first four moments, and H_0 for the Delli Gatti and Grazzini (2020) test (DGG2020) is 'ergodicity'. The figures are rounded to one or two decimal places.

Appendix B. Machine learning evolutions

The following evolutions are coded according to labels in Table 1:

- Markov-switching multifractal model:
  - BSME: ABS-AC15 → RAW-AC1 → RAW-AC3 → SQR-AC15 → RAW-KURT → SQR-AC5 → ABS-AC10 → SQR-AC20 → SQR-AC1 → SQR-AC10 → ABS-HILL2.5 → ABS-AC50 → RAW-AC2 → ABS-AC5 → ABS-AC100 → ABS-AC1 → ABS-AC25 → ABS-AC20 → ABS-HILL5 → (beginning of the overall best set #246) SQR-AC25 → (beginning of the effective set #250) ABS-MEAN → RAW-VAR,
  - FSMS: RAW-VAR → ABS-MEAN (end of the effective set #27) → SQR-AC20 → SQR-AC25 → SQR-AC10 → SQR-AC5 → ABS-HILL5 → ABS-AC5 (end of the overall best set #140) → RAW-AC3 → ABS-AC25 → ABS-AC100 → RAW-AC1 → SQR-AC1 → ABS-AC15 → ABS-HILL2.5 → ABS-AC20 → ABS-AC50 → ABS-AC1 → RAW-KURT → SQR-AC15 → RAW-AC2 → ABS-AC10.
- Alfarano et al. (2008):
  - BSME: ABS-AC10 → ABS-AC15 → SQR-AC20 → ABS-AC20 → ABS-AC1 → ABS-AC50 → RAW-AC3 → SQR-AC25 → ABS-AC5 → SQR-AC15 → SQR-AC10 → RAW-AC1 → (beginning of the overall best set #190) RAW-AC2 → ABS-HILL5 → SQR-AC1 → ABS-AC25 → SQR-AC5 → (beginning of the effective set #238) ABS-HILL2.5 → ABS-AC100 → ABS-MEAN → RAW-KURT → RAW-VAR,
  - FSMS: ABS-AC100 → ABS-AC25 → RAW-VAR → RAW-KURT → ABS-MEAN → ABS-HILL2.5 → SQR-AC1 (end of the effective set #128) → SQR-AC5 → ABS-HILL5 → RAW-AC1 → RAW-AC2 → ABS-AC15 (end of the overall best set #192) → SQR-AC20 → RAW-AC3 → ABS-AC20 → SQR-AC15 → ABS-AC50 → SQR-AC10 → ABS-AC10 → SQR-AC25 → ABS-AC5 → ABS-AC1.
- Franke and Westerhoff (2012):
  - BSME: ABS-AC5 → RAW-AC2 → ABS-AC50 → SQR-AC15 → RAW-KURT → SQR-AC25 → ABS-AC1 → ABS-AC15 → RAW-AC1 → ABS-AC20 → SQR-AC20 → (beginning of the overall best set #187) SQR-AC5 → SQR-AC1 → ABS-HILL2.5 → ABS-AC100 → (beginning of the effective set #224) SQR-AC10 → RAW-AC3 → ABS-AC10 → ABS-AC25 → ABS-MEAN → RAW-VAR → ABS-HILL5,
  - FSMS: SQR-AC20 → SQR-AC25 → ABS-MEAN → ABS-AC100 → RAW-KURT → ABS-HILL5 → RAW-VAR (end of the effective set #118) → ABS-HILL2.5 → ABS-AC25 → ABS-AC50 → RAW-AC1 → RAW-AC2 → SQR-AC10 → ABS-AC5 → ABS-AC10 → SQR-AC5 → ABS-AC1 → RAW-AC3 (end of the overall best set #239) → ABS-AC20 → ABS-AC15 → SQR-AC15 → SQR-AC1.

Appendix C. Evolution of the parameter estimates

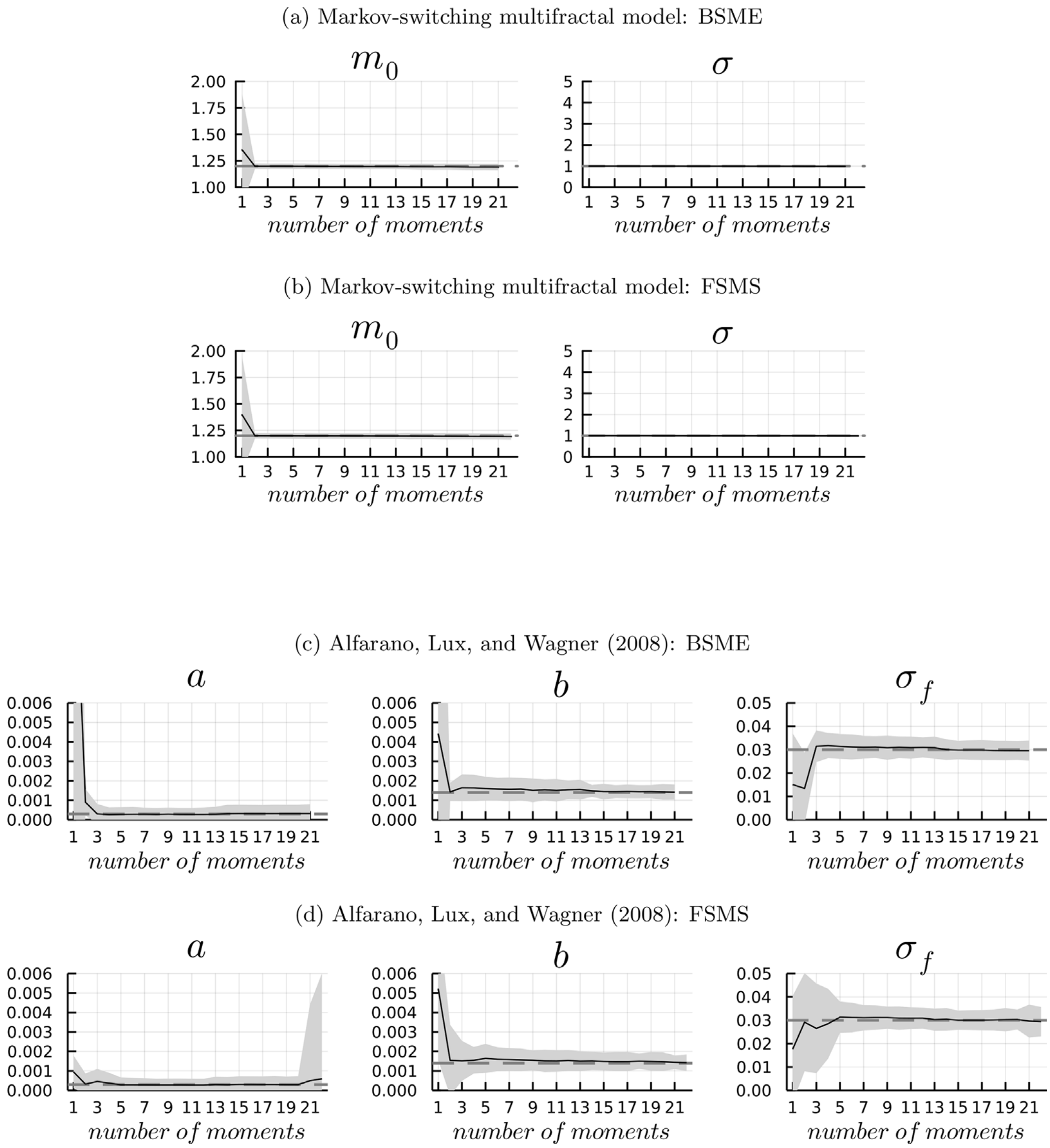
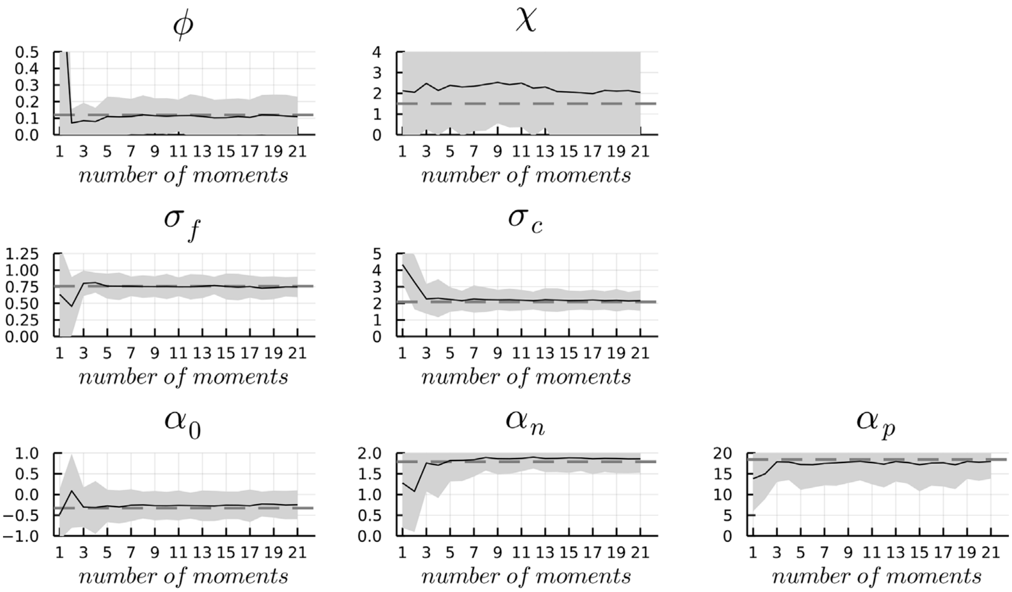


Fig. C.4. Evolution of the parameter estimates across machine learning rounds.
Note: The dashed horizontal line represents the true parameter value, the full line corresponds to its estimate for the best moment set of the respective round, and the gray area shows its 95% sample confidence band.

(a) Franke and Westerhoff (2012): BSME



(b) Franke and Westerhoff (2012): FSMS

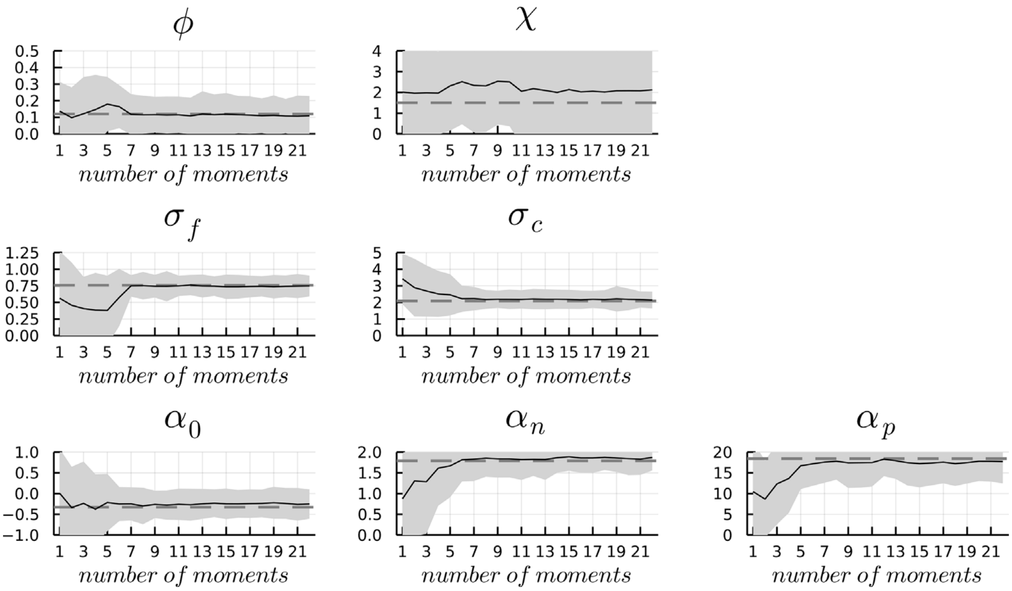


Fig. C.4. Continued

Appendix D. Alternative weighting matrices

Table D.7
Franke and Westerhoff (2012): alternative weighting matrices.

		FW9	*CHL15*
Identity matrix			
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.767	0.843
BSME—Effective set #224	0.732	4.6%	13.2%
FSMS—Effective set #118	0.814	−6.1%	3.4%
Diagonal overlapping block bootstrap weighting matrix			
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.610	0.521
BSME—Effective set #224	0.530	13.1%	−1.7%
FSMS—Effective set #118	0.549	10.0%	−5.4%
Non-overlapping block bootstrap weighting matrix			
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.580	0.671
BSME—Effective set #224	0.405	30.2%	39.6%
FSMS—Effective set #118	0.408	29.7%	39.2%
History sampling bootstrap weighting matrix			
Benchmark	$\text{RMSE}_{\hat{\theta}}$	0.557	0.443
BSME—Effective set #224	0.445	20.1%	−0.5%
FSMS—Effective set #118	0.409	26.6%	7.7%

Note: Estimation improvements compared to the benchmark moment sets are reported in %.

References

Alfarano, S., Lux, T., Wagner, F., 2005. Estimation of agent-based models: the case of an asymmetric herding models. Comput. Econ. 26, 19–49.
Alfarano, S., Lux, T., Wagner, F., 2006. Estimation of a simple agent-based model of financial markets: an application to Australian stock and foreign exchange data. Physica A 370 (1), 38–42.
Alfarano, S., Lux, T., Wagner, F., 2007. Empirical validation of stochastic models of interacting agents. Eur. Phys. J. B 55 (2), 183–187.
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.
Altonji, J.G., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. J. Bus. Econ. Stat. 14 (3), 353–366.
Bargigli, L., Riccetti, L., Russo, A., Gallegati, M., 2020. Network calibration and metamodeling of a financial accelerator agent based model. J. Econ. Interact. Coord. 15 (2), 413–440.
Boivin, J., Giannoni, M.P., 2006. Has monetary policy become more effective? Rev. Econ. Stat. 88 (3), 445–462.
Brock, W.A., Hommes, C.H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J. Econ. Dyn. Control 22, 1235–1274.
Calvet, L.E., Fisher, A.J., 2004. How to forecast long-run volatility: regime switching and the estimation of multifractal processes. J. Financ. Econom. 2 (1), 49–83.
Carrasco, M., Florens, J.P., 2002. Simulation-based method of moments and efficiency. J. Bus. Econ. Stat. 20 (4), 482–492.
Chen, S., Desiderio, S., 2022. A regression-based calibration method for agent-based models. Comput. Econ. 59 (2), 687–700.
Chen, S.-H., Chang, C.-L., Du, Y.R., 2012. Agent-based economic models and econometrics. Knowl. Eng. Rev. 27, 187–219.
Chen, Z., Lux, T., 2018. Estimation of sentiment effects in financial markets: a simulated method of moments approach. Comput. Econ. 52 (3), 711–744.
Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quant. Finance 1 (2), 223–236.
Cont, R., 2007. Volatility clustering in financial markets: empirical facts and agent-based models. In: Teyssiere, G., Kirman, A.P. (Eds.), Long Memory in Economics. Springer Berlin Heidelberg, pp. 289–309.
Day, R.H., Huang, W., 1990. Bulls, bears and market sheep. J. Econ. Behav. Organ. 14 (3), 299–329.
Delli Gatti, D., Grazzini, J., 2020. Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic agent based models. J. Econ. Behav. Organ. 178, 875–902.
Dieci, R., He, X.Z., 2018. Heterogeneous agent models in finance. In: Hommes, C., LeBaron, B. (Eds.), Handbook of Computational Economics. In: Handbook of Computational Economics, vol. 4. Elsevier, pp. 257–328.
Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61 (4), 929–952.
Fabretti, A., 2013. On the problem of calibrating an agent based model for financial markets. J. Econ. Interact. Coord. 8 (2), 277–293.
Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., Roventini, A., 2019. Validation of agent-based models in economics and finance. In: Beisbart, C., Saam, N.J. (Eds.), Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives. Springer International Publishing, pp. 763–787.
Farmer, J.D., Joshi, S., 2002. The price dynamics of common trading strategies. J. Econ. Behav. Organ. 49 (2), 149–171.
Franke, R., 2009. Applying the method of simulated moments to estimate a small agent-based asset pricing model. J. Empir. Finance 16, 804–815.
Franke, R., 2018. Competitive moment matching of a New-Keynesian and an Old-Keynesian model. J. Econ. Interact. Coord. 13 (2), 201–239.
Franke, R., Jang, X.-S., Sacht, S., 2015. Moment matching versus Bayesian estimation: backward-looking behaviour in a New-Keynesian baseline model. North Am. J. Econ. Finance 31, 126–154.
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.
Franke, R., Westerhoff, F., 2016. Why a simple herding model may generate the stylized facts of daily returns: explanation and estimation. J. Econ. Interact. Coord. 11 (1), 1–34.
Ghonghadze, J., Lux, T., 2016. Bringing an elementary agent-based model to the data: estimation via GMM and an application to forecasting of asset price volatility. J. Empir. Finance 37, 1–19.

Gilli, M., Winker, P., 2003. A global optimization heuristic for estimating agent based models. Comput. Stat. Data Anal. 42, 299–312.
Grazzini, J., 2012. Analysis of the emergent properties: stationarity and ergodicity. J. Artif. Soc. Social Simul. 15 (2), 7.
Grazzini, J., Richiardi, M., 2015. Estimation of ergodic agent-based models by simulated minimum distance. J. Econ. Dyn. Control 51, 148–165.
Grazzini, J., Richiardi, M.G., Tsionas, M., 2017. Bayesian estimation of agent-based models. J. Econ. Dyn. Control 77, 26–47.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.
Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50 (5), 1269–1286.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. second ed., New York.
Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. J. Finance 54 (6), 2143–2184.
Jalali, M., Rahmandad, H., Ghoddusi, H., 2015. Using the Method of Simulated Moments For system Identification. MIT Press, Cambridge, MA.
Jang, T.-S., Sacht, S., 2016. Animal spirits and the business cycle: empirical evidence from moment matching. Metroeconomica 67 (1), 76–113.
Jang, T.-S., Sacht, S., 2021. Forecast heuristics, consumer expectations, and New-Keynesian macroeconomics: a horse race. J. Econ. Behav. Organ. 182, 493–511.
Kirman, A., 1991. Epidemics of opinion and speculative bubbles in financial markets. In: Taylor, M. (Ed.), Money and Financial Markets. Macmillan, New York, USA, pp. 354–368.
Kirman, A., 1993. Ants, rationality, and recruitment. Q. J. Econ. 108 (1), 137–156.
Kocherlakota, N.R., 1990. On tests of representative consumer asset pricing models. J. Monet. Econ. 26 (2), 285–304.
Kukacka, J., Kristoufek, L., 2020. Do 'complex' financial models really lead to complex dynamics? Agent-based models and multifractality. J. Econ. Dyn. Control 113, 103855.
Kukacka, J., Kristoufek, L., 2021. Does parameterization affect the complexity of agent-based models? J. Econ. Behav. Organ. 192, 324–356.
Lamperti, F., Roventini, A., Sani, A., 2018. Agent-based model calibration using machine learning surrogates. J. Econ. Dyn. Control 90, 366–389.
Lee, B.-S., Ingram, B.F., 1991. Simulation estimation of time-series models. J. Econom. 47 (2), 197–205.
Lee, J.S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, G., Sun, Z., Parker, D.C., 2015. The complexities of agent-based modeling output analysis. J. Artif. Soc. Social Simul. 18 (4), 4. doi:10.18564/jasss.2897.
Lux, T., 2008. The Markov-switching multifractal model of asset returns. J. Bus. Econ. Stat. 26 (2), 194–210.
Lux, T., 2009. Chapter 3: stochastic behavioral asset-pricing models and the stylized facts. In: Hens, T., Schenk-Hoppé, K.R. (Eds.), Handbook of Financial Markets: Dynamics and Evolution. Handbooks in Finance, San Diego: North-Holland, pp. 161–215.
Lux, T., 2018. Estimation of agent-based models using sequential Monte Carlo methods. J. Econ. Dyn. Control 91, 391–408.
Lux, T., 2022a. Approximate Bayesian inference for agent-based models in economics: a case study. Stud. Nonlinear Dyn. Econom. doi:10.1515/snde-2021-0052.
Lux, T., 2022b. Bayesian estimation of agent-based models via adaptive particle Markov chain Monte Carlo. Comput. Econ. 60, 451–477.
Lux, T., 2022c. Inference for nonlinear state space models: acomparison of different methods applied to Markov-switching multifractal models. Econom. Stat. 21, 69–95.
Lux, T., Zwinkels, R.C., 2018. Empirical validation of agent-based models. In: Hommes, C., LeBaron, B. (Eds.), Handbook of Computational Economics. In: Handbook of Computational Economics, vol. 4. Elsevier, pp. 437–488.
Mandelbrot, B.B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. J. Fluid Mech. 62 (2), 331–358.
Mandes, A., Winker, P., 2017. Complexity and model comparison in agent based modeling of financial markets. J. Econ. Interact. Coord. 12 (3), 469–506.
Manzan, S., Westerhoff, F.H., 2007. Heterogeneous expectations, exchange rate dynamics and predictability. J. Econ. Behav. Organ. 64, 111–128.
Mao, C.S., 1990. Hypothesis Testing and Finite Sample Properties of Generalized Method of Moments Estimators: A Monte Carlo Study. Working Paper. Federal Reserve Bank of Richmond.
McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57 (5), 995–1026.
Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.
Platt, D., 2020. A comparison of economic agent-based model calibration methods. J. Econ. Dyn. Control 113, 103859.
Platt, D., 2022. Bayesian estimation of economic simulation models using neural networks. Comput. Econ. 59 (2), 599–650.
Salle, I., Yıldızoğlu, M., 2014. Efficient sampling and meta-modeling for computational economic models. Comput. Econ. 44 (4), 507–536.
Schmitt, N., Westerhoff, F., 2017a. Herding behaviour and volatility clustering in financial markets. Quant. Finance 17 (8), 1187–1203.
Schmitt, N., Westerhoff, F., 2017b. Heterogeneity, spontaneous coordination and extreme events within large-scale and small-scale agent-based financial market models. J. Evol. Econ. 27 (5), 1041–1070.
Schwartz, I., & Kirstein, M. (2022). Time is Limited on the Road to Asymptopia. Technical report.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. 58 (1), 267–288.
Tubbenhauer, T., Fieberg, C., Poddig, T., 2021. Multi-agent-based VaR forecasting. J. Econ. Dyn. Control 131, 104231.
Vandin, A., Giachini, D., Lamperti, F., Chiaromonte, F., 2022. Automated and distributed statistical analysis of economic agent-based models. J. Econ. Dyn. Control 143, 104458.
Welch, P.D., 1983. The statistical analysis of simulation results. In: The Computer Performance Modeling Handbook, vol. 22. Academic Press, Inc., pp. 268–328.
Winker, P., Gilli, M., Jeleskovic, V., 2007. An objective function for simulation based inference on exchange rate data. J. Econ. Interact. Coord. 2 (2), 125–145.
Zegadło, P., 2021. Efficient calibration of a financial agent-based model using the method of simulated moments. In: Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (Eds.), Computational Science – ICCS 2021. Springer International Publishing, Cham, pp. 316–329.
Zhang, W., Valencia, A., Chang, X.B., 2023. Synergistic integration between machine learning and agent-based modeling: a multidisciplinary review. IEEE Trans. Neural Netw. Learn. Syst. 34 (5), 2170–2190.