Estimation of a Structural Stochastic Volatility Model of Asset Pricing

Reiner Franke [a,*] *Frank Westerhoff* [b]

November 2009

[a]University of Kiel, Germany
[b]University of Bamberg, Germany

Abstract

The paper estimates an elementary agent-based financial market model recently put forward by the same authors. Invoking the two trader types of fundamentalists and chartists, it comprises four features: price determination by excess demand; a herding mechanism that gives rise to a macroscopic adjustment equation for the population shares of the two groups; a rush towards fundamentalism when the price misalignment becomes too large; and, finally, differently strong noise components in the demand per chartist and fundamentalist trader, which implies a structural stochastic volatility in the returns. The estimation is performed using the method of simulated moments. Combining it with bootstrap and Monte Carlo methods, it is found that the model cannot be rejected by the empirical daily returns from a stock market index and a foreign exchange rate. Measures of the matching of the single moments are satisfactory, too, while the behavioural parameters are well identified and are able to discriminate between the two markets.

JEL classification: D84; G12; G14; G15.

Keywords: Method of simulated moments; daily returns; autocorrelation patterns; fundamentalist and technical trading.

1. Introduction

Over the past ten or twenty years, a rich literature on agent-based asset pricing models has developed as an alternative to the rational expectations framework. Even simple heuristic strategies of two (or a few) groups of traders have proven to be quite successful in generating dynamic phenomena that share central characteristics with the time series

from real financial markets. Numerical investigations have recently tended to become the rule, and an increasing number of these models are now not only simulated to illustrate their basic mechanisms, but they also attempt to match certain stylized facts in quantitative ways. [1] Extending research in the latter direction appears to be particularly important since it promises progress in coping with the problem of the arising anecdotal wilderness within the bounded rationality paradigm.

In many cases, the models use a calibration methodology in the following (limited) sense: the researcher selects a number of summary statistics for the prices or (mostly) returns of a set of financial markets, presents numerical values for the model's parameters, shows a few sample runs, and finally points out that the model-generated statistics are quantitatively of a similar order of magnitude to the empirical ones. In this way, a model can be considered validated. More generally, the analysis enables bad models to be sorted out from the good ones. [2] However, it may also be desirable to distinguish between good and "very good" models, so that, for example, future research may prefer to concentrate on refining or generalizing the better of these models. This higher ambition for judging a model, then, leads us from the still more or less informal calibration to the issue of estimation.

In the present context, estimation in the first instance means the existence of a single criterion that evaluates a model's performance under alternative sets of numerical parameter values. In other words, estimation is based on an objective function that is to be optimized across different parameters. In this paper we are concerned with a simulation-based objective function, the outcome of which is also directly comparable across quite different (and non-nested) models. The corresponding estimation procedure is known as the method of simulated moments (MSM). [3] It is, in fact, a more rigorous version of what we had roughly described as calibration, which refers to several summary statistics or, synonymously, "moments" from the model simulations. The idea that these "moments" should come close to their empirical counterparts is captured by an objective function that, briefly speaking, simply adds up the deviations of the simulated moments from the empirical ones. Each of these components is, however, weighted by a factor determined by the empirical characteristics of the corresponding moment. [4]

As an estimation method, MSM is frequently criticized because of its arbitrary choice of moments in the criterion function. For this reason, methods such as indirect inference or the efficient method of moments have been developed where, in effect, a set of moment

---

[1] A recent overview by Chen at al. (2008) refers to about 50 such models.
[2] This is not to say that these "bad" models might not have other merits.
[3] Opinions seem to be divided over whether to call this method the 'method of simulated moments' or the 'simulated method of moments'.
[4] More exactly, as explained later on, the objective function is a quadratic function of the moment deviations. Entering it is a weighting matrix derived from the data.

conditions is endogenously defined by means of an auxiliary model.[5] It may, however, be argued that this does not completely resolve the problem of arbitrariness, or judgement; it is only shifted from the choice of moments to the choice of the auxiliary model. Taking into account that structural models are necessarily misspecified, one may furthermore prefer MSM to the other methods because of its higher degree of transparency. After all, a model is not meant to resemble or mimic any actual economy or financial market; the best we can expect from it is that it matches *some* of the 'stylized facts'. While MSM requires the researcher to make up his or her mind about the dimensions along which the model should be most realistic, the implications of a decision for a particular auxiliary model are usually less clear. Besides the fact that in certain cases MSM is possibly faster or more robust than other methods, and in addition to the option of comparing its performance across different models, we consider the explicitness of MSM and its easy interpretation as the main argument for its use in estimation.

The short characterization of MSM and the accentuation of its merits brings us to the question of what specific model to fit, so that we can learn more about the method and the model. The present authors have put forward a small agent-based model (Franke and Westerhoff, 2009) which ultimately was calibrated in the way indicated above. Since its matching of some of the central moments of the daily returns of the S&P 500 stock market index looked so good to us, we decided it was time to put the model to a more stringent test, that is, to estimate it by MSM. As will be seen, the method allows us to pin down the range of suitable parameter values, to measure the model's overall goodness-of-fit by a single number, and to compare the relative fit of the single moments to each other. Apart from highlighting the generally good properties of our model, the quantitative figures we thus obtain may be taken as a benchmark that other models of similar complexity will still have to meet or, accepting it as a challenge, to surpass.

To put the aspect of a future model contest in a more specific perspective, we would like to point out the main innovative feature of our model, which we called structural stochastic volatility (SSV) and which was identified in Franke and Westerhoff (2009) as the pivotal mechanism leading to the good results. Since SSV can be easily transferred to other models (albeit with a similar philosophy), this device may prove equally worthwhile to them or, regarding the matching of the stylized facts, even superior.[6]

To give a brief impression of this concept of structural stochastic volatility, it may be mentioned that the model, as so many others, stipulates fundamentalism and chartism as two simple trading strategies between which the agents can switch on a day-to-day basis

---

[5] Carrasco and Florens (2002) provide a succinct overview of MSM and the other two aforementioned methods.

[6] Without placing particular emphasis on it, SSV has already been employed in a richer and, in some other specification details, quite different model by Westerhoff (2008). Although the model is "only" calibrated and moment matching is not thoroughly discussed, the diagrams shown look very promising with respect to estimation.

(according to a herding mechanism that is modified by a rush towards fundamentalism when the price misalignment becomes too large). The average demand per agent in each group, however, is split into a deterministic "core demand" and a random term, which is supposed to capture the heterogeneity within the groups. Given that the two noise terms differ in their variance, variations of the population shares will induce variations in the noise level of total demand, which in turn determines the change in prices. The model's stochastic component is therefore akin to a GARCH-type of modelling and should produce similar effects in the returns, though it is not just technically imposed but is derived from a structural argument.

Returning to the approach of the method of simulated moments, it seems that it has found application in only a few estimations of asset pricing models.[7] On the one hand, there are several papers by a research group around M. Gilli and P. Winker (see Gilli and Winker, 2003; Winker et al., 2007). Although their work is methodologically interesting, the models they consider are not yet very well suited to reproduce the desired stylized facts (at least not with the authors' exogenous setting of the other model parameters that were not estimated). On the other hand, Franke (2009) tested a more appropriate model by Manzan and Westerhoff (2005) with MSM and obtained quite encouraging results, showing the great potential of the estimation concept. The present paper distinguishes itself from this study in two respects. First, the SSV modelling framework is more flexible and has a wider scope than the somewhat special threshold mechanism in Manzan and Westerhoff. Second, Franke (2009) relied on the asymptotic econometric theory of MSM, whereas here we use bootstrap methods that can take better account of the small-sample properties of the estimations and that also give us more specific information about the model's general goodness-of-fit.[8]

The remainder of this paper is organized as follows. The next section briefly recapitulates the SSV model from Franke and Westerhoff (2009). Section 3 presents the MSM approach in its bootstrap version and then introduces a benchmark parameter set, which we decide to be our final estimation result for the daily returns of the S&P 500 index. By comparing the bootstrap distribution of the objective function to the corresponding distribution from a battery of simulations of the model, Section 4 arrives at an assessment of the overall fit of the model. In particular, it finds that the model cannot be rejected by the empirical data. Section 5 goes into finer detail and evaluates the matching of the single moments. Turning to the issue of the identification of the six parameters to be estimated, Section 6 computes their one- and two-dimensional confidence regions. Section 7 uses the same investigation methods to estimate an alternative financial mar-

---

[7] For a recent overview of the use of estimation methods in agent-based asset pricing models, see Chen et al. (2008; especially Table 3 on p. 22).

[8] We do not put 'small-sample' in inverted commas since, although our samples over more than 25 years of daily data may appear huge from a macroeconomic point of view, it will be seen that the bootstrapped distributions of the single moments are still not very narrow.

ket, namely the US dollar / German mark exchange rate, and obtains markedly different results. Section 8 concludes.

2. Formulation of the model

In this section, we briefly reiterate the financial market model put forward and analyzed in Franke and Westerhoff (2009). Asset price changes are determined by excess demand on a market populated by two types of speculative traders—fundamentalists and chartists. Fundamentalists have long time horizons and base their demand on the differences between the current price and the fundamental value. Even though they might expect the gap between the two prices to widen in the immediate future, they do not trade on the likeliness of this event and rather choose to place their bets on an eventual rapprochement. Chartists, on the other hand, have a short-term perspective and bet on the most recent price movements, buying (selling) if prices have been rising (falling). However, the agents are allowed to switch from one type to the other, where their choice is governed by a herding mechanism combined with the evaluation of the most recent price levels.

Let us start with the demand for the asset.[9] We join numerous examples in the literature and, for the core of each group, postulate a demand per average trader that is kept as simple as possible. For the fundamentalists, it is inversely related to the deviations of the price from its fundamental value. That is, in period t their core demand is proportional to the gap $(p^\star - p_t)$, where p_t is the log price of the asset known at the beginning of the period, and $p^\star$ is the (constant) fundamental log value. The core demand of the group of chartists is hypothesized to be proportional to the returns they have just observed, i.e. $(p_t - p_{t-1})$.

The innovative part of the model is that we add a noise term to *each* of these demand components (and not just their sum). The two terms are supposed to reflect the many individual digressions from the simple rules, which we do not wish to describe in full detail. This noise, which therefore captures a certain within-group heterogeneity, is conveniently represented by normally distributed random variables ε_t^f and ε_t^c for the fundamentalists and chartists, respectively. Combining the deterministic and stochastic elements, the average fundamentalist and chartist trader has the following net demand for the asset in period t:

$$d_t^f = \phi(p^\star - p_t) + \varepsilon_t^f \qquad \varepsilon_t^f \sim N(0, \sigma_f^2) \qquad \phi > 0,\ \sigma_f > 0 \qquad (1)$$

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

Total demand on the market (normalized by the population size) results from multiplying d_t^f and d_t^c by the population shares of the two groups.

---

[9] To be exact, by demand we mean the orders (positive or negative) per trading period, not the desired positions of the agents.

Since it will be important later in Section 7, we stress that the polar case $\chi = \sigma_c = 0$ also makes economic sense. Under these circumstances 'chartism' is tantamount to not trading at all. In other words, the agents would then choose between fundamentalist strategies and complete inactivity. [10]

Concerning the market fractions of the traders, let n_t^f and n_t^c be the number of fundamentalists and chartists. For the analysis of the model it is convenient to scale the population size with $2N$ and define $x_t := (n_t^f - n_t^c)/2N$ as the majority index of the fundamentalists. By construction, x_t is contained between -1 (all traders are chartists) and $+1$ (all traders are fundamentalists). It is easy to show that total excess demand is thus given by $(1+x_t)\,d_t^f/2 + (1-x_t)\,d_t^c/2$.

In general, the market is in disequilibrium. A market maker is assumed to absorb any excess of supply, and to serve any excess of demand from his inventory. He reacts to this imbalance by adjusting the price for the next period with a (constant) factor $\mu > 0$ in the direction of excess demand. Accordingly, the equation determining the price for the next period $t+1$ results from (1) and (2) as

$$p_{t+1} \;=\; p_t \;+\; \frac{\mu}{2}\left[(1+x_t)\,\phi\,(p^\star - p_t) \;+\; (1-x_t)\,\chi\,(p_t - p_{t-1}) \;+\; \varepsilon_t\right] \tag{3}$$

$$\varepsilon_t \;\sim\; N(0, \sigma_t^2)\,, \qquad \sigma_t^2 \;=\; \left[(1+x_t)^2\,\sigma_f^2 + (1-x_t)^2\,\sigma_c^2\right]/2 \tag{4}$$

Equation (4) is derived from the fact that the sum of the two normal distributions in (1) and (2), which are to be multiplied by the market fractions $(1 \pm x_t)/2$, is again normally distributed, with mean zero and the variance being equal to the sum of the two single variances. Obviously, if σ_f^2 and σ_c^2 are distinct, σ_t^2 will alter with the changes in the majority index x_t. This feature of a time-varying variance σ_t^2 in the returns (i.e. the log differences in prices) is what was coined *structural stochastic volatility* (SSV) in Franke and Westerhoff (2009).

It remains to set up the motions of the majority index x_t. To this end, we first introduce a so-called switching index $s = s_t$, which is uniform across the agents. The basic idea is that an increase of the switching index increases the probability of an individual chartist becoming a fundamentalist, and decreases the probability of an individual fundamentalist becoming a chartist. Assuming a large population (and two further natural specification details), it can be demonstrated that, at the macroscopic level, the adjustments of the majority index are governed by the deterministic equation,

$$x_{t+1} \;=\; x_t \;+\; \nu\left[(1-x_t)\exp(s_t) - (1+x_t)\exp(-s_t)\right] \tag{5}$$

where 'exp' is the exponential function and the (positive) coefficient ν may generally be characterized as a *flexibility parameter*. Hence, in a situation $x_t \le 0$ when chartists are not in the minority, the market moves towards fundamentalism if s_t is positive, i.e. x_t

---

[10] The situation for $\phi = 0$ and, possibly, $\sigma_f = 0$ would be formally analogous. In this case, however, the price dynamics would no longer be anchored in the fundamental value.

increases. Conversely, there is unambiguously a movement towards chartism if $x_t \geq 0$ and s_t is negative. If both x_t and s_t are equally signed, it depends on the precise circumstances in which direction the majority index changes. [11]

The switching index itself is specified as follows:

$$s_t = \alpha_o + a_x x_t + \alpha_d \cdot (p_t - p^\star)^2 \qquad (6)$$

Here, the coefficient α_o can be interpreted as a *predisposition parameter*, since in a balanced state $x_t = 0$ where also the other effects in (6) cancel out, a positive (negative) α_o would give rise to an aggregate motion in the direction of fundamentalism (chartism).

The other two parameters α_x and α_d are understood to be positive. Thus, the second term on the right-hand side of (6) captures the idea of herding: the more traders are already fundamentalists (i.e. the higher x_t), the higher the probability that also the remaining chartists convert to fundamentalism (and *vice versa*, since $x_t < 0$ if chartists are in the majority). The coefficient α_x may correspondingly be called a *herding parameter*. The third term in (6) measures the influence of price misalignment. It expresses the idea that when the price is further away from its fundamental value, "professionals tend more and more to anticipate" its "mean-reversion towards equilibrium" (Menkhoff et al., 2009, p. 251). In the present context, this means that the probability of becoming a fundamentalist rises so that, ultimately, with the overall increase in x_t, the price in (3) should be driven back to normal.

To sum up, the two central dynamic equations of the model are (i) the price adjustments (3), (4) with the structural stochastic volatility component σ_t^2, and (ii) the changes in the majority index x_t described in (5) and (6), which basically represent a herding dynamics curbed by a control for strong price misalignment. The system has an obvious recursive structure and can easily be iterated forward.

3. The estimation method

The model has been designed to explain—at least partially—the most important stylized facts of financial markets. [12] Referring to price changes at a daily frequency, we aim to check the four features that have received the most attention in the literature on agent-based models. These are the absence of autocorrelations in the raw returns, fat tails in their frequency distributions, volatility clustering, and long memory (see Chen et al., 2008, p. 19). The returns themselves are specified in percentage points, so we may write

---

[11] In more structural detail, the expressions $\nu \exp(\pm s_t)$ in (5) have the interpretation of individual transition probabilities. In the simulations below, these terms are therefore bounded from above by unity (so that, strictly speaking, ν should not be isolated outside the brackets). As also checked by the computer program, it is then no problem for the trajectories to stay away from the ± 1 boundaries.

[12] Detailed descriptions of the statistical properties of asset prices can be found in Cont (2001), Lux and Ausloos (2002), or Lux (2009).

$$r_t \; := \; 100 \cdot (p_t - p_{t-1}) \,, \qquad v_t \; := \; |r_t| \tag{7}$$

For a quantitative analysis, we measure these features by a number of summary statistics or, synonymously, moments. The first moment is the volatility of the returns, which we define as the mean value of the absolute returns v_t (here and in the autocorrelations below it makes no great difference whether one works with the absolute or squared returns). This is a (necessary) matter of scaling and, in the first instance, will have a bearing on the admissible general noise level in the model, as it is brought about by the two variances σ_f and σ_c. The second moment is the first-order autocorrelation of the raw returns. The requirement that it be close to zero should limit the chartists' extrapolations in the form of the parameter χ. On the other hand, we checked that if this moment is matched, the autocorrelations at the longer lags will all practically vanish, too. This lack of additional information is also the reason why we only use one moment of the raw returns.

Next, in order to capture the long memory effects, we invoke the autocorrelation function (ACF) of the absolute returns up to a lag of 100 days. As it is slowly decaying without becoming insignificant at the long lags, we have an entire profile to match. We view it as being sufficiently well represented by the six coefficients for the lags $\tau = 1, 5, 10, 25, 50, 100$. The sample variability involved in their computation is reduced by using the centred three-lag averages. [13] Lastly, the fat tail property is measured by the well-known Hill estimator of the tail index of the absolute returns, where the tail is conveniently specified as the upper 5 per cent. Thus, on the whole, we evaluate the performance of the model on the basis of nine moments, which we summarize in a (column) vector $m = (m_1, \ldots, m_9)'$ (the prime denotes transposition).

In Franke and Westerhoff (2009), a set of numerical parameter values was proposed and, referring to these moments, an illustrative sample run of the model demonstrated that the resulting return series was fairly close to the daily returns of the S&P 500 stock market index. In the present paper, we employ a formal estimation procedure to make the informal summary of "fairly close" more precise. As this issue has been put forward, and from the arguments mentioned in the Introduction, it is only natural for us to utilize the method of simulated moments.

For this purpose, we first of all need an objective function that allows us to define a distance between two moment vectors. It is given by a quadratic function, which is characterized by a weighting matrix $W \in \mathbb{R}^{9 \times 9}$ (to be specified shortly). Having a moment vector $m \in \mathbb{R}^9$ to compare to another reference vector $m^{ref} \in \mathbb{R}^9$, the function reads,

$$J \; = \; J(m, m^{ref}; W) \; := \; (m - m^{ref})' \; W \; (m - m^{ref}) \tag{8}$$

---

[13] That is, at lag τ the mean of the three autocorrelation coefficients for $\tau-1$, τ, $\tau+1$ is computed, except for $\tau = 1$, where it is the average of the first and second coefficient. It may also be noted that volatility clustering, which describes the tendency of large changes in the asset price to be followed by large changes, and small changes to be followed by small changes, is closely related to these long-term dependencies between the returns.

The weighting matrix takes the sampling variability of the moments into account. The basic idea is that the higher the sampling variability of a given moment i, the larger the differences between m_i and m_i^{ref} that can still be deemed insignificant. This can primarily be achieved by a correspondingly small diagonal element w_{ii}. In addition, the matrix W should provide for possible correlations between the single moments. An obvious candidate for a suitable weighting matrix is the inverse of an estimated variance-covariance matrix $\widehat{\Sigma}$ of the moments,

$$W \;=\; \widehat{\Sigma}^{-1} \tag{9}$$

One such estimate $\widehat{\Sigma}$ to set up W may be obtained from a Newey-West estimator of the long-run covariance matrix of the empirical moments (see, e.g., Lee and Ingram, 1991, p. 202, or the application of the method of simulated moments in Franke, 2009, Section 2.2). In this paper, we follow Winker et al. (2007) and choose a bootstrap approach to construct additional empirical samples of the moments, from which the covariances can then be derived.

More exactly, a block bootstrap is appropriate due to the long-range dependence in the series. An immediate idea for choosing the block length is one year, i.e. 250 days (Winker et al., 2007, Section 3.2). Thus we reduce our empirical return series of S&P 500 (which comprises 6866 data points) to $T = 6750$ observations, subdivide it into 27 fixed blocks of 250 days, and construct a new series, block by block, from 27 random draws (with replacement). Subsequently, all of the desired moments can be computed from this string of the blocks of returns.

Repeating this 5000 times, a frequency distribution for each of the moments is obtained. Ideally, the empirical moment should be more or less in the centre. For the ACF of the absolute returns at longer lags $\tau \geq 10$, however, the bootstrapped coefficients show a non-negligible tendency towards lower values. This bias suggests that the blocks in the bootstrap are too short. We therefore also tried longer blocks of 500 and 750 days. [14] For the latter it was indeed found that the bias was considerably mitigated. On the basis of these explorations of the bootstrap distributions, we decided to use a block length of 250 days for the first five "short memory" moments (the Hill estimator, volatility, first-order autocorrelation of the raw returns, and the autocorrelations at lags $\tau = 1$ and $\tau = 5$ for the absolute returns), and a block length of 750 days for the last four "long memory" moments (the ACF of $v_t = |r_t|$ at lags $\tau = 10, 25, 50, 100$).

The frequency distributions of these bootstrapped moments can be used to obtain an estimate $\widehat{\Sigma}$ of the moments' variance-covariance matrix. Formally, let $b = 1, \dots, B$ identify the bootstrap samples ($B = 5000$ presently), let $m^b = (m_1^b, \dots, m_9^b)'$ be the resulting moment vectors, and compute the vector of their mean values $\overline{m} := (1/B) \sum_b m^b$. From this, $\widehat{\Sigma}$ results like

---

[14] Unfortunately, blocks of 1000 days seem too long compared to the empirical sample period.

$$\widehat{\Sigma} \ := \ \frac{1}{B} \sum_{b=1}^{B} (m^b - \overline{m})(m^b - \overline{m})' \tag{10}$$

As the weighting matrix W in (9) is now well-defined, we can turn to an estimation of the model. Denote m_T^{emp} as the vector of the empirical moments from a return series of length T. Furthermore, let θ denote the vector of the model parameters to be estimated, and let $m(\theta; S)$ be the moment vector from a simulation of the model with a specific parameter set θ over S periods. Since long time horizons reduce the sample variability, we follow a common practice and choose $S = 10 \cdot T$. It is also understood that several hundred observations from an initial transition period have been discarded before, and that all simulations use the same random number sequence for the demand shocks. [15] On this basis, the estimation amounts to solving the following minimization problem across the admissible parameter vectors θ,

$$J[m(\theta; S), m_T^{emp}; W] \ = \ [m(\theta; S) - m_T^{emp}]' \, W \, [m(\theta; S) - m_T^{emp}] \ = \ \min_{\theta}! \tag{11}$$

We employ the Nelder-Mead simplex algorithm to carry out these optimizations (see Press et al., 1986, pp. 289–293). To avoid getting prematurely trapped in a local minimum, we repeatedly restart the procedure upon convergence several times with a relatively large new starting simplex which hopefully reaches beyond the "hills" surrounding such a local "valley", until no further noteworthy improvement in the minimization occurs. If in the course of the search some parameter θ_i violates a non-negativity constraint and becomes negative, we reset it to zero to run the simulation. However, after computing J from its moments, a penalty is added that proportionately increases with the extent of the violation.

Regarding the parameters, we can fix the price impact factor μ, the predisposition coefficient α_o, and the herding coefficient α_x. While a free choice of μ is an obvious matter of scaling, it was argued in Franke and Westerhoff (2009) when discussing the phase diagrams that the effects of variations of α_o and α_x could at least be qualitatively compensated by suitable changes in the other parameters (of α_d in the first instance). In fact, no compelling reason was found to try further improvements over the numerical values we decided to choose after a moderate number of initial explorations. Six parameters therefore remain to be estimated, namely,

$$\theta \ := \ (\phi, \chi, \nu, \alpha_d, \sigma_f, \sigma_c) \tag{12}$$

As already indicated, we try to fit the daily returns of the S&P 500, for which 6866 observations from January 1980 to mid-March 2007 are considered. Proceeding for the minimization (11) as described above (though in this case not too extensively) and rounding

---

[15] For the normally distributed ε_t with variance σ_t^2, this more precisely means that for each simulation run at time t the same random number $\tilde{\varepsilon}_t$ is drawn from the standard normal distribution $N(0,1)$ and ε_t is set as $\sigma_t \tilde{\varepsilon}_t$.

the solution to two digits after the decimal point, we concentrate on the result summarized in Table 1. In the rest of the discussion of the present estimation problem, we also refer to this vector as our benchmark parameter set θ^o.

ϕ	χ	μ	ν	α_o	α_x	α_d	σ_f	σ_c
0.18	2.35	0.01	0.57	-0.15	1.35	11.40	0.79	1.91

Table 1: Benchmark parameter set θ^o for S&P 500.

Given the underlying random number sequence, the value of the objective function brought about by θ^o is $J = 10.33$. Extending the minimization procedure to an interactive trial-and-error re-initialization of the search algorithm, two additional, and still superior, local minima have been found. One at $J = 8.00$ is achieved by $\phi = 0.195$, $\nu = 0.570$ and almost negligible changes in $\chi, \alpha_d, \sigma_f, \sigma_c$. The other is constituted by $\phi = 0.180$, $\nu = 0.634$ and minor changes in $\chi, \alpha_d, \sigma_f, \sigma_c$, which beats the first one with $J = 7.79$.

However, employing another random number sequence in the estimation is bound to lead to different results; in particular, the minimum value of J may well be lower than 7.79 or higher than 10.33. Hence an improvement in a single estimation of the kind just reported, which could only be detected with a certain additional effort, should not be overrated. The following analysis will rather focus on the question of how a specific parameter set, on which we finally have to settle for a succinct message, stands up to different random shocks. In this respect, the benchmark coefficients θ^o may be quite as good. Reference to θ^o might also emphasize that we take a broader view and do not rely on the results from a single estimation.

4. Assessment of the overall model fit

The process of bootstrapping the weighting matrix in the objective function J provides us with a large sample of empirically derived moments. If the weighting matrix is applied to their deviations from the moments actually observed, we also get a frequency distribution of values of J itself. From this a critical value of J can be established that, so it seems, allows us to assess the values of J that we obtain from the simulations of the model. The straightforward idea is that this bootstrap distribution proxies the distribution of J that would materialize if the hypothetical construct called the real-world data generation process (DGP) could produce a large number of different samples of return series. Accordingly, if a given model-generated series and its moments yield a value of J within the range of the bootstrapped values of J, we would not be able to tell it apart from a real-world series. Conversely, and taking account of rare extreme events with a significance level of 5%, a simulated return series will be found to be inconsistent with

11

the real-world data, and will therefore be rejected, if its J-value exceeds the 95% quantile of the bootstrapped J.

In the present case, we compute $J = J_{0.95} = 22.5$ for this critical value. However, it cannot not serve as a yardstick for the minimized J-values from (11) in the estimation procedure, where (as already mentioned at the end of the previous section) our benchmark parameter set θ^o gave rise to $J = 10.3$. Here it must not be forgotten that, just in order to reduce the sample variability, these returns were simulated over a time horizon that is ten times as long as the samples of the returns entering the bootstrap. Rather, for a meaningful comparison to the bootstrapped $J_{0.95}$, the J-values of simulations of equal length should be used.

It is obvious that the information from a single simulation run over the relatively short bootstrap period $T = 6750$ is not reliable enough for a final evaluation of the model fit. Nevertheless, it is easy to undertake a Monte Carlo (MC) experiment and repeat these simulations many times. In this way an entire distribution of model-generated J-values is obtained, which can subsequently be contrasted with the bootstrap distribution.

To describe our precise method, and to distinguish it clearly from the experiment below, let us emphasize that the bootstrapped moments are extracted from the empirical data $\{r_t^{emp}\}_{t=1}^{T}$ by writing $m^b = m^b(\{r_t^{emp}\})$, and that the weighting matrix is deduced from the same set of data by writing $W = W^{emp}$.[16] Furthermore, indicate the different random number sequences underlying the MC simulations by the letter c, giving rise to moments $m = m^c(\theta^o; T)$ (as in the bootstrap, we work with $c = 1, \ldots, 5000$). The frequency distributions from the following two sets of J-values are then considered:

$$
\begin{aligned}
&\text{Bootstrap:} \quad \{\, J[\, m^b(\{r_t^{emp}\}),\ m_T^{emp};\ W^{emp}\,]\, \}_{b=1}^{5000}\ , \qquad T = 6750 \\
&\text{Monte Carlo:} \quad \{\, J[\, m^c(\theta^o; T),\quad m_T^{emp};\ W^{emp}\,]\, \}_{c=1}^{5000}
\end{aligned}
\tag{13}
$$

The outcome of this experiment is shown in Figure 1.[17] The bootstrap distribution, plotted as the thin solid (green) line, is much narrower than the MC distribution, which is the bold (blue) line. A large part of the simulated series would therefore be rejected. How many exactly would be rejected is marked by the vertical line at $J = 22.5$, which is the 95% quantile of the bootstrap distribution. Regarding the MC distribution, we compute that 27.7% of its J-values are less than this rejection level. For a succinct summary of the general moment matching, it may therefore be tempting to say that our model has a p-value of 27.7%, with respect to the S&P 500 index and the parameter set θ^o. This means that roughly three-fourths of the simulations of the model over $T = 6750$

---

[16] The moments m^b are the same as those that already underlaid the estimate of the variance-covariance matrix $\widehat{\Sigma}$ in (10), which in turn yielded W^{emp} as its inverse by eq. (9).

[17] The density functions in this and the other diagrams are estimated by means of the Epanechnikov kernel; see Davidson and MacKinnon (2004, pp. 678–683) for the computational details.

periods would be classified as being inconsistent with the empirical returns, while roughly one-fourth of these simulations would not be rejectable on these grounds.
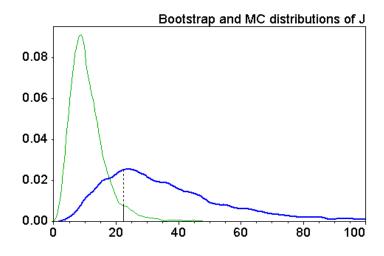

Figure 1: Distributions of J from the bootstrap (solid line) and the MC simulations of θ^o (bold line) described in (13).

Note: The vertical line at $J = 22.5$ indicates the 95% quantile of the bootstrap distribution, and the 27.7% quantile of the MC distribution.

The conclusion, so far, is that the model is not rejected by the data. The assumption on which it rests was clarified at the beginning of the section, namely, that the bootstrap distribution of J is a proxy for the distribution of J that we (hypothetically) would procure from different samples of the real-world DGP. The extent to which this assumption is justified, and how the conclusions drawn from it may have to be modified, can be checked by an elementary thought experiment where we have everything under control. It is based on the question of how the bootstrap and MC distribution of J would be related if both come from the *same* DGP, and this 'true' DGP is given by our model with parameter set θ^o.

Thus, let us take the simulated returns obtained in the estimation procedure with the near-optimal parameters θ^o and use the first $T = 6750$ values as an artificial 'empirical' sample. [18] To highlight the fact that it has been generated from the special random number sequence underlying the estimation, we indicate this sequence by $c = 0$ and write $\{r_t^o(\theta^o)\}_{t=1}^T$ for the resulting return series (which could also be viewed as the outcome of a special MC simulation run with $c = 0$). The corresponding 'empirical' moments are $m^o(\theta^o; T)$, for which we may more briefly write m_T^o.

---

[18] We repeated the following experiment several times with other 'empirical' samples from the model and found no great differences in the results. The numerical figures reported below are therefore sufficiently representative.

In the first step, we follow the bootstrap prescriptions of eqs (9), (10) (again with the differentiated block lengths of 250 and 750 days, as described in the previous section) and construct a weighting matrix from the returns $\{r_t^o(\theta^o)\}$. To distinguish it from the matrix underlying (13), we designate it $W = W^o$, and write $m^b = m^b(\{r_t^o(\theta^o)\})$ for the bootstrapped moments from which W^o was composed. Subsequently, the J-values based on W^o are computed from these 5000 moments. On the other hand, the simulation runs in the MC part of the experiment can be the same as above. The only difference is that their moments are now related to m_T^o (instead of m_T^{emp}), and that these deviations are evaluated using the weighting matrix W^o. To sum up, this experiment is implemented by the following two distributions:

$$\text{Bootstrap:} \quad \{\, J[\, m^b(\{r_t^o(\theta^o)\}),\ m_T^o;\ W^o\,]\, \}_{b=1}^{5000}$$

$$\text{Monte Carlo:} \quad \{\, J[\, m^c(\theta^o; T), \quad m_T^o;\ W^o\,]\, \}_{c=1}^{5000} \tag{14}$$

Figure 2 presents the distributions resulting from (14). The most important thing to note is that the bootstrap and MC distribution do *not* coincide, not even approximately so. Although it is so consequential, the phenomenon that the bootstrap distribution (again the thin solid line) is narrower than the MC distribution (the bold line) may not be too surprising. After all, the bootstrap distribution is constructed from a single realization of the DGP, so that the variability in J is exclusively due to the stochastic mixing and selection of the bootstrap blocks in the return series. By contrast, the MC distribution has more heterogeneity in its data since the random shocks that are at work in each of its samples feed back on the price and therefore, in interaction with the deterministic core, propagate in the entire model. [19]

As a matter of fact, in the present setting it is the MC distribution that would be the appropriate distribution to set up a critical value of J, simply because it is the distribution of the objective function from a large number of realizations of the true DGP. Apparently, the bootstrap distribution, with which we would try to help ourselves if we had only one sample of the same DGP available, is not a very representative proxy for it.

In other words, in the beneficial case where we have the data of 5000 realizations of the DGP on hand, it would not be correct to take the 95% quantile of the bootstrap distribution in Figure 2, where $J = 23.2$, as the critical value of J; it is rather $J = 66.8$, which constitutes the 95% quantile of the MC distribution. [20] By the same token, it

---

[19] In particular, near the isoclines of the majority index x_t, even small shocks of opposite sign can drive the market toward either chartism or fundamentalism with their different volatility regimes, where it then will stay for quite a while; see Figures 1 and 2 in Franke and Westerhoff (2009).

[20] The quantile is so high because the distribution stretches out further to the right. Although appropriate, it is not indicated in Figure 2 since a vertical line at $J = 66.8$ would hardly be visible.
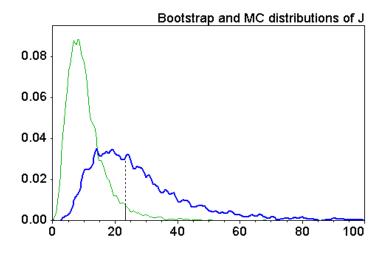
Figure 2: Distributions of J from the bootstrap (solid line) and
the MC simulations of θ^o (bold line) described in (14).

Note: The vertical line at $J = 23.2$ indicates the 95% quantile of the bootstrap distribution,
and the 47.3% quantile of the MC distribution.

would not be correct to refer to the observation that $J = 23.2$ is the 47.3% quantile of
the MC distribution and infer from it that the model has a p-value of 47.3%. Since here
we know that the model is the true DGP, its p-value is 95% by construction.

Two things can thus be learned from the experiment of eq. (14). Generally, the boot-
strap distribution of J, since it is based on a single empirical data series, is very likely
to be narrower than the distribution we would obtain if we had at our disposal a large
number of different realizations of the real-world DGP.[21] The true critical value of J
would therefore be higher than $J_{0.95} = 22.5$, the value obtained in Figure 1. Likewise, the
true p-value of the MC distribution in Figure 1 would be higher than the 27.7% induced
by this value of J. That is, what we have tentatively called the "p-value" of our model
can only be a proxy p-value, to which we resorted for lack of something more precise.
This also means that $J_{0.95} = 22.5$ is a most conservative guess of the true critical value
of J, so that our proxy p-value is (perhaps considerably) downward biased. Hence, our
model is rejected even less by the data than stated before.

A second insight from the experiment is more specific, though somewhat speculative.
In the thought experiment with the model as the true DGP, the bootstrap distribution
proposed 23.2 as a candidate for a critical value of J, whereas from the 95% quantile
of the MC distribution we know that the true critical value is 66.8 . If we now dare
to hypothesize that the 95% quantile of the empirical bootstrap, $J_{0.95} = 22.5$, and the

---

[21] This supposition is indirectly confirmed here by the fact that the narrow model-based bootstrap
distribution in Figure 2 is very close to the empirical one in Figure 1. In particular, with 22.5
versus 23.2, their 95% quantiles are only marginally different.

unknown 95% quantile $J_{0.95}^{rw}$ of the MC distribution of samples from the real-world DGP are similarly related, then the equation $J_{0.95}^{rw}/22.5 = 66.8/23.2$ can be set up to obtain $J_{0.95}^{rw} = 64.8$. Turning to the MC distribution in Figure 1, where the model-generated samples are evaluated at the empirical weighting matrix W^{emp}, we find that it has its 89.0% quantile at this value of J. So our guess of a more appropriate p-value of the model would be that it is even in as high a region as 89%.

While we do not mean to take this p-value literally, it may underline the good fit of the model. Furthermore, an argument of this kind could be interesting if, in the estimation of another model, a proxy p-value less than the original 27.7% in Figure 1 were obtained. On the other hand, in order not to exaggerate the 89% p-value, we recall that the objective function is composed from a limited choice of moments. The favourable p-values, to which of them so ever we might refer, will certainly decrease if additional moments (i.e., additional dimensions of reality) were included to test the model.

5. Matching the single moments

Despite the positive evaluation of the model as a whole in the previous section, the matching of the empirical moments is, of course, not perfect. For a more detailed analysis of the model's goodness-of-fit, it is enlightening to compare the (marginal) MC distributions of the single simulated moments $m^c(\theta^o; T)$ computed above to the moments $m^b(\{r_t^{emp}\})$ from the empirical bootstrap. From the deviations or a close coincidence of the two distributions, we can hopefully learn along which dimensions the model performs well or where it may still exhibit certain shortcomings.

Figure 3 presents the distributions of six selected moments. There is only one moment, namely the mean value of the absolute returns (vMean), for which the MC distribution nearly impeccably reproduces the bootstrap distribution. A minor shift to the right is obtained for the MC distribution of the autocorrelation of the absolute returns with a lag of 25 days (vACF–25), while the coefficients with a lag of 5 and 100 days (vAC–5 and vAC–100, respectively) show a minor shift to the left.

A stronger downward bias has to be noticed for the first-order autocorrelation of the absolute returns (vAC–1). The deficiency of the model responsible for this and the slightly weaker autocorrelation effects at lag 5 can be understood from a direct comparison of the profile of the empirical ACF with the ACF from a single (but long) simulation run. [22] Corresponding to the often discussed power law behaviour, the human eye perceives a slight convex curvature over the first (more or less) 30 lags of the empirical ACF. In contrast, the decay in the simulated ACF occurs in an almost linear manner. Hence, if we think of the estimation problem as basically fitting a straight line to the curved line of

---

[22] Not shown for reasons of space, but see Figure 3 in Franke and Westerhoff (2009).
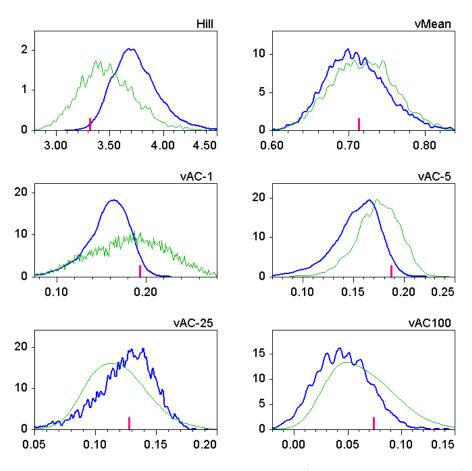
Figure 3: Distributions of selected moments (based on empirical data).

Note: The bold (blue) line represents the distributions of the moments resulting from the MC simulations of θ^o, the solid (green) line depicts the corresponding bootstrap distributions from $\{r_t^{emp}\}$; both with a time horizon of 6750 days. The (red) vertical bar indicates the empirical moments.

the empirical ACF, there will be a tendency to underestimate the lower-lag coefficients and to overestimate the higher-lag coefficients.

In addition to the quasi-linear decay of the ACF, a second potential problem of the model is revealed in the top-left panel of Figure 3. The upward bias in the MC distribution of the Hill estimator indicates that the model with its parameter set θ^o tends to produce a lower fatness in the tails of the returns. Nevertheless, the six panels in Figure 3 also make it clear that the deviations in the distributions of the simulated from the bootstrapped moments are associated with a relatively high sample variability. This is something that has to be accepted if the time series are "only" 27 years long (when $T = 6750$). It is, however, somewhat astonishing that both types of distributions here have a similar spread, in contrast to the distributions in Figure 1 of the objective function itself, which in a certain sense "adds up" the single moments.

17

The diagrams in Figure 3 show the effects in relation to the width of the entire distributions and therefore might over- or underemphasize the deviations somewhat. Besides, three of the nine moments to be estimated are omitted. To supplement Figure 3 and to give a quantitative short summary information about the order of magnitude in the imperfect moment matching, Table 2 contrasts the median values of the moment distributions from the MC simulations with those from the bootstrap distributions. Looking at the figures of the autocorrelation coefficients of the absolute returns, their underestimation at the low lags 1 and 5 may now appear less pronounced than in the graphical representations of the frequency distributions. The deviations of the simulated moments at the medium and long lags should be even more tolerable. hence the Hill estimator seems to remain the moment with the strongest bias.

| | rAC–1 | Hill | vMean | vAC | | | | | |
				1	5	10	25	50	100
median MC :	0.006	3.72	0.707	0.160	0.156	0.148	0.127	0.094	0.045
median B :	−0.009	3.46	0.718	0.183	0.175	0.148	0.117	0.094	0.059
t-statistic :	1.39	1.01	−0.26	−0.53	−0.89	0.06	0.43	−0.08	−0.49

Table 2: Moment statistics.

Note: B and MC refer to the samples of length $T = 6750$ from the bootstrap and the MC simulations with θ^o. The t-statistics is derived from eq. (15).

A comparison of the median values should not only be based on the mere figures but may also take the sampling variability of the single moments into account. As the latter is captured by the diagonal elements in the estimated variance-covariance matrix of the moments from eq. (10), it is straightforward to compute the (analogue of the) t-statistics of the differences in the medians. [23] They are given, in obvious notation, by

$$t_i = (\text{MC-median}_i - \text{B-median}_i) \Big/ \sqrt{\widehat{\Sigma}_{ii}} \qquad (15)$$

Their values are reported in the last row of Table 2. Since we have no knowledge of their distributional properties, (15) cannot be used as a test statistic to infer whether or not the MC and bootstrap distributions are "significantly" different. However, these figures can give us information about the moments' goodness-of-fit relative to each other. An exception is the autocorrelation of the raw returns, where the values of 0.006 and −0.009

---

[23] As in ordinary hypothesis testing in finite samples, using the expression "t-statistic" does not mean that it will follow the Student's t-distribution.

are practically zero for us; therefore this t-statistic, although it is the highest, can be safely neglected. The matching of the Hill estimator and the autocorrelation coefficient of $v_t = |r_t|$ at lag 5 are, then, the least satisfactory features of the model. Apart from this, the deviations in the ACF of v_t at very low and (in our context) very long lags can perhaps be evaluated as mild, while there are no problems worth mentioning at the medium lags or with respect to the scaling of the returns in form of the mean value of v_t.

When we mentioned "deviations" in the last paragraph, it may have been understood as implying that these phenomena would not occur if the Monte Carlo moments in Table 2 had been generated from the true DGP. This presumption can be readily checked with the model-generated 'empirical' return series that was already used in the experiment in the previous section. That is, as explained when introducing eq. (14), we take a specific random number sequence, simulate a return series designated $\{r_t^o(\theta^o)\}_{t=1}^T$ on this basis, and bootstrap 5000 moment vectors $m^b = m^b(\{r_t^o(\theta^o)\})$ from it. Rather than applying the objective function J to this sample, we now consider the frequency distributions of the single moments and contrast them with the moment distributions from the model's Monte Carlo simulations $m^c = m^c(\theta^o; T)$. The outcome for the six selected moments considered in Figure 3 is presented in Figure 4. The distributions of the MC moments are the same as in Figure 3, and again drawn as bold (blue) lines. The solid (green) lines represent the alternative bootstrap distributions from the artificial data series $\{r_t^o(\theta^o)\}$.

Possibly, the first impression gained from the six panels in Figure 4 is the perfect congruence of the two distributions for the Hill estimator; the deviations displayed in Figure 3 have completely disappeared. However, the top-right panel reveals that this improvement is not without a side effect: whereas in Figure 3 it was the bootstrap and MC distributions of the mean volatility that almost coincided, this feature is not preserved and vMean in Figure 4 shows a similar deviation to that of the Hill estimator in Figure 3.

We should add here that the bootstrap distribution from an alternative artificial series $\{r_t^o(\theta^o)\}$ can easily give rise to a further shift of, in particular, the two moments just discussed. The reason for this is that (put somewhat loosely) the model dynamics irregularly switches between a fundamentalist and a chartist regime, with a low and high volatility, respectively. It is in this respect that, as stated above, the time horizon $T = 6750$ is still rather short, so that two arbitrary samples may have different numbers of such regimes, with different durations (cf. also footnote 19). Accordingly, the series may have quite distinct Hill estimators and mean volatilities, and this will carry over to the (support of the) moment distributions bootstrapped from them. Hence, if the dissimilarities in a bootstrap and MC distribution are not too dramatic, it is not obvious whether they are mainly due to different DGPs underlying them or to the peculiarities of the return series itself from which the bootstrap samples derive.

Apart from a shift of the MC *versus* the bootstrap distributions of the moments, the
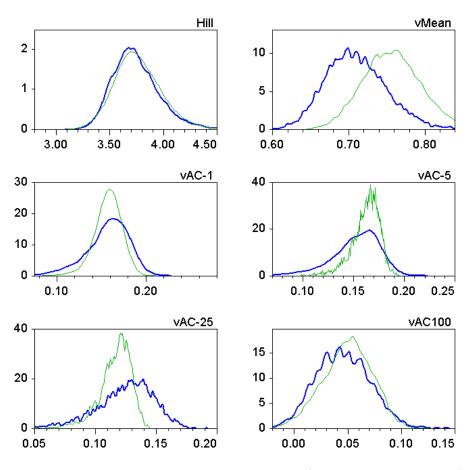
Figure 4: Distributions of selected moments (based on artificial data).

Note: The bold (blue) line represents the distributions of the moments resulting from the MC simulations of θ^o, and the solid (green) line depicts the corresponding bootstrap distributions from $\{r_t^o(\theta^o)\}$; both with a time horizon of 6750 days.

results for J in the previous section seem to suggest narrower bootstrap distributions here, too (since they use information from a single history of returns, whereas the MC distributions evaluate a large sample of different series). This idea is in fact borne out by the autocorrelations of the absolute returns at lags 1, 5 and 25 in Figure 4. However, it may also be recalled that the single empirical return series gave rise to wider bootstrap distributions in these cases (see Figure 3), so there is no general rule in this respect either. [24] We take these observations as further evidence against premature conclusions from, or overinterpretations of, possible differences between the bootstrap and MC distributions of the moments.

In sum, the (at the worst) moderate dissimilarities in the bootstrap and MC moment distributions documented in Figure 3 and Table 2 might (at the worst) be interpreted as

---

[24] The bootstrap distribution of vAC–1 in Figure 3 is not only wider than in Figure 4, but even wider than the MC distribution.

being weakly indicative of the fact that the model is not sufficiently compatible with the real-world DGP. However, even if we therefore wished to improve the model, we could not be sure along which dimensions this should primarily be done.

6. Parameter identification

When introducing the benchmark parameters θ^o, three questions were hinted at (implicitly at least). (A) Can we be sure that a local minimum of the objective function (11) found by the search procedure is also a global minimum? (B) How large is the "valley" where a minimum has been found? And (C), how different will the solution be if another random number sequence is employed in the model simulations. All three questions are aspects of the problem of how well the model parameters can be identified, a special sub-problem being whether all of the parameters truly contribute to the market dynamics and are hence significantly distinct from zero.

Let us pick up on question (B), which involves the local nature of the objective function around the minimum. A long valley extending in the direction of a coefficient or a combination of coefficients would be indicative of an identification problem; one parameter set would be equally good at matching the moments as another very different one. [25] A most straightforward method for a first exploration of the local neighbourhood around the minimum solution is to fix all parameters except one at their estimated values, and to compute J over the long simulation horizon $S = 10 \cdot 6866$ as a function of the variable parameter θ_i (maintaining, of course, the random number sequence, designated $c = 0$, underlying the estimation with its moments $m = m^c(\theta; S)$). The results are very similar if, instead of the parameter estimates, reference is again made to the benchmark vector θ^o. To be precise, if a vector $\theta_{\mathrm{ex}(i)}$ collects all components of the vector θ except the i-th one, then locally around θ_i^o we consider the functions defined by

$$J^o(\theta_i) \; := \; J[\, m^o(\theta_i, \theta_{\mathrm{ex}(i)}^o; S), m_T^{emp}; W^{emp}\,] \, , \qquad\qquad i = 1, \ldots, 6 \qquad (16)$$

An important feature of the model is that the market irregularly switches between two regimes—one dominated by chartists and the other by fundamentalist traders. Although with respect to a given random number sequence the objective function is purely deterministic, one may therefore expect the system to be fairly sensitive to changes in the initial conditions or, which amounts to the same, to slight variations in the parameters. This is confirmed by the graphs of the functions $J^o = J^o(\theta_i)$, which are all plotted in Figure 5. In fact, the raggedness they display is immediately visible. [26] The accordingly

---

[25] An example of this problem is the MSM estimation of the Manzan–Westerhoff (2005) asset pricing model in Franke (2009). There it was impossible to pinpoint the reaction parameter ϕ of the fundamentalists (which was only natural in that modelling framework and with the moments considered). It could not even be recognized as being bounded away from zero.

[26] Other examples of this behaviour can be found in Franke (2009) or Gilli and Winker (2003). Incidentally, the graphs for our two predetermined parameters $\theta_i = \alpha_o, \alpha_x$ are noticeably smoother.

high number of local minima is typical for applications of the method of simulated moments to agent-based asset pricing models, and often the functions do not become much smoother if they are computed over smaller intervals with a higher resolution. It would therefore not be very pertinent to put too much strain on a search algorithm to determine the "absolute" minimum. [27]
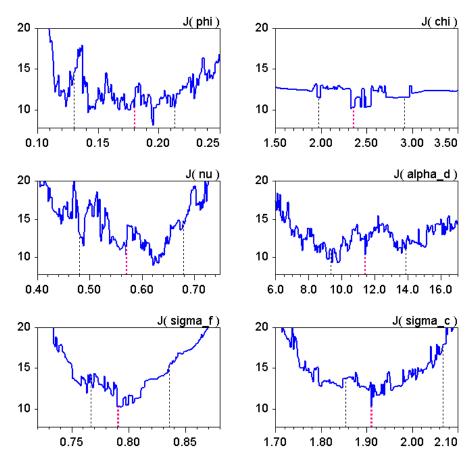

Figure 5: Graphs of the functions $\theta_i \mapsto J^o(\theta_i)$, for $\theta_i = \phi, \chi, \nu, \alpha_d, \sigma_f, \sigma_c$.

Note: Middle (red) dashed line depicts the benchmark value θ_i^o of the parameter, the two outer dashed lines indicate the 2.5% and 97.5% quantiles of the estimations presented below. The underlying time horizon is $S = 68660$.

Generally, the diagrams in Figure 5 make us more confident that a solution of the estimation problem is situated within the parameter intervals shown there (we also checked this over wider intervals, especially for $J^o = J^o(\chi)$ in the top-right panel). In detail, the graphs give us a clue where there are prospects for improvements. Regarding the benchmark θ^o, this is clearly seen for the coefficients ϕ, ν, α_d, and plots of these functions

---

[27] Unless one is interested in the properties of the algorithm itself, as in Gilli and Winker (2003), where the authors test their deterministic analogue of a simulated annealing procedure.

were indeed the basis for the interactive search mentioned at the end of Section 3. Recall that we reported the existence of two local minima superior to θ^o, where one of them is essentially characterized by $\phi = 0.195$ and the other by $\nu = 0.634$. However, are the corresponding gains in J (which was lowered in this way from 10.33 to 8.00 and 7.79) significant? Or, to return to the formulation above, what exactly would be "too much" strain in the determination of a global minimum?

These considerations lead us to the problem of how to construct a confidence interval for the single parameters. Unfortunately, we have no critical level J^{crit} of the objective function J, analogous, for example, to that from a likelihood ratio test, which would specify the confines of a parameter region or interval beyond which $J > J^{crit}$ and where parameters entailing $J > J^{crit}$ could be regarded as being "significantly different" from the vector that had minimized J. Instead of attempting to construct such a value J^{crit} from simulated data (which could only be valid for a given random number sequence), we can use a Monte Carlo re-estimation as a more direct approach. Accordingly, we devise a large number of different shock sequences, re-estimate the model on each of them, and compute the frequency distributions of the parameters thus obtained. This experiment will also answer question (C) from above. To address question (A) for the global minimum, we need not be overly eager to find it with rigorous accuracy in these estimations, since there is no indication that the errors are very systematic when a parameter vector is prematurely accepted as a minimum solution (provided, of course, the search algorithm has already proceeded far enough; with two re-initializations, say).

Specifically, a sample size of 1000 of these re-estimations proves to be sufficient for our purpose. The resulting marginal distributions of the six parameters ϕ, χ, ν, α_d, σ_f, σ_c are displayed in Figure 6, together with the median values and the 2.5% and 97.5% quantiles (the dashed lines for the latter two are not always clearly visible if the tails are too flat). Despite perhaps having preferred a smoother shape in the distribution of χ, which would probably require a (much) larger sample, the distributions are sufficiently informative. All are unimodal and at least so regular that the three quantiles are suitable statistics for a short characterization.

Table 3 gives a synopsis of the numerical values of the three quantiles (rounded). Contrasting the medians with our benchmark vector θ^o in the first two rows, it is seen that θ^o was a decent choice of a representative parameter set.[28] The intervals between the two outer quantiles can now be regarded as the confidence intervals that we have been looking for. This is also why these values were additionally plotted as the outer dashed

---

[28] We decided on θ^o as our benchmark parameters at an early stage of our investigations and subsequently came across no reason why it should be replaced by another, in some sense better vector. More systematically, we could also have employed the median values from Table 3 as a benchmark. In fact, this would hardly make any difference. In particular, the proxy p-value of the model established in Section 4 would then only slightly decrease (rather than increase) from 27.7% to 26.8%.
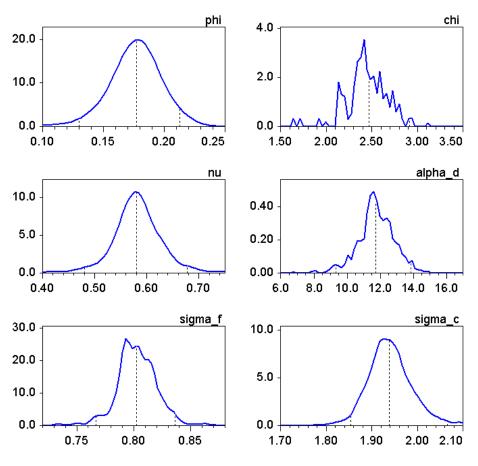
Figure 6: Marginal distributions of the re-estimated parameters.

Note: The vertical dashed lines indicate the 2.5%, 50% and 97.5% quantiles, respectively. The underlying sample size is 1000.

lines in Figure 5. Returning to these diagrams, it can now be seen that the increase from a minimum J to which the two quantiles give rise are not necessarily symmetric (which would make it even more problematic to establish one critical level of the function, J^{crit}, for all of the parameters).

The last row in Table 3 computes the ratio of the standard deviation of the distributions and the medians. It shows that regarding their estimation, the parameters fall into two categories with a, respectively, similar relative width of their distributions. Remarkably, the coefficients in the deterministic skeleton of the model all belong to the same class, that with the wider distributions, while the two noise parameters σ_f and σ_c in the stochastic part both have much narrower distributions. In any case, all of the six parameters are well identified and definitely bounded away from zero.

Although the marginal distributions of the parameters provide valuable information about the identification of the model, it may also be recognized that generally the parameter estimates are not completely independent of each other. Limiting our attention

	ϕ	χ	ν	α_d	σ_f	σ_c
benchmark:	0.18	2.35	0.57	11.40	0.79	1.91
median:	0.18	2.47	0.58	11.74	0.80	1.94
2.5% quantile:	0.13	1.97	0.48	9.32	0.77	1.85
97.5% quantile:	0.21	2.91	0.68	13.86	0.83	2.07
sd/median:	0.114	0.092	0.083	0.092	0.023	0.027

Table 3: Summary statistics of the estimated parameters.

Note: 'sd' means standard deviation of the corresponding marginal distribution.

to possible interdependencies between pairs of coefficients, a convenient measure is given by the correlation coefficients of the 1000 estimates. The compilation of these statistics in Table 4 shows a wide range of what is possible. There are relatively strong correlations, which may be positive or (mostly) negative, while others are weak or practically non-existent. To pick out three cases: the reaction coefficients of the fundamentalists and chartists, ϕ and χ, exhibit a certain complementarity: a higher estimate of one of them tends to go hand in hand with a lower estimate of the other; likewise, higher estimates of the flexibility parameter ν tend to be associated with lower estimates of the noise level σ_f in fundamentalist demand, but higher noise levels σ_c in chartist demand. On the other hand, estimations of the noise levels of the two groups of traders themselves are perfectly independent (which is in some contrast to the 'complementarity' between their deterministic responsiveness ϕ and χ).

	ϕ	χ	ν	α_d	σ_f	σ_c
ϕ:	1.00	−0.45	−0.14	−0.09	0.07	−0.34
χ:		1.00	−0.26	−0.41	−0.01	−0.07
ν:			1.00	−0.11	−0.59	0.36
α_d:				1.00	0.14	0.03
σ_f:					1.00	−0.01
σ_c:						1.00

Table 4: Pairwise correlations of parameter estimates.

Figure 7 translates this information into a graphical representation by constructing two-dimensional confidence regions for selected pairs of parameters. To this end, we make
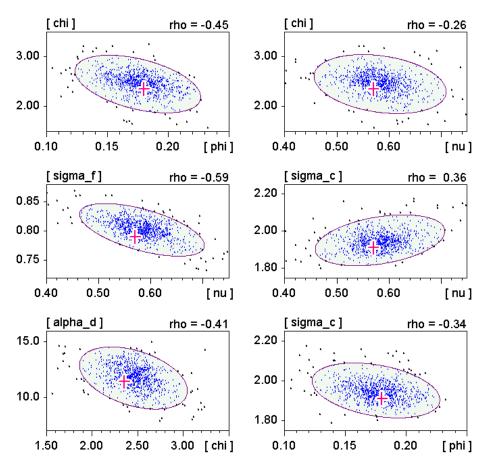
Figure 7: Two-dimensional confidence regions of the parameter estimates.

Note: The (red) crosses indicate the benchmark values of the parameters, and the dots are the re-estimated pairs $(\hat{\theta}_i^c, \hat{\theta}_j^c)$ (some of which may be outside the panels); 'rho' is the correlation coefficient.

use of the Wald statistic (see, e.g., Davidson and MacKinnon, 2004, pp. 189–196). With respect to two (distinct) components i and j of the parameter vectors, let $\{(\hat{\theta}_i^c, \hat{\theta}_j^c)\}$ be the set of their estimates for the random number sequences designated $c = 1, \ldots, 1000$, and V the corresponding 2×2 variance-covariance matrix. Furthermore, let $\bar{\theta}_k$ be the mean of the 2.5% and 97.5% quantiles of the $\hat{\theta}_k^c$ ($k = i, j$). Writing $\hat{\eta}^c := (\hat{\theta}_i^c, \hat{\theta}_j^c)'$ and $\bar{\eta} := (\bar{\theta}_j, \bar{\theta}_j)'$ for compactness, we then obtain the 1000 "test statistics"

$$\tau^c = \tau(\hat{\eta}^c) := (\hat{\eta}^c - \bar{\eta})' \, V \, (\hat{\eta}^c - \bar{\eta}), \qquad c = 1, \ldots, 1000 \qquad (17)$$

Sorting τ^c from smallest to largest and denoting their 95% quantile by $\tau^\star$, the 95% confidence region is specified as the set

$$C_{ij} := \{ (\theta_i, \theta_j) : [(\theta_i, \theta_j)' - \bar{\eta}]' \, V \, [(\theta_i, \theta_j)' - \bar{\eta}] \le \tau^\star \}, \qquad i \ne j \qquad (18)$$

Hence, by construction, the confidence regions in Figure 7 have an elliptical shape and are symmetrical around $(\bar{\theta}_i, \bar{\theta}_j)$, not the benchmark (θ_i^o, θ_j^o). We already know that the

estimates are not uniformly distributed, which implies that the C_{ij} are not filled out evenly by the pairs $(\hat{\theta}_i^c, \hat{\theta}_j^c)$. Characterizing the density of the clouds of dots in the two-dimensional parameter plane, we note that roughly $80\text{--}85$ per cent of the estimates are concentrated in a considerably smaller area than the entire C_{ij} region. Hence, when applied to other financial markets, the model seems to offer good prospects of being able to classify them as being more or less similar to the S&P 500, or essentially distinct from it.

7. Estimation of a foreign exchange market

Having established the positive parameter identification in the estimations of the S&P 500, we are now interested in the problem of whether the model can also differentiate between (at least) stock and foreign exchange markets. In this respect, we hope it will prove at least as efficient as the totally different (and much simpler) asymmetric herding model estimated in Alfarano and Franke (2007). The maximum likelihood method that could be applied there characterized the stock markets by a consistently higher population share of noise traders vis-à-vis the stabilizing fundamentalists. To check whether statements of a similar kind hold true for the present model, we concentrate on a representative foreign exchange market, namely the US dollar / Deutsche mark exchange rate over the same period as above (6861 daily observations from January 1980 to mid-March 2007).[29] Of course, in addition to telling the USD–DEM and S&P 500 indices apart, we also wish to investigate whether the previous good estimation properties are basically maintained.

After a few explorations to guide the search algorithm towards a suitable parameter region, it became clear that chartism will have to play a weaker role in the market dynamics. This finding incited us to go to the extreme and eliminate the chartists altogether. As already remarked on the demand equations (1) and (2), by setting the coefficients specifying the chartist behaviour to zero, $\chi = \sigma_c = 0$, the agents choose between fundamentalist strategies and not trading at all. More in line with empirical evidence, the two components of the hitherto 'fundamentalist' demand (1) could also be interpreted as modelling pure fundamentalists (the deterministic part) and—in the aggregate—pure noise traders (the stochastic part).[30]

Usually the objective function was not yet strictly minimized under this zero constraint. Since, on the other hand, its relaxation led to only relatively minor improvements, we prefer the greater clarity of the radical assumption. The benchmark parameters on

---

[29] Following euro adoption, the USD–DEM rates are pseudo rates imputed by applying the euro locking rate to the current euro exchange rate.

[30] It would be interesting to find out if the performance of the model could be enhanced by considering a choice between inactivity, fundamentalists, and chartists proper (similar to, in this respect, Westerhoff, 2008).

ϕ	χ	μ	ν	α_o	α_m	α_d	σ_f	σ_c
1.75	0.00	0.01	0.92	-0.15	1.35	37.10	1.55	0.00

Table 5: Benchmark parameter set θ^o for USD–DEM exchange rate.

which we thus decide are reported in Table 5. Comparing them to the set in Table 1, it may be noted that besides the disappearance of the chartists, the reaction coefficient ϕ of the fundamentalists is now almost ten times higher and the parameter α_d more than three times as high. Since both coefficients reinforce the mean-reverting forces in the market, the price deviations from the fundamental value as well as the volatility of the returns will be more limited. [31] It furthermore makes good sense that *a priori*, as expressed by the higher value of the flexibility parameter ν, the agents have a stronger tendency to switch to inactivity ($\nu = 0.92$) than to chartism ($\nu = 0.57$ in Table 1 for the stock market).

To evaluate the model's overall goodness-of-fit with respect to the benchmark parameters, we proceed as in Section 4. That is, as described in eq. (13), we bootstrap the empirical returns and contrast the resulting values of the function J with the MC distribution of the objective function (based on θ^o, of course). However, here two versions of specifying J will be considered: the same function as above, which continues to underlie the estimations to follow, and, second, a version $\widetilde{J}$ where the first-order autocorrelation of the raw returns (rAC–1) is excluded from the moments entering the function.

The reason for the second version is that the empirical coefficient is 0.013 with a confidence interval from -0.011 to 0.037, so that we still tend to view this correlation as negligible. [32] The MC distribution, on the other hand, is centred around -0.003 and therefore bears out the model's complete absence of autocorrelation in r_t. While these differences do not seem serious, they nevertheless cause a noticeable increase in the J-values of the MC *versus* bootstrap distribution, part of which effect is due to the corresponding diagonal element in the weighting matrix, which is almost twice as high as in the weighting matrix derived from the S&P 500.

Using, therefore, the function $\widetilde{J}$ as our main reference, the bootstrap and MC distributions are shown in Figure 8. The picture is qualitatively similar to Figure 1 in Section 4, although now the MC distribution turns out to be wider. As a consequence, we get a lower proxy p-value of the model, as we called it. The value $\widetilde{J} = 25.8$, which constitutes the 95% quantile of the bootstrap, corresponds to the 18.4% quantile of the MC distribution. If instead we take the original objective function J with rAC–1 included, the additional penalty reduces the proxy p-value to 12.8%. This is worse than in the

---

[31] Although the weaker price misalignment is not shown in this paper, we checked it.

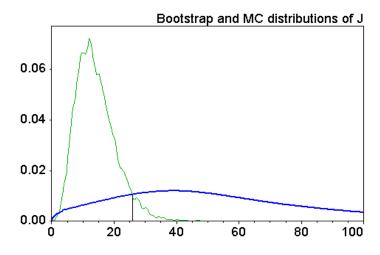[32] The Bartlett confidence band is $0.013 \pm 2/\sqrt{6750}$.

Figure 8: Distributions of $\widetilde{J}$ for USD–DEM from the bootstrap
(solid line) and the MC simulations of θ^o (bold line).

Note: The vertical line at $\widetilde{J} = 25.8$ indicates the 95% quantile of the bootstrap distribution, and the 18.4% quantile of the MC distribution ($\widetilde{J}$ does not include the rAC–1 moment).

estimation of the S&P 500 (where this proxy p-value amounted to 27.7%). But still, and all the more if we refer to the discussion in the second part of Section 4, the model is not rejected by the data either.

For a more detailed diagnostics to identify the moments with the greatest problems, we are less exhaustive here than in Section 5, and content ourselves with the counterpart of Table 2. The statistics reported in Table 6 buttress the negative, even dominant role of the rAC–1 moment in the objective function, since all other moments have a substantially lower (in modulus) t-statistic.

The second-worst match is the first-order autocorrelation (smoothed) of the absolute returns v_t. Certain problems with it are, however, preconditioned. Whereas we were happy enough for the model to produce a systematic decay in the ACF(v_t), something widely considered to be a stylized fact of financial markets, the USD–DEM exchange rate displays the curiosity of an increase over the first few lags. The relatively high (in modulus) t-statistic of the coefficient at lag 100 has a similar explanation: empirically, the ACF tends to increase slightly rather than decrease from lag 90 on. Since the model, as any other model we know of, is not designed to mimic these counterexamples to a generally monotonic behaviour in the ACF, the overall fitting of the model may appear in a more positive light than before.

Regarding the identification of the parameters, we directly turn to their re-estimation when using 1000 different random number sequences in the simulation runs. The main characteristics of the resulting marginal distributions of the four parameters are given in Table 7. Again, all of the coefficients have well-confined confidence intervals. The different

	rAC–1	Hill	vMean	vAC					
				1	5	10	25	50	100
empirical:	0.013	3.98	0.496	0.084	0.101	0.091	0.060	0.034	0.022
median B:	0.013	4.04	0.500	0.080	0.098	0.087	0.053	0.029	0.016
median MC:	−0.003	4.23	0.491	0.100	0.092	0.082	0.057	0.031	0.008
t-statistic:	2.42	0.76	−0.50	1.21	−0.43	−0.35	0.40	0.22	−0.83

Table 6: Moment statistics for USD–DEM exchange rate.

Note: B and MC refer to the samples of length $T = 6750$ from the bootstrap and the MC simulations with θ^o. The *t*-statistics derive from eq. (15).

	ϕ	χ	ν	α_d	σ_f	σ_c
benchmark:	1.75	0.00	0.92	37.1	1.55	0.00
median:	2.08	−−	0.91	37.8	1.60	−−
2.5% quantile:	1.41	−−	0.84	25.5	1.49	−−
97.5% quantile:	2.98	−−	0.94	44.9	1.86	−−
sd/median:	0.196	−−	0.026	0.126	0.060	−−

Table 7: Summary statistics of the parameter estimates for USD–DEM.

Note: 'sd' means standard deviation of the corresponding marginal distribution.

nature of the USD–DEM and S&P 500 markets is also further corroborated, as none of them shows the slightest overlap with a confidence interval of the re-estimates in Table 3 for the stock market. The relative widths of the intervals, specified as the ratio of the standard deviation of the distribution to the median, are perhaps interesting to note but cannot be too strictly compared to the ratios in Table 3 since for USD–DEM there are fewer parameters to estimate.

Table 8 shows that again the estimates of the parameters are by no means independent. The inverse relationship between estimates of the flexibility parameter ν and the (now only) noise level σ_f is even of an almost perfect kind. Compared to the correlation coefficients in Table 4, the relationship between α_d and ν (and σ_f), which is negative in Table 4 (weakly positive, for both σ_f and σ_c), is now positive (negative, respectively). The

	ϕ	ν	α_d	σ_f
$\phi:$	1.00	−0.72	−0.17	0.74
$\nu:$		1.00	0.65	−0.98
$\alpha_d:$			1.00	−0.67
$\sigma_c:$				1.00

Table 8: Pairwise correlations of USD–DEM parameter estimates.

other relationships maintain their sign but are moderately, or even drastically, stronger. Together Tables 7 and 8 convey concise information about the parameter values that most suitably describe the agents' behaviour on the foreign exchange market in our model.

8. Conclusion

Although any structural model is false and therefore, trivially, cannot be expected to reproduce reality in all its facets, one would often like to employ formal estimation methods to identify (a range of) suitable numerical parameter values. "Suitable", accordingly, means that the model captures reality along certain dimensions to a certain degree. One advantage of using the method of simulated moments (MSM) for estimation is that it obligates the researcher to put forward summary statistics, or moments, that can clarify his or her preferred dimensions. The estimation results themselves are also transparent and easy to interpret. In addition, the MSM approach has no problems in treating unobservable variables and nonlinearities in the model dynamics, and it usually requires less computational effort than other simulation-based estimation procedures. Having a small-scale asset pricing model at hand that looked very promising regarding its ability to match the salient 'stylized facts' of the daily returns of (at least) the S&P 500 stock market index (Franke and Westerhoff, 2009), these were all good reasons for us to apply MSM for a more rigorous test. The moments we chose for this purpose capture, in particular, the fat tails and the long memory in the absolute returns.

Rather than relying on the asymptotic theory of MSM, in this paper we used bootstrap methods that can better account for the small-sample properties of the estimation. In particular, we thus got an empirically based distribution of the objective function J, to which one could compare a Monte Carlo (MC) distribution of J obtained from a large number of simulation runs with the same parameter set. The percentage of the MC distribution, the values J of which are less than the 95% quantile of the bootstrap distribution, can then be conceived of as a proxy for the model's p-value. It was furthermore argued that this proxy is (probably strongly) downward biased. In any case, it was

31

a safe conclusion that at least one-fourth (if not many more) of the sample runs of the model over the empirical time horizon cannot be recognized as being inconsistent with the empirical returns. In this sense, our model was not rejected by the data.

A comparison of the bootstrap and MC distributions of the single moments gave us more information along which dimensions the model performs better or worse, although we also conducted an experiment that warns against premature conclusions. A battery of re-estimations of the model with different random number sequences for the simulations provided us with distributions of the parameters. As all of the one- and two-dimensional confidence intervals constructed from them were relatively narrow, identification of the model proved to be no problem. This feature was of special importance in the last section when the model was estimated on the US dollar / German mark exchange rate. Here two of the six parameters could be fixed at zero, and the confidence intervals of the remaining four did not overlap with those from the S&P 500. Considering the economic meaning of the parameters, the estimation therefore revealed that these stock and foreign exchange markets are different, offering us an interpretation of in which respect.

A further virtue of MSM is that the values of its objective function are directly comparable across models. Being now fairly satisfied with the results for our model, we wish to see how it will perform relative to other models of similar complexity. These can be existing "rival" models, or developments or modifications of the present model, or combinations of some of its elements (basically that of structural stochastic volatility) with the building blocks from other models. In general, a wider perspective is thus opened up. By providing a reliable criterion to evaluate models of very different kinds, MSM allows us to make progress in coping with the anecdotal wilderness that, after this research has been firmly established, now threatens the field of agent-based models of financial markets.

References

ALFARANO, S. AND FRANKE, R. (2007): A simple asymmetric herding model to distinguish between stock and foreign exchange markets. Working Paper, University of Kiel. (www.bwl.uni-kiel.de/gwif/downloads_papers.php?lang=en).

CARRASCO, M. AND FLORENS, J.-P. (2002): Simulation-based method of moments and efficiency. *Journal of Business and Economic Statistics*, 20, 482–492.

CHEN, S.-H., CHANG, C.-L. AND DU, Y.-R. (2008): Agent-based economic models and econometrics. Paper presented at the Econophysics Colloquium, Kiel, August 2008 (www.bwl.uni-kiel.de/gwif/econophysics/abstracts/chen.pdf).

CONT, R. (2001): Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1, 223–236.

DAVIDSON, R. AND MACKINNON, J.G. (2004): *Econometric Theory and Methods*. Oxford University Press: Oxford.

FRANKE, R. (2009): Applying the method of simulated moments to estimate a small agent-based asset pricing model. *Journal of Empirical Finance*, 16, 804–815.

FRANKE, R. AND WESTERHOFF, F. (2009): Validation of a structural stochastic volatility model of asset pricing. Working Paper, Universities of Kiel and Bamberg. (`www.bwl.uni-kiel.de/gwif/downloads_papers.php?lang=en`).

GILLI, M. AND WINKER, P. (2003): A global optimization heuristic for estimating agent based models. *Computational Statistics and Data Analysis*, 42, 299–312.

LEE, B.-S. AND INGRAM, B.F. (1991): Simulation estimation of time series models. *Journal of Econometrics*, 47, 197–205.

LUX, T. (2009b): Stochastic behavioural asset-pricing models and the stylized facts. In: Hens, T. and Schenk-Hoppé, K.R. (eds.): *Handbook of Financial Markets: Dynamics and Evolution.* North-Holland: Amsterdam; pp. 161–216.

LUX, T. AND AUSLOOS, M. (2002): Market fluctuations I: Scaling, multiscaling, and their possible origins. In: Bunde, A., Kropp, J. and Schellnhuber, H. (eds): *Science of Disaster: Climate Disruptions, Heart Attacks, and Market Crashes.* Springer: Berlin; pp. 373–410.

MANZAN, S. AND WESTERHOFF, F. (2005): Representativeness of news and exchange rate dynamics. *Journal of Economic Dynamics and Control*, 29, 677–689.

MENKHOFF, L. AND TAYLOR, M. (2007): The obstinate passion of foreign exchange professionals: technical analysis. *Journal of Economic Literature*, 45, 936–972.

PRESS, W.H. ET AL. (1986), *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press: Cambridge, UK.

WESTERHOFF, F. (2008): The use of agent-based financial market models to test the effectiveness of regulatory policies. *Jahrbücher für Nationalökonomie und Statistik* (Journal of Economics and Statistics), 228, 195–227.

WINKER, P., GILLI, M. AND JELESKOVIC, V. (2007): An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination*, 2, 125–145.