Structural Stochastic Volatility in Asset Pricing Dynamics: Estimation and Model Contest

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

October 2011

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

Abstract

In the framework of small-scale agent-based financial market models, the paper starts out from the concept of structural stochastic volatility, which derives from different noise levels in the demand of fundamentalists and chartists and the time-varying market shares of the two groups. It advances several different specifications of the endogenous switching between the trading strategies and then estimates these models by the method of simulated moments (MSM), where the choice of the moments reflects the basic stylized facts of the daily returns of a stock market index. In addition to the standard version of MSM with a quadratic loss function, we also take into account how often a great number of Monte Carlo simulation runs happen to yield moments that are all contained within their empirical confidence intervals. The model contest along these lines reveals a strong role for a (tamed) herding component. The quantitative performance of the winner model is so good that it may provide a standard for future research.

JEL classification: D84; G12; G14; G15.

Keywords: Method of simulated moments; moment coverage ratio; herding; discrete choice approach; transition probability approach.

1. Introduction

Over the past ten or twenty years, a rich literature on small-scale asset pricing models has come into bloom that abandon the rational expectations framework and rather consider the financial markets as being populated by a few groups of agents who rely on simple

heuristic trading strategies.[2] Currently numerical simulations tend to become the rule. They do not only illustrate the basic mechanisms but also attempt to match certain stylized facts in quantitative ways. Contributing to this research, the present paper has three main and, as we may claim, innovative goals: (1) the specification of different versions of a promising modelling approach featuring what we call structural stochastic volatility; (2) the estimation of these models on daily returns from a major stock market index by two versions of the transparent and easy-to-apply method of simulated moments; and (3) a competitive comparison of the estimation results. We are confident that the methods and the results thereby obtained can set a new standard for the evaluation of asset pricing models, although of course the reader may judge this on his or her own at the end of the paper.

Let us begin with a brief characterization of the concept of structural stochastic volatility. The most elementary form in which we model it here is based on the usual archetypes of financial markets participants, i.e. chartists who extrapolate past price trends and fundamentalists who bet on mean reversion. However, the asset demand of the two types of agents is noisy and the corresponding noise levels may also be different. If then the market shares of fundamentalists and chartists are varying over time, the overall noise level in the demand-induced price changes will be varying, too. In addition, a dominance of chartists favours the emergence of bubbles and a dominance of fundamentalists is conducive to more tranquil market periods. In this way there is scope for the returns to exhibit stochastic volatility. In contrast to the technical GARCH models and their refinements, this phenomenon has now a structural, although parsimonious, underpinning.

To make this general idea workable, we next have to ask what governs the endogenous switches between the two trading strategies. Four socio-economic principles will be considered in this respect: (a) *predisposition* as a behavioural bias towards one of the trading strategies; (b) *differential wealth* that hypothetically would have been earned by the two strategies over the past (making a strategy more attractive the more successful it has been); (c) *herding*, which means that the attractiveness of a group rises with the share it already has in the total population; (d) a *misalignment* correction mechanism, according to which high deviations of the price from its fundamental value increase the fears that the bubble will eventually burst and, as a result, fundamentalism is expected to become more profitable again.

Choosing in various ways from this set, the single mechanisms are additively combined in a switching index which summarizes the relative attractiveness of fundamentalism versus chartism. The models are then completed by incorporating the index into one of two different devices. The first one is the well-known discrete choice approach along the lines of Brock and Hommes (1997), where the switching index determines the fractions

---

[2] Recent surveys on this research are Hommes (2006), LeBaron (2006), Chiarella et al. (2009), Lux (2009) and Westerhoff (2009), among others.

of the two groups of traders directly by a nonlinear functional relationship. In the other approach, which goes back to Weidlich and Haag (1983) and Lux (1995), the switching index influences the probabilities with which the agents switch to the opposite group. In the aggregate it is here the *change* of the market fractions that is determined by the index. On the whole, seven model variants will be studied in greater detail.

While until a few years ago researchers have contended themselves with calibrating their models in informal ways, there are now a number of attempts to estimate them in a more systematic manner.[3] A most suitable approach for the present kind of models is the method of simulated moments (MSM). 'Moments' refers to the time series of one or several variables and means certain summary statistics computed from them, the empirical values of which the model-generated moments should try to match. In the present context the moments will reflect what is considered to be the most important stylized facts of the daily stock market returns, in particular, volatility clustering and fat tails. After all, this is what the evaluation of the models in the literature usually centres around.[4]

In pursuing the second goal of the paper, we first follow the standard specification of MSM. It searches for the parameter values of a model that minimize the distance between the empirical and simulated, i.e. model-generated moments, where the distance is defined by a quadratic loss function. In addition, we will introduce the joint *moment coverage ratio* (MCR) as an alternative evaluation criterion. This is the percentage ratio of a great number of simulation runs of the model the moments of which are all contained within their empirical confidence intervals. Since the models are thus required to match nine (rather diverse) moments, a ratio of more than five per cent, say, would already be quite an achievement.

After the ground has thus been prepared, we can turn to the third goal of the paper and ask the most obvious question: which of our models is the best? In particular, we find out here that the two estimation criteria, although being similar in spirit, convey different information, so that even the ranking of the models may be affected. As a consequence, we estimate the models for each criterion separately. To anticipate the answer to our question, in both cases the winner is the discrete choice model that includes herding, predisposition, and the misalignment effect. Defining a (moment-specific) bootstrapped *p*-value to quantify a model's goodness-of-fit, it turns out that roughly one-third of all simulation runs cannot be directly rejected by the data. On the other hand, the joint MCR amounts to more than 25 per cent, which we think is a fairly respectable order of

---

[3] See Gilli and Winker (2003), Alfarano et al. (2005), Manzan and Westerhoff (2007), Winker et al. (2007), Boswijk et al. (2007), Amilon (2008), Franke (2009), Li et al. (2010), Chiarella et al. (2011), Franke and Westerhoff (2011).
[4] The choice of MSM does not rule out that also other estimation approaches may be tried. For a brief summary of the comparative advantages of MSM see Franke (2009, pp. 804f). In our opinion, its main merit is the high transparency in the evaluation of a model's goodness-of-fit.

magnitude as well.

It is furthermore remarkable that both estimation procedures also yield the same model in second place. This is the model that incorporates the same three effects into the transition probability approach. We learn from this result that what makes the notion of structural stochastic volatility most successful is an appropriate choice of the socio-economic principles governing the switching between the two trading strategies, whereas the differences between the discrete choice and transition probability approach seem to be of second order importance.

The remainder of the paper is organized as follows. Section 2 reiterates the constituent parts of the method of simulated moments and specifies the set of moments that will be underlying the estimations in the subsequent sections. Section 3 presents the model variants that we select for estimation. It is here also illustrated that they can give rise to rather different patterns of the switching between chartism and fundamentalism. The estimation results and their comprehensive discussion are contained in Section 4. Section 5 concludes, and an appendix details some technical issues.

2. The method of simulated moments

The models that we will be studying further below have been designed to explain—at least to some extent—the most important stylized facts of financial markets at a daily frequency.[5] Regarding the empirical data, we will limit ourselves to the S&P 500 stock market index with $T = 6866$ daily observations from January 1980 to mid-March 2007. With respect to the (log) price changes, or returns, the models aim to match the four features that have received 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).[6] The returns themselves are specified in percentage points, so with respect to the log prices p_t we may write

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

For a quantitative analysis, the features of interest are assessed by a number of summary statistics or, synonymously, moments. The first moment that the models in the following sections will seek to match is the first-order autocorrelation coefficient of the raw returns. The requirement that it be close to zero should limit the chartists' price extrapolations in

[5] Detailed descriptions of the statistical properties of asset prices can be found in Cont (2001), Lux and Ausloos (2002), or Lux (2009).
[6] Generally, one might also include a negative skewness of stock returns. Our fairly parsimonious models, however, will not provide for any asymmetry in this respect. A referee has pointed out that the existence of jump behaviour is another stylized fact (see Chan and Maheu, 2002, and Eraker, 2004), which our models will not be able to account for, either. These aspects might be a route for further and more ambitious research.

the models. We have checked that if the model-generated coefficient becomes insignificant, then all autocorrelations at longer lags will practically vanish, too. This lack of additional information is the reason why only one moment of the raw returns is used.

The other moments are concerned with the volatility issues and so have the absolute returns underlying. To begin with, we have to take care that the models can suitably scale the overall volatility. To this end we employ the mean value of the absolute returns. Next, the fat tail property is measured by the well-known Hill estimator of the tail index of the absolute returns.[7]

The Hill estimator is known to suffer from a small-sample bias. Fortunately, we need not be much concerned about this since the expressions describing the bias are typically invariant with respect to the relative tail size, for which we therefore universally assume a customary 5 per cent ratio. Hence, more scrupulously, what the models are trying to match is a summary statistic that, as an estimate of a 'true' tail index, is subject to the same bias in the empirical and model-generated data.[8]

In order to capture the long memory effects, we invoke the autocorrelation function (ACF) of the absolute returns up to a maximal lag length of 100 days. As the function is slowly decaying without becoming insignificant at the long lags, we have an entire profile to match. However, just because of this regularity, not all of the autocorrelation coefficients need to be considered. We view the profile as being sufficiently well represented by the coefficients for the six lags $1, 5, 10, 25, 50, 100$. To reduce the influence of accidental outlier effects that might show up at one of the selected lags, we also smooth them by computing the centred three-lag averages.[9]

The performance of the models that will be set up further below is evaluated on the basis of these nine moments. We summarize them in a (column) vector $m = (m_1, \ldots, m_9)'$ (the prime denotes transposition). They are to come close to their empirical counterparts, which may be designated m^{emp} (the values of the latter can be found in Table A1 in the Appendix). The distance between m and m^{emp} is defined by a quadratic function with a suitable weighting matrix $W \in \mathbb{R}^{9 \times 9}$ (to be specified shortly). We thus have the following loss function for the model-generated (and other) moments,

$$J \;=\; J(m) \;:=\; (m - m^{emp})'\, W\, (m - m^{emp}) \tag{2}$$

---

[7] There is no need to distinguish between extreme positive and negative returns because, as mentioned in the previous footnote, all of the models are symmetrical in this respect.

[8] The precise formula for the bias term depends upon the null hypothesis for the distribution function of the extreme returns. For examples see Huisman et al. (2002) or Caeiro et al. (2005).

[9] 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-range dependencies between the returns. To take additional account of the latter in future work, an anonymous referee has suggested including a "fractional integration parameter" in the list of moments. For an agent-based model that seeks to match this statistic, see He and Li (2007, Section 4.2).

In the first instance, this function serves to estimate a model by the method of simulated moments (MSM). To this end, it is simulated over S periods, or days (after an initial period of several hundred days is discarded to rule out any transitory effects). The simulations are repeated for alternative sets of the structural parameters of the model, and estimation means to find parameters that minimize the associated loss. It goes here without saying that for each parameter set the same random number sequence is used, which in the following discussion will be supposed to be given.[10] Collecting the parameters to be estimated in a vector θ, letting Θ be the admissible set, and denoting the moments of the model to which a vector θ gives rise by $m = m(\theta; S)$, the estimated parameters $\widehat{\theta}$ are determined by

$$\widehat{\theta} \;\; = \;\; \arg\min_{\theta \in \Theta} \; J[m(\theta; S)] \tag{3}$$

To reduce the sample variability in the stochastic simulations, the time horizon is chosen longer than the empirical sample period T. As it is common practice, we will work with $S = 10 \cdot T$. Regarding the weighting matrix W entering the objective J, 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^{emp} 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,

$$W \;\; = \;\; \widehat{\Sigma}^{-1} \tag{4}$$

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, pp. 201f, or the application of MSM in Franke, 2009, Section 2.2). In this paper, we follow Winker et al. (2007) and Franke and Westerhoff (2011) and choose a bootstrap approach to construct from the empirical returns a large number of samples of the moments, from which subsequently the covariances can be derived. The details are given in Appendix A2.

After solving the minimization problem in (3), it would be possible to determine the asymptotic standard errors of the parameter estimates (see Lee and Ingram, 1991, p. 202). Since the main focus of this paper is on the fitting properties of alternative models, we will, however, leave the issue of the precision of the single estimates aside. Suffice it to note that all of the parameter values that we will report are significantly different from zero.

---

[10] Entering the models will be normally distributed random terms ε_t with variance σ_t^2. Identical random number sequences then mean 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 subsequently obtained as $\sigma_t \tilde{\varepsilon}_t$.

3. Alternative switching mechanisms to get structural stochastic volatility

3.1. *The common model building blocks of demand and its price impact*

The two groups of fundamentalist and chartist traders, their specification of demand for the asset, and the impact of aggregate demand on the price are common to all of the model versions that we consider.[11] To begin with the latter, the market is generally allowed to be in disequilibrium. A market maker is assumed to hold inventory, from which he serves any excess of demand and to which he puts any excess of supply. He reacts to this imbalance by proportionately adjusting the price for the next period with a (constant) factor $\mu > 0$ in the direction of excess demand. Thus, letting p_t be the (log) price that he quotes at the beginning of period t, d_{t-1}^f and d_{t-1}^c the demand in the previous period of an average fundamentalist and chartist trader, respectively, and n_{t-1}^f, n_{t-1}^c the market fractions of the two groups in that period, the price impact equation reads,

$$p_t \;=\; p_{t-1} \;+\; \mu\,(n_{t-1}^f\,d_{t-1}^f \;+\; n_{t-1}^c\,d_{t-1}^c) \tag{5}$$

Regarding the formulation of demand, we join numerous examples in the literature and, in the first step, postulate two extremely simple deterministic rules. They govern what we may call the core demand of an average trader in each group. For the fundamentalists, this demand is inversely related to the deviations of the price from its fundamental value. That is, in period t it is proportional to the gap $(p^\star - p_t)$, $p^\star$ being the (log of the) fundamental value, which we treat as an exogenously given constant (for simplicity and to show that no random walk behaviour of the fundamental value is required to obtain the stylized facts). The core demand of the group of chartists is hypothesized to be proportional to the returns they have just observed, i.e. $(p_t - p_{t-1})$.

A crucial feature of our models 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 a certain within-group heterogeneity, which we do not wish to describe in full detail. Since the many individual digressions from the simple rules as well as their composition in each group will more or less accidentally fluctuate from period to period, it is a natural short-cut to have this heterogeneity represented by two independent and normally distributed random variables ε_t^f and ε_t^c for the fundamentalists and chartists, respectively.[12] Combining the

---

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

[12] For example, individual and presently active traders with a fundamentalist strategy may adopt different values for their fundamental price, they react with different intensities to their trading signal, or they experiment with more complex trading rules which may also be continuously subjected to further modifications. Similarly so for the chartists, which explains the independence of ε_t^f and ε_t^c. In short, the two noise variables can be conceived of as a most convenient short-cut of certain aspects that are more specifically dealt with in models with hundreds or thousands

deterministic and stochastic elements, the net demands of the average fundamentalist and chartist trader for the asset in period t are given by

$$d_t^f = \phi\,(p^\star - p_t)\, +\, \varepsilon_t^f\,, \qquad\qquad \varepsilon_t^f \sim N(0, \sigma_f^2) \qquad\qquad (6)$$

$$d_t^c = \chi\,(p_t - p_{t-1})\, +\, \varepsilon_t^c\,, \qquad\qquad \varepsilon_t^c \sim N(0, \sigma_c^2) \qquad\qquad (7)$$

where here and in the following, the Greek symbols denote constant and nonnegative parameters.

Plugging the lagged equations (6) and (7) into the price impact function (5), we do get one single noise term acting on the price changes, or returns (as the sum of two normal distributions). Its variance σ_t^2 is, however, dependent on the variations of the market fractions of the fundamentalists and chartists, $\sigma_t^2 = (n_{t-1}^f)^2\,\sigma_f^2 + (n_{t-1}^c)^2\,\sigma_c^2$, even if the two group-specific variances σ_f^2 and σ_c^2 were equal. This feature of a time-varying variance in the returns is what was coined *structural stochastic volatility* (SSV) in Franke and Westerhoff (2009). Nevertheless, whether in this way also the most important stylized fact of the returns can be matched by the model will, in the first instance, be a matter of the specific switching mechanisms between the two groups of traders. A number of different proposals for them will be introduced next.

Before, a general feature is worth pointing out that will prevail in all of the model versions. First, in a pure chartist regime, $n_t^c \equiv 1$, the two-dimensional price process is easily seen to have a zero and a unit root. Second, in a pure fundamentalist regime, $n_t^f \equiv 1$, the root of the one-dimensional price dynamics is $1 - \mu\phi$, where in the estimations the product $\mu\phi$ turns out to be around 0.01 or less. [13] Hence in all of the models there is broad scope for persistent price misalignment, which is another stylized fact of financial markets although it is not included in the list of our moments. [14]

3.2. *The transition probability and discrete choice approaches*

There are two by now fairly standard formulations in the literature that specify how the agents in a simple model may switch between two or more attitudes, or groups: the transition probability approach and the discrete choice approach. For easier reference we may also use the acronyms TPA and DCA, respectively.

To describe the transition probability approach, we first introduce the concept of a switching index $a = a_{t-1}$, which is uniform across all agents. For the moment being it suffices to note that a_{t-1} represents the relative attractiveness of fundamentalism *versus* chartism at the end of period $t-1$ (precise specifications are discussed below). The basic

---

of different agents that one would have to keep track of over time (see Farmer and Joshi, 2002; LeBaron, 2006).

[13] We owe these observations to one of the referees.

[14] A historical perspective on the recurrent phenomenon of price misalignments is given by Kindleberger (2000). Shiller (2005, Ch. 1) demonstrates that also the S&P 500 continually disconnects from its fundamental value.

idea of TPA is that the switches of the individual agents only occur with a certain probability. More specifically, rising values of the index increase the probability π^{cf} of a chartist becoming a fundamentalist in period t, and decrease the probability π^{fc} of a fundamentalist becoming a chartist; and *vice versa* if the index is falling. Assuming a large population together with linearity and symmetry in the relative changes of the two switching probabilities, it can be demonstrated that, at the macroscopic level, the effects of the probabilistic elements disappear and the adjustments of the market fractions are governed by the deterministic equation, [15]

$$
\begin{aligned}
n_t^f &= n_{t-1}^f + n_{t-1}^c \, \pi^{cf}(a_{t-1}) - n_{t-1}^f \, \pi^{fc}(a_{t-1}) \\
n_t^c &= 1 - n_t^f \\
\pi^{cf}(a_{t-1}) &= \min[\,1, \nu \, \exp(a_{t-1})\,] \\
\pi^{fc}(a_{t-1}) &= \min[\,1, \nu \, \exp(-a_{t-1})\,]
\end{aligned}
\tag{TPA}
$$

The application of the discrete choice approach in the present framework of fundamentalist and chartist strategies usually involves payoff indices u^f and u^c that are derived from the past capital gains of the two groups. On this basis, their market shares n_t^s ($s = f, c$) are determined by the formula $n_t^s = \exp(\beta u_{t-1}^s) \, / \, [\exp(\beta u_{t-1}^f) + \exp(\beta u_{t-1}^c)]$, where β is the so-called intensity of choice. Dividing numerator and denominator by $\exp(\beta u_{t-1}^f)$ we can, with respect to the fundamentalists, also write $n_t^f = 1 \, / \, \{1 + \exp[-\beta(u_{t-1}^f - u_{t-1}^c)]\,\}$. That is, only the differential payoffs need to be considered. Moreover, $(u_{t-1}^f - u_{t-1}^c)$ may generally represent the difference of any payoff or utility variables. Since this term would measure the relative attractiveness of fundamentalist trading, we may change the notation and replace $(u_{t-1}^f - u_{t-1}^c)$ with the attractiveness index a_{t-1}. In this way the discrete choice approach becomes,

$$
\begin{aligned}
n_t^f &= \frac{1}{1 + \exp(-\beta \, a_{t-1})} \\
n_t^c &= 1 - n_t^f
\end{aligned}
\tag{DCA}
$$

It will be noted that, as it should be and as it is the case in (TPA), an increase in the index a_{t-1} increases the market share of the fundamentalists. It thus makes sense to use the same specification of a_{t-1} in (DCA) and (TPA). The main difference between the two approaches is then that in (DCA) the impact of a_{t-1} is directly on the *level* of n_t^f (via the ratio $1/[1 + \exp(-\beta \, a_{t-1})]$), whereas in (TPA) it is on the *change* $(n_t^f - n_{t-1}^f)$ (via the term $n_{t-1}^c \pi^{cf}(a_{t-1}) - n_{t-1}^f \pi^{fc}(a_{t-1})$). This is, of course, not negligible in a hypothetical situation where the relative attractiveness stays constant over a certain period of time. On the other hand, in the fully dynamic models with the endogenous deterministic and stochastic interactions of the variables, the differences might possibly be less substantial,

---

[15] See, e.g., Franke (2008, pp. 309ff) or Franke and Westerhoff (2009, Section 2) for more details.

provided that some of the parameters are suitably adjusted. This is indeed an important topic that we will have an eye on in the estimations below.

3.3. *Specification of the relative attractiveness*

We consider four principles that may play a role in the agents' choice of one of the two strategies, which as we have seen is tantamount to the determination of the relative attractiveness. Correspondingly, we put forward four components some of which, in various combinations, are added to yield the level of attractiveness a_t at the end of period t. The first principle is herding (**H**), which means that joining a group becomes the more attractive the more adherents it already has. This idea is straightforwardly represented by a term that is proportional to the most recent difference $(n_t^f - n_t^c)$ between the two market fractions.

The second principle is based on the abovementioned differential profits, where the modelling literature usually incorporates some inertia.[16] With respect to strategy $s = f, c$, let to this end g_t^s be the short-term capital gains that an average agent in this group could realize at day t. They derive from the (average) demand formulated at $t-2$ and executed at the price p_{t-1} of the following day. Furthermore, let u_t^s be the 'utility' obtained from these capital gains and η a memory coefficient between zero and one. The specification then typically reads $u_t^s = g_t^s + \eta\, u_{t-1}^s$. We follow the same idea but consider it more meaningful to have a weighted average of g_t^s and u_{t-1}^s on the right-hand side of this equation. So we work with

$$
\begin{aligned}
g_t^s &= [\exp(p_t) - \exp(p_{t-1})]\, d_{t-2}^s\,, \\
w_t^s &= \eta\, w_{t-1}^s + (1-\eta)\, g_t^s\,,
\end{aligned}
\qquad s = c, f \qquad (8)
$$

The reason for our slight modification and for replacing the symbol u with w is the interpretation to which (8) gives rise. In fact, the second equation in (8) is equivalent to the infinite sum

$$
w_t^s = (1-\eta) \sum_{k=0}^{\infty} \eta^k\, g_{t-k}^s
$$

Apart from the rescaling by $(1-\eta)$, w_t^s therefore represents the accumulated profits, discounted by the coefficient $\eta < 1$, that would have been earned by a hypothetical agent who consistently had followed strategy s over the infinite past on a day-to-day basis. In other words, w^s is the hypothetical wealth attributable to strategy s. Equipped with the concept of eq. (8), the second component determining a_t is proportional to the difference $(w_t^f - w_t^c)$. This feature may be identified by the letter (**W**).

---

[16] Occasionally, reference is also made to the squared forecast errors of the agents (in Brock and De Fontnouvelle, 2000, for example). We view this as a more indirect specification of differential profits (which would derive from them) and neglect it to limit our investigations.

In addition to herding and differential wealth, there may also be a certain predisposition ($\mathbf{P}$) towards one of the strategies. This is simply expressed by a constant α_o, which is positive (negative) if this *a priori* preference is for fundamentalism (or chartism, respectively). Finally, the fourth principle captures the idea that chartism may appear progressively riskier as prices move further away from the fundamental value. Hence a_t would rise in proportion to the price misalignment ($\mathbf{M}$), which is conveniently measured by the squared deviations of p_t from $p^\star$.

These four components to set up a_t can be combined in many different ways (15, to be exact) and substituted in TPA or DCA. Principally, including all four of them should yield the best matching of the moments that we are interested in. As it turns out, however, for both TPA and DCA the improvement of a model version over a version that it nests, or that is more parsimonious, is not always very substantial. On the other hand, satisfactory results for the versions including the herding component require a certain taming by the simultaneous effects from predisposition as well as misalignment.[17] Since it is somewhat cumbersome, unless tiring, to report all details in this respect, we will concentrate on the results that we found most relevant. Concerning the determination of the relative attractiveness, the following four combinations are considered:

wealth:
$$a_t = \alpha_w \left(w_t^f - w_t^c\right) \tag{W}$$

wealth + predisposition:
$$a_t = \alpha_w \left(w_t^f - w_t^c\right) \,+\, \alpha_o \tag{WP}$$

wealth + herding + predisposition:
$$a_t = \alpha_w \left(w_t^f - w_t^c\right) \,+\, \alpha_n \left(n_t^f - n_t^c\right) \,+\, \alpha_o \tag{WHP}$$

herding + predisposition + misalignment:
$$a_t = \alpha_n \left(n_t^f - n_t^c\right) \,+\, \alpha_o \,+\, \alpha_p \left(p_t - p^\star\right)^2 \tag{HPM}$$

where it is understood that the coefficients α_w, α_n and α_p are strictly positive, while α_o may attain any sign.

3.4. *Some preliminary comparative results*

The results that we find worth discussing for the discrete choice approach include all four of these combinations, while for the transition probability approach the results from (WHP) can be safely neglected. Thus, on the whole, we will discuss seven variants of our model family featuring structural stochastic volatility. Since this is from now on understood, it will be more convenient to refer to these variants as 'models' of their own.

---

[17] If in the discussions of the estimation results below "herding" is compared to the other versions, predisposition and misalignment may be meant to be included.

For all models, we normalize the demand terms in the price impact function (5) by setting $\mu = 0.010$. For the DCA models, we fix the intensity of choice at $\beta = 1.00$, which is just a matter of scaling our measure a_t of relative attractiveness of fundamentalism. For the TPA models, we fix the flexibility parameter at $\nu = 0.050$. This approximately amounts to scaling a_t (it would be exact if $\exp(\cdot)$ were a linear function). Together with the other numerical parameters, our choice for ν implies that the upper limit for the individual transition probabilities in (TPA) practically never becomes binding.

	ϕ	χ	η	α_w	α_o	α_n	α_p	σ_f	σ_c
DCA									
W :	1.00	1.20	0.991	1580	—	—	—	0.681	1.724
WP :	1.00	0.90	0.987	2668	2.100	—	—	0.752	1.726
WHP :	1.00	0.90	0.987	2668	2.100	1.28	—	0.741	1.705
HPM :	0.12	1.50	—	—	−0.327	1.79	18.43	0.758	2.087
TPA									
W :	1.15	0.81	0.987	1041	—	—	—	0.715	1.528
WP :	1.00	0.83	0.987	1078	0.376	—	—	0.736	1.636
HPM :	0.18	2.30	—	—	−0.161	1.30	12.50	0.790	1.900

Table 1: Parameters of seven model variants estimated by MSM.

Note: Common to all models is the normalization $\mu = 0.01$. In addition, $\beta = 1$ is common to the DCA versions and $\nu = 0.05$ to the TPA versions.

With these conventions, we choose an arbitrary random number seed for the simulation of the stochastic trajectories and subject each model to the MSM estimation procedure described in the previous section. For a more detailed analysis of these results, we settle down on the rounded parameter values collected in Table 1. We report no standard errors since as mentioned before, we are mainly interested in the absolute and comparative goodness with which the empirical moments can be matched.[18] All of the parameters are, however, significantly different from zero.

It is interesting to note that in DCA as well as TPA the predisposition parameter α_o changes sign from WP (and WHP) to HPM. This shows that one has to be careful when comparing single parameters across models. On the other hand, it is remarkable that

---

[18] Since these evaluations will be based on a (very) large number of simulation runs, their reliability is not compromised by the specific random number seed underlying the first MSM estimation. If anything, skillfully chosen slight changes in the parameters of Table 1 might lead to slight improvements in the goodness-of-fit.

the parameters of corresponding versions of DCA and TPA have an order of magnitude that is not dramatically different (except α_o). This property of the model estimations may be taken as a first indication that the differences between the basic approaches of DCA and TPA are less important than different specifications of the index of relative attractiveness itself.

Figure 1 presents sample runs of four models over 7000 periods, i.e. 7000 days. The comparison of the evolution of the market fraction of chartists illustrates that the models can give rise to fairly different patterns of the switching between chartism and fundamentalism, although the random number sequences in the demand terms are identical. The smallest differences are observed between the most elementary versions of DCA and TPA, when only the wealth effect enters the determination of the relative attractiveness. On the whole, they produce rather similar movements of n_t^c, but DCA exhibits more variability over shorter periods of time. This can be explained by the fact that in DCA the switching index directly determines n_t^f, while in TPA it acts on the change $n_t^f - n_{t-1}^f$ and thus slows down the adjustments of the market fractions (cf. the remark on eq. (DCA) in Section 3.2).

The distinguishing feature of the two DCA models in the upper two panels in Figure 1 is a positive predisposition of the agents towards fundamentalism in DCA–WP. Since the other parameter values are not very different, the latter model has on average a lower weight of the chartists in the market. For example, we have a similarly pronounced chartist regime around $t = 5000$ in the two models, but it sooner and more dramatically breaks down in DCA–WP. Nevertheless, the overall noise level of n_t^c seems to be about the same.

The bottom panel in the figure is another indication that the main differences between the model versions are not so much to be sought in the two approaches of discrete choice and the transition probabilities, but whether in the determination of the index of relative attractiveness herding is present or not. In fact, the TPA–HPM model shows very few episodes of a chartist regime, and also otherwise chartists have a very low weight in the market.

Figure 2 selects the herding variant of the discrete choice approach, i.e. DCA–HPM, to show the implications of the switching pattern for the evolution of prices and returns. First, there are some more chartist episodes than we observed for TPA–HPM, but on the other side the market fractions of this group are less conspicuous here. Second, despite the overall dominance of fundamentalists, prices are persistently disconnected from the fundamental value. It was already noted at the end of Section 3.1 that this phenomenon is primarily brought about by the low values of the product $\mu\phi$ that we obtain. This result is remarkable since price alignment was none of our moments in the estimations.

As a third point, the returns do seem to exhibit a certain clustering in volatility, although to the naked eye it does not come out so nicely as in the empirical return series
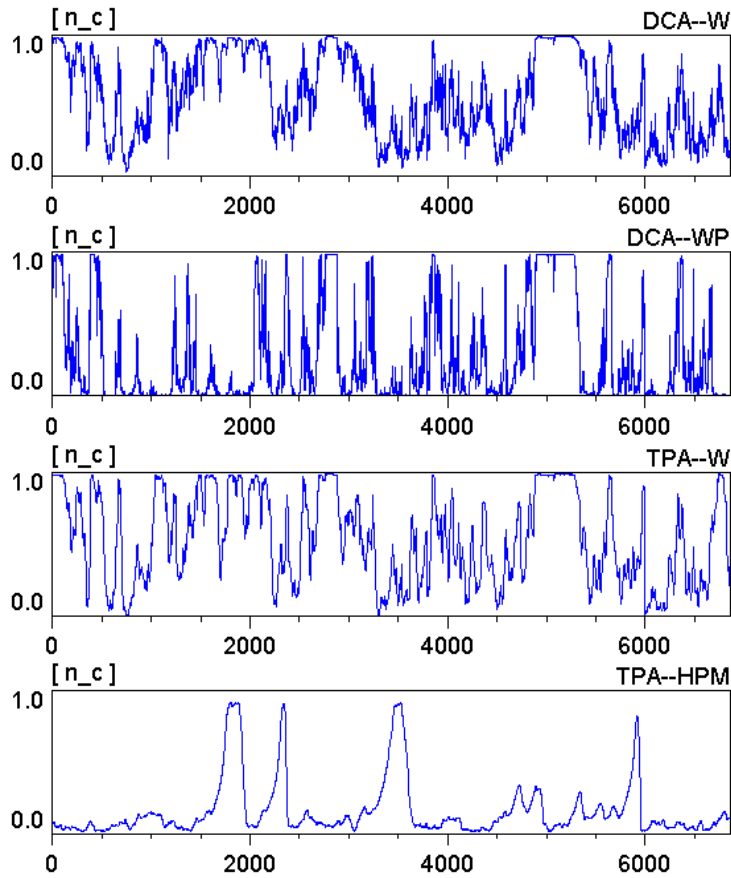
Figure 1: Market fraction of chartists in four model variants.

in the bottom panel.[19] Conforming to our intuition, the periods of higher volatility in the third panel are clearly seen to be caused by a temporary dominance of chartists in the market. This interrelationship can be observed in all of the models and we consider it a great theoretical merit of our approach to structural stochastic volatility. However, whether the model-generated clustering is satisfactorily close to the empirical phenomenon is an issue for which the precise summary statistics need to be computed and compared.

4. Moment matching and model contest

4.1. *Specification of a moment-specific p-value*

After the straightforward MSM estimations have provided us with numerical values for the structural parameters in each of our seven models, we are in the rest of the paper

---

[19] However, we could have also presented a more beautiful example in this respect.
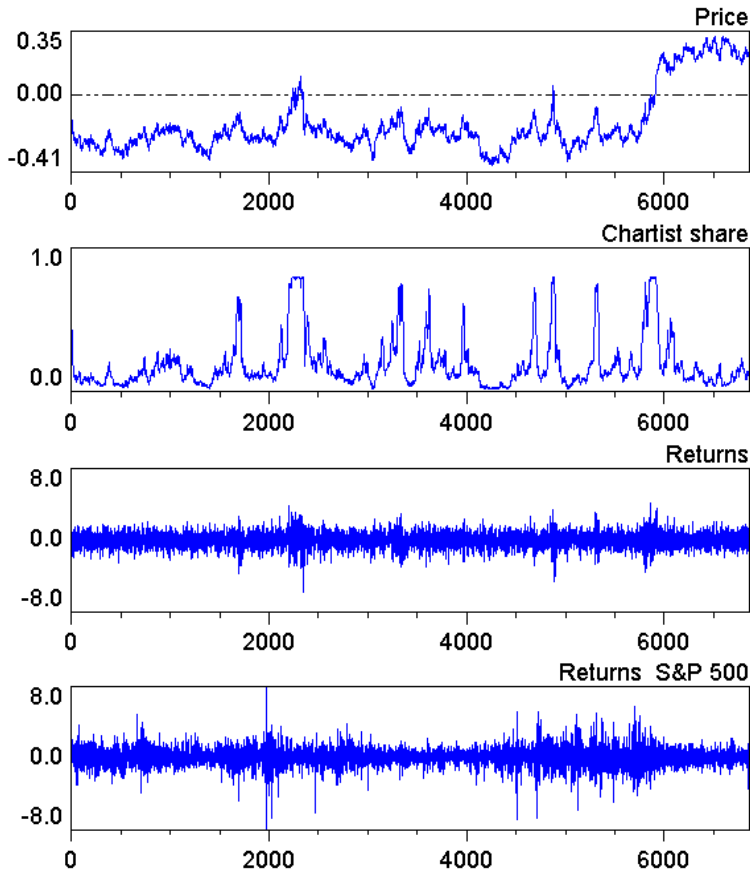
Figure 2: Simulated p, n^c, r of DCA–HPM and empirical daily returns.

concerned with a more detailed assessment of the quality of the moment matching thus achieved. The investigation begins with the most elementary question if some or all of the models would be rejected by the data. Although if taken literally this is a rather academic question since every stylized model, by definition, is 'false' anyway, it is a very convenient one that can be addressed from different angles.

Let us start out from the process of bootstrapping the moments of the empirical returns (which is described in Appendix A2). While so far it served us to derive an estimate of the variance-covariance matrix of the moments, the large sample of $B = 5000$ moment vectors obtained by this procedure can also be used to apply the loss function J to them. Doing this for all of these vectors $m^1, \ldots, m^B$, we have an entire frequency distribution of values of J at our disposal. From it, a critical value of J can be established that allows an evaluation of the values of J that characterize the numerical simulations of the models.

The basic idea is that the bootstrap distribution of J proxies the distribution that would materialize if the hypothetical construct called the real-world data generation

15

(DGP) process 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 the real-world data, and will therefore be rejected, if its J-value exceeds the 95% quantile $J_{0.95}$ of the bootstrapped J.

In finer detail it must be taken into account that the 95% quantile just described cannot serve as a yardstick for the minimized J-values from (3) in the estimation procedure. The reason is that, just in order to reduce the sample variability, these returns were simulated over a time horizon S that was more than ten times as long as the samples of the returns from the bootstrap, where $T' = 6750$. Rather, for a meaningful comparison to the bootstrapped $J_{0.95}$, the J-values of model simulations of equal length T' must be utilized.

It goes without saying that the information from a single simulation run over the relatively short bootstrap period is not reliable enough for a final evaluation of a model fit. Nevertheless, it is easy to undertake a Monte Carlo (MC) experiment and repeat these simulations many times. In this way a distribution of model-generated J-values is obtained, which can subsequently be contrasted with the bootstrap distribution.

For a precise description of this method we emphasize that the bootstrapped moments are extracted from the empirical returns by writing $m^b = m^b(\{r_t^{emp}\}_{t=1}^{T'})$.[20] Furthermore, let $\widehat{\theta}$ be the estimated vector of a model's structural parameters and indicate the different random number sequences underlying the MC simulations by the letter c, giving rise to moments $m = m^c(\widehat{\theta}; T')$ (as in the bootstrap, we work with $c = 1, \dots, 5000$). The frequency distributions from the following two sets of J-values are then considered:

$$
\begin{aligned}
\text{Bootstrap:} &\quad \{\, J[\, m^b(\{r_t^{emp}\}_{t=1}^{T'})] \,\}_{b=1}^{5000} \\
&\hspace{10em} T' = 6750 \quad\quad (9) \\
\text{Monte Carlo:} &\quad \{\, J[\, m^c(\widehat{\theta}; T')] \,\}_{c=1}^{5000}
\end{aligned}
$$

Take model DCA–W as a first example. Here we obtain $J_{0.95} = 22.45$ as the critical value of the bootstrapped loss function. As expected, the model-generated MC distribution of J is considerably wider to the right, such that $J_{0.95}$ corresponds to its 22.7% quantile. For a succinct summary of the general moment matching it may therefore be said that the model has a p-value of 22.7%, with respect to the parameter vector $\widehat{\theta}$ from Table 1, the data of the S&P 500 index, and the specific nine moments that we have chosen. This means that more than three-fourths of the simulations of the model over $T' = 6750$ days would be rejected, while this is not possible for a bit more than one-fifth of them. Hence, according to the conventional significance criteria, the DCA–W model would not

---

[20] It will be clear that these moments are those that have already been underlying the estimate of the variance-covariance matrix $\widehat{\Sigma}$ in (A16), which in turn yielded W as its inverse by eq. (4).

be discarded as being obviously incompatible wit the empirical data. In fact, a p-value of 22.7% does not seem to be a bad performance. [21]

	DCA				TPA		
W	WP	WHP	HPM		W	WP	HPM
22.7	23.5	24.1	32.6		12.7	21.3	24.0

Table 2: Moment-specific p-values with respect to the bootstrap distribution of J of the empirical returns.

The moment-specific p-values of this and the other models are documented in Table 2, and the clear winner by far with a respectable value of more than 30 per cent is: DCA–HPM. Moreover, within both switching approaches DCA and TPA, the herding variant augmented by the predisposition and misalignment effects outperforms the variants with the differential wealth effects. On the other hand, for each of the common specifications of the attractiveness index, DCA outperforms TPA.

Nevertheless, if the present limitation to our nine moments as being sufficiently representative of the most relevant stylized facts of a major stock market index is accepted, and likewise the concept of the corresponding p-value as a criterion for not rejecting a model, then all seven models have passed a first important test. We view this as substantial progress over the merely qualitative assessments of a structural asset pricing model that deemed satisfactory only a couple of years ago.

4.2. *Small-sample variability in pairwise model comparisons*

The concept of the moment-specific and simulation-based p-value of a model averages across MC simulation runs with more realistic and less realistic outcomes, thus minimizing the influence of sample variability. In this subsection we slightly change our perspective and actually start out from the phenomenon of sample variability in a direct comparison of two models. Here we are aware that although one model may be superior in terms of the p-value, it is nonetheless possible that its value of J from an arbitrarily selected simulation run is worse than that from an arbitrary simulation of the rival model. We ask how often this can happen if the experiment is repeated many times. In

---

[21] In addition, there are good reasons to expect that the, so to speak, MC simulations of the real-world DGP would yield a wider distribution of J than the bootstrap from the one realization actually observed. In this respect, the p-value of 22.7% would be an underestimation rather than an overestimation. See Franke and Westerhoff (2011, Section 4) for an elaboration of this argument.

this way we get another quantitative measure of how much better one model is than another.

	DCA			TPA		
	WHP	WP	W	HPM	WP	W
DCA						
HPM	59.4	59.9	56.6	58.7	59.4	64.3
	82.1	**81.8**	**63.4**	**77.9**	**76.3**	**81.1**
WHP		50.1	45.9	48.9	49.0	52.6
	×	48.8	25.0	44.1	38.7	41.7
WP			45.6	48.7	48.8	52.5
		×	25.6	45.3	39.9	43.0
W				53.0	53.3	58.0
			×	**69.2**	**66.6**	**71.7**
TPA						
HPM					50.1	54.0
				×	45.2	48.8
WP						54.2
					×	54.0

Table 3: Small-sample variability in pairwise comparisons of J: percentage of cases where model in row is superior to model in column.

Note: Based on 5000 MC replications for each model. First line in each row: all possible pairings of J; second line: all possible pairings of the 500 values obtained from averaging J across 10 sample runs (see text). Boldface figures emphasize the good performance of a model.

Specifically, we take the same 5000 MC simulations for each model over $T' = 6750$ days that have been underlying Table 2. We now consider all of the possible pairings of two models M1 and M2, and for each pair the values $J_1(c)$, $J_2(c')$ of the loss function from the MC runs $c, c' = 1, \ldots, 5000$. Varying the influence of their sample variability, we make use of them in two different ways. First, we compare $J_1(c)$ to $J_2(c')$ for all pairs c, c' and count the number of cases where $J_1(c) < J_2(c')$. The first entry in each cell of Table 3 reports them in per cent of all of these comparisons, which amount to a total of $5000^2 = 25$ millions.

Since the MSM estimations were based on simulations over a horizon that is ten times as long as the empirical return series, we perform a second battery of comparisons where we pool the 5000 values of the objective function of each model in (non-overlapping) samples $d = 1, \ldots, 500$ of size 10 and compute the average value for each of them. The

ratio where $J_1(d) < J_2(d')$ $(d, d' = 1, \ldots, 500)$ yields the second percentage ratio in each cell of the table, which thus derives from the pairwise comparisons of $500^2 = 250,000$ values of J.

In setting Table 3 against the results in the previous subsection, the most obvious question is whether the ranking of the models is preserved. This certainly holds true for DCA–HPM, which is superior to the rest of the models as indicated by the bold face figures in the first row of Table 3 for the pooled model comparisons. Numerically, however, the dominance might appear less pronounced, at least if we consider the simulations over the empirical time horizon.

What is also preserved in the latter context is the overall inferiority of TPA–W. The pooled comparisons, however, make it superior to DCA–WP and DCA-WHP. Interestingly, the significance of the pure wealth effects increases in greater generality. Integrating them into DCA, which formerly yielded the third-worst position in Table 2, leads now to a better result; DCA–W gains second rank since it outperforms all models except DCA–HPM. In particular, it beats the more general versions DCA–WP and DCA–WHP, and this superiority is remarkably strong for the pooled comparisons where DCA–W has a lower value of J in roughly 75% of all cases (the complement of 25.0 and 25.6 in the third column of Table 3.

It is similarly worth pointing out that TPA–WP improves from second-worst in Table 2 to third-best in Table 3, while TPA–HPM deteriorates in relation to TPA–WP and TPA–W. These examples show that although being based on the same MC simulations, Table 2 and 3 process different information: Table 3 with the direct comparisons of two models and the values of the loss function, and Table 2 with the comparison of one summarizing statistic of the frequency distributions of J. [22]

The overall lesson that we take from these observations is that, even though Table 2 can appear more meaningful because it limits the role of sample variability as far as possible, its more detailed quantitative results might not be overrated. This note of caution rather prompts us to consider an alternative criterion to evaluate the goodness of the models' moment matching.

4.3. *Moment coverage ratios*

The evaluation of the models in the previous subsections was based on the values of the objective function J. While this allowed us to compare the models relative to each other, we could not tell how "well" they are able to match the moments. An immediate method to address this question is to test the minimized value J of a model for the overidentifying restrictions, as there are more moments than parameters. If it is less than a certain critical

---

[22] Nevertheless, the deterioration in the ranking of DCA–WP and DCA–WHP might lead one to repeat their estimation with different random number seeds. We abstain from this extension of the analysis for reasons of space and since DCA–HPM is so dominant over the other DCA versions.

value derived from the χ^2-distribution, the model would be considered to be valid with respect to the selected moments. On the other hand, it is well-known that the converse case is not necessarily an indication that the model fails to mimic the empirical data in several dimensions. [23]

In this paper we use a more direct way to assess the degree of moment matching of a model, which can also take each single moment into account. All we need is the concept of a confidence interval of the empirical moments. Although it would be easiest to adopt the diagonal entries of the bootstrapped covariance matrix (A16) for this purpose, we choose an alternative approach in case one feels uneasy about the bootstrapping of the ACF at the longer lags. Except for the Hill moment, for which a theoretical formula is available, the standard errors of the moments are obtained by conceiving the autocorrelation coefficients as (nonlinear) functions of the means, or time averages, of simpler expressions of the returns and applying the delta method to them (the details of these computations are given in Appendix A1). On this basis, the 95% confidence interval of a moment is defined as the interval with boundaries ± 1.96 times the standard error around the empirical estimate. In addition, the slight smoothing of the single coefficients in Section 2 with the centred three-lag averages is taken over (the lower and upper bounds of the confidence intervals are reported in Table A1 in the appendix).

In this way we have a simple and intuitive criterion for a qualitative assessment of a given return series that, as we may emphasize, is simulated over the (relatively short) empirical time horizon: the series cannot be rejected as being incompatible with the data if all of its moments are contained in the confidence intervals. Nevertheless, because of the sample variability a single simulation is clearly not sufficient to evaluate a model as a whole.

Focussing on the ACF of the absolute returns, this is illustrated in the upper part of Figure 3. The shaded area is its confidence band, and the two bold lines result from two sample runs of model TPA–HPM. The upper (red) line is an example showing that the all-in criterion can be satisfied (with respect to the moments considered here), which is indeed a remarkable property of the model. The second (blue) line exhibits a similar pattern and is still bounded away from zero over the first 100 lags, but larger segments are outside the confidence band. The general 'shift' in the ACF can be ascribed to the number of chartist regimes occurring in the two samples over the time span of the 6866 days. The simulation producing the upper line gives rise to seven episodes where the chartists form a fair majority, while the lower bold line originates from a simulation run with only four chartist regimes. Besides, the lower thin line in Figure 3, which displays the ACF of the raw returns of one of the simulation runs, is perfectly representative and exemplifies once more that this—and the other models as well—have no difficulty in

---

[23] See Davidson and MacKinnon (2004, p. 368) for a short summarizing statement in this respect. Two elaborate studies on the subject are Kocherlakota (1990) and Mao (1990).
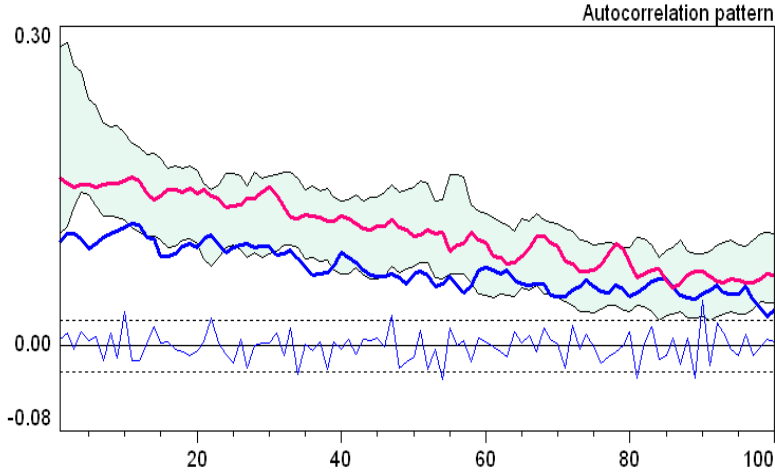
Figure 3: Sample autocorrelation functions from TPA–HPM.

Note: Shaded area: (smoothed) confidence band of the ACF of the empirical absolute returns; bold lines: (smoothed) ACF of the absolute returns from two simulation runs over the empirical time horizon; lower solid line: ACF of the raw returns from one sample; dotted lines: the corresponding Bartlett confidence band.

keeping it insignificant.

It is now obvious to think of many simulation runs of a model for each of which the confidence interval check is repeated. This leads us to the concept of a coverage ratio. That is, we count the number of Monte Carlo simulation runs for which the single moments, or all nine moments jointly, are contained in their confidence intervals and define the corresponding percentage numbers as the model's *moment coverage ratios*. More briefly, they may also be denoted by the acronym MCR.

As the concept has been introduced, the central coverage ratio is certainly the joint MCR. Its values for the seven models are presented in the first row of Table 4. It will perhaps be expected that comparatively low values of the objective function J go along, more or less, with comparatively high coverage ratios. The table, however, reveals that the minimized values of J and the corresponding joint MCRs convey different information, which even leads to a different ranking of the models. In particular, the formerly dominant model DCA–HPM now only ranks fourth, whereas TPA–HPM as the model that in Tables 2 and 3 was second or worse, now shows the best performance—and this with a distance from the other models.

There is one model (TPA–W) that misses the obvious benchmark value of 5%, and DCA–W (which formerly was among the best) is just at the margin. This is quite in contrast to their p-values with respect to the empirical bootstrap distribution of J in Table 2, which were distinctly higher than 5%. The other five models are not in open contradiction to the data, neither on the basis of J nor on the basis of the joint MCR criterion.

	bootstrapped	DCA				TPA		
	r_t^{emp}	W	WP	WHP	HPM	W	WP	HPM
joint MCR	**32.6**	5.1	12.4	13.7	10.1	0.5	6.8	**22.1**
in % of 32.6	—	15.5	38.1	41.9	31.0	1.6	20.8	**67.9**
1: r AC–1	99.8	98.0	98.6	98.1	98.4	98.8	98.7	97.1
2: Hill	81.2	10.4	61.6	**79.5**	23.4	2.0	21.0	**53.3**
3: v Mean	66.8	40.3	72.4	**75.8**	54.4	71.4	72.5	64.9
4: v AC–1	95.3	99.9	95.2	98.5	97.5	99.3	98.5	92.0
5: v AC–5	92.0	**89.6**	49.7	65.5	**72.6**	60.3	70.1	68.2
6: v AC–10	84.2	**96.0**	63.3	73.7	**88.7**	87.5	86.2	83.1
7: v AC–25	80.4	**87.3**	51.6	59.5	**87.4**	80.3	78.4	65.2
8: v AC–50	69.4	**86.6**	35.0	39.4	**76.4**	53.3	55.1	73.5
9: v AC–100	67.3	**76.3**	29.3	32.4	**69.1**	38.5	40.1	66.5

Table 4: Moment coverage ratios (in %).

Note: Boldface figures emphasize the good performance of a model with respect to the particular moment.

Beyond this qualitative statement, we would like to put the figures of the joint coverage ratios in Table 4 into a quantitative perspective. To this end, consider the events that the sample moments from the true DGP process happen to fall into their confidence intervals. If it were assumed that these events are all independent of each other, the true DGP would have a joint MCR of $0.95^9 = 63.0\%$. If more conservatively the fact of a certain dependence among the events is approximated by the assumption that only five of them are independent—say: r AC–1, Hill, v Mean, v AC–1 and v AC–100—then a joint MCR of 77.4% would be obtained. Using the latter as a conservative benchmark to which the fitting ratios from the Monte Carlo experiments may be related, the 10.1% of the DCA model with herding, predisposition and misalignment (HPM) corresponds to an effective MCR of $10.1/77.4 = 13.0\%$, and the 22.1% of the analogous TPA model to an effective MCR of 28.6%.

Another problem is that all of these statements are based on asymptotic theory. The confidence intervals of the moments that we thereby obtained may or may not be a good approximation to the ones appropriate for small samples of the real-world DGP. Above, to get more information about the small sample properties, we applied the bootstrap to

the empirical returns. We can once more make use of the 5000 bootstrap samples for our present purpose, that is, we compute the joint MCR for them and employ it as a yardstick against which we can measure the MCRs of our models. The result, as shown in the first column of the numbers in Table 4, is a ratio of 32.6%, which is considerably lower than the hypothetical reference figures of 77.4% or 63.0% that have just been mentioned. On the other hand, as it should be, the ratio of 32.6% is still distinctly higher than the MCRs of the models, though it does not degrade the latter completely.

The most obvious way to relate the models' joint MCRs to that of the bootstrap is to express them as a fraction of it (in %). This is done in the second row of Table 4. Referring to these statistics, we only would reject TPA–W; already the figure of 15.5% for DCA–W as the second-worst model appears fairly acceptable, even though the remaining five models perform much better. The best model, TPA–HPM, now reaches a level of even 67.9%. Nonetheless, this figure is probably an overestimation of a more appropriate relative coverage statistic and should not be taken too literally. In any case, whatever measure we refer to, the degree of the model's ability to match the moments we have chosen is highly remarkable.

Returning to the model ranking, TPA–HPM clearly outperforms DCA–HPM, which was formerly so dominating. To see what mainly causes the reversal in the ranking, we should have a look at the coverage ratios of the single moments in the lower part of Table 4 and compare them across the models. It is thus found that there are only two moments in which DCA–HPM is inferior to TPA–HPM, namely the Hill moment and the mean of the absolute returns. In all other moments, DCA–HPM has an edge over TPA–HPM, especially at the longer lags in ACF(v). Its lower joint MCR results from the fact that the superiority in the latter is less pronounced than the inferiority in v Mean and, most dramatically, in the Hill moment.

The overall impression from the detailed model comparisons is the omnipresence of trade-off effects. No model is, so to speak, Pareto inferior or superior to another model (except perhaps the WP and WHP versions of DCA, if we neglect the first moment which is almost perfectly matched by all models). In particular, the second-worst model DCA–W has the highest MCRs of all models at the longer lags of ACF(v). Unfortunately, this positive result is wrecked by the bad match of v Mean and Hill. The more general DCA–WHP is clearly better in this respect (even the best of all models), but at the price of a serious deterioration of the last three moments. There are similar, though more moderate examples for other pairwise model comparisons. On the whole, with the originally estimated parameters, TPA–HPM happens to have found the best compromise in the trade-offs.

23

4.4. *The joint coverage ratio as an alternative estimation criterion*

Table 4 has disproved the expectation that minimal values of the objective function J also imply a near-optimal ratio of a model's joint moment matching. In particular, the formerly best model DCA–HPM only ranks fourth when evaluated in terms of the joint MCR. Hence the question arises if a different choice of its structural parameters could enhance its position; or, with respect to the presently best model TPA–HPM, if we could still improve its joint MCR of 22.1%. Most obviously, we are thus asking for the maximization of the joint MCR as an alternative estimation criterion. Unfortunately, a direct maximization would incur an extremely high computational cost. In addition, the many trade-offs that we have seen are likely to cause a larger (or even infinite) number of local maxima of the joint coverage ratio.

To reduce the computational burden we resort to a heuristic procedure. Our idea is to modify the original loss function J and set up a version of which we hope that it is better suited for achieving high coverage ratios. To begin with, we disregard the cross-dependencies between the single moments in the weighting matrix. Since our goal is to have the moments contained in their confidence intervals, the width of which are given by roughly two times the standard error (SE_i) of the moments i, we think of a diagonal weighting matrix with weights $w_{ii} = 1/(\mathrm{SE}_i)^2$.

With such an objective function, a deviation of the simulated from the empirical moment i by one or two SE_i, respectively, would yield a contribution of this moment to the total loss of the function of 1 or 4, respectively. However, in our explorations of this idea we found out that penalizing the deviations in a quadratic way is not sufficiently strong.

As a more flexible approach, the following specification proved useful then. Starting from the observation that the expressions $d_i := |\, m_i(\theta; T, s) - m_i^{emp} \,| \,/\, SE_i$ should fall short of 1.96 if possible (s being an index to represent the random seed in the simulation of the model), we expect better chances if values of d_i close to or above this threshold are disproportionately penalized. The simplest way to do so is to apply a piecewise linear transformation function to d_i, where the change in the slope is described by two parameters $\widetilde{d_i}$ and κ_i. The slope for values of d_i between zero and some critical value $\widetilde{d_i}$ somewhat below 1.96 is thus relatively flat compared to the slope when $d_i > \widetilde{d_i}$. Without loss of generality, the slope over the first segment can be unity, and it is $\kappa_i \gg 1$ when $d_i > \widetilde{d_i}$. After some trial and error, we uniformly put $\widetilde{d_i} = \widetilde{d} = 1.90$, so that our transformation function reads,

$$
F(d_i; \kappa_i) \;=\; \begin{cases} d_i & \text{if } 0 \le d_i \le \widetilde{d} \\ \widetilde{d} + \kappa_i(d_i - \widetilde{d}) & \text{if } d_i \ge \widetilde{d} \end{cases} \qquad \widetilde{d} = 1.90 \tag{10}
$$

Given the moment-specific slopes κ_i, and given the index s of the random seed of a simulation run over the empirical time horizon T $(s = 1, 2, \dots)$, the total loss to which

24

a parameter set θ gives rise is the sum of the losses of the single moments in (10),

$$\ell(\theta; T, s) \quad := \quad \sum_{i=1}^{9} F\left[\,|\,m_i(\theta; T, s) - m_i^{emp}\,|\,/\,\mathrm{SE}_i\,;\,\kappa_i\,\right] \tag{11}$$

To reduce the sample variability, these losses are finally averaged over a larger number S of simulations. Dropping the explicit reference to T, this yields the loss function $L = L(\theta; S)$ with which we will work in the following:

$$L(\theta; S) \quad := \quad \frac{1}{S} \sum_{s=1}^{S} \ell(\theta; T, s) \tag{12}$$

We first applied the function to model TPA–HPM and explored whether it would thus be possible to raise its joint MCR noticeably above the maximal 22.1% from Table 4. We began with $S = 10$ and later sharpened it to $S = 100$. Playing around with different combinations of the slope parameters κ_i, the trade-off effects were also (not surprisingly) found for the individual model. That is, an increase in the coverage ratios of one or two moments goes at the expense of some other moments. Generally, of course, the implications for the joint coverage ratio are ambiguous. It nevertheless proved promising to try to improve on the coverage ratio of the Hill estimator. Although this tends to deteriorate the ratios of the autocorrelations of the absolute returns with the longer lags, sufficiently high κ_i for the latter can restrain these negative effects such that, on the whole, the joint coverage ratio increases. These trade-offs are illustrated for four different parameter sets in Table 6. For better control in the explorations, we here limit ourselves to the variations of four parameters, which are reported in Table 5 (this limitation will, of course, be dropped in the final optimization).

Set	χ	α_p	σ^f	σ^c
A	2.300	12.500	0.790	1.900
B	2.176	11.981	0.786	1.912
C	2.160	11.768	0.794	2.044
D	2.323	12.148	0.779	1.937

Table 5: Alternative parameter sets for TPA–HPM
(other parameters as in Table 1).

The first parameter set is the one from our original MSM estimation of TPA–HPM; see Table 1. Accordingly, column A in Table 6 reproduces the statistics in the last column of Table 4. As observed before, the Hill estimator (the second moment) has the worst fit. The optimization under a higher tuning parameter κ_2 for it leads to parameter set B and

procures a moderate improvement in this respect, at the price of a moderate worsening of the long-lag autocorrelations. The effects brought about by set C (from somewhat different values of the κ_i) go in the same direction, but the quantitative changes are larger. In both cases, the overall impact of these trade-offs on the joint coverage ratio is positive, being mild in the first and stronger in the second case.

	A	B	C	D
joint MCR	22.1	22.3	22.9	23.7
1: r AC–1	97.1	97.2	96.9	97.0
2: Hill	53.3	58.9	76.1	67.2
3: v Mean	64.9	63.2	60.1	62.6
4: v AC–1	92.0	93.6	95.0	94.3
5: v AC–5	68.2	73.3	79.5	77.0
6: v AC–10	83.1	84.5	77.2	82.2
7: v AC–25	65.2	59.3	48.5	53.6
8: v AC–50	73.5	69.1	60.0	64.5
9: v AC–100	66.5	64.7	59.5	62.4

Table 6: Moment coverage ratios of TPA–HPM
from the alternative parameter sets of Table 5.

Regarding parameter set C one may nevertheless feel that the ratios of the long-lag autocorrelations are already unduly low. This induced us to try higher penalties for these moments. They worked out as expected, the primary trade-off again affecting the coverage of the Hill moment. On the whole, however, a further increase of the joint MCR up to 23.6% could be achieved. Column D in the table and its comparison with the other columns shows the kind of compromise that makes this success possible. The underlying moment-specific penalty parameters are as follows (with respect to the moment numbering in Table 6),

$$\kappa_i = \begin{cases} 10 & i = 1,4,5,6 \\ 50 & i = 2,3 \\ 100 & i = 7,8,9 \end{cases} \tag{13}$$

These are also the values that we adopt for our final optimization, together with $S = 100$ for the number of sample runs in (12). Having found a minimum in this way, we subsequently rounded the parameter values, ran the 5000 MC simulations, and computed the

coverage ratios. We also made further explorations with slight modifications of the parameters in order to check that we have not missed a higher joint MCR in this parameter region.

	ϕ	χ	η	α_w	α_o	α_n	α_p	σ^f	σ^c
TPA									
HPM:	0.180	2.255	—	—	−0.165	1.291	12.924	0.782	1.927
DCA									
WP:	0.945	0.998	0.987	3018	1.469	—	—	0.690	1.759
WHP:	0.935	1.148	0.987	3018	1.506	0.431	—	0.690	1.759
HPM:	0.120	1.500	—	—	−0.336	1.839	19.671	0.708	2.147

Table 7: Parameters to maximize the joint coverage ratio.

Note: Common to all models is the normalization $\mu = 0.01$. In addition, $\beta = 1$ is common to the DCA versions and $\nu = 0.05$ is underlying TPA–HPM.

The presentation of our results is limited to the four model versions with the best performance. This is the herding version without differential wealth for TPA as well as DCA, and for the latter approach the two more elaborated versions with differential wealth. The parameter values to which we commit ourselves in the end are collected in Table 7. It is seen that the values for TPA–HPM are not very different from the original values in Table 1. Common to the three DCA models is the opposite shift of the noise levels in the fundamentalist and chartist demand, which are lower for the former and higher for the latter group. In the two wealth versions DCA–WP and DCA–WHP we notice similar modifications of the fundamentalist and chartist reaction coefficients ϕ and χ. It is also interesting that in DCA–WHP the herding effect is now considerably weakened (a strong decrease in α_n).

The moment coverage ratios to which the new parameters give rise are reported in Table 8. In comparing them to Table 4, two results are most noteworthy. First, we see that slight changes in the parameters of TPA–HPM improve the joint MCR by more than two percentage points. Second, the same variant under DCA can do much better than in Table 4. Indeed, DCA–HPM overtakes TPA–HPM and achieves a top joint MCR of 26.6%. This is more than 80 per cent of the MCR of the bootstrap samples, which already served us as a yardstick above. Notice that, in particular, the new parameters come off with an extraordinarily higher MCR of the Hill moment (71.6% versus the

	TPA–HPM	DCA–WP	DCA–WHP	DCA–HPM
'p-value'	**24.0**	10.2	8.7	19.5
joint MCR	24.4	23.8	24.1	**26.6**
in % of 32.6	74.8	73.0	73.9	**81.6**
1: r AC–1	97.0	97.4	96.9	97.5
2: Hill	62.9	86,4	86,7	71.6
3: v Mean	64.7	66.8	65.9	55.8
4: v AC–1	94.7	100.0	100.0	99.5
5: v AC–5	75.9	95.8	96.4	93.5
6: v AC–10	84.6	89.8	88.5	90.0
7: v AC–25	59.0	79.8	79.7	79.4
8: v AC–50	70.2	67.8	69.7	84.4
9: v AC–100	65.2	49.5	51.1	70.5

Table 8: Moment coverage ratios (in %)
obtained from minimizing the loss function $(10) - (12)$.

Note: The figures in the first row are the moment-specific p-values with respect to the bootstrap distribution of J of the empirical returns, described in Section 4.1. The figure of 32.6% in the third row is the joint MCR of the bootstrap samples (see Table 4). Boldface figures emphasize the good performance of a model.

previous 23.4%). Our heuristic search procedure to optimize on the coverage ratios proved therefore remarkably successful.

The first row in Table 8 takes up the concept of the moment-spcific p-value of a model with respect to the bootstrap distribution of J of the empirical returns, which was discussed in Section 4.1. It shows that this criterion when applied to the same parameter sets would reverse the ranking, as here TPA–HPM performs best and DCA–HPM comes out second. Despite its higher joint MCR, model TPA–HPM has even the same p-value as in Table 2. The different implications of the two criteria of the p-value and MCR are most strikingly seen in the comparison of TPA–HPM to DCA–WHP; while their joint MCRs are about equal, DCA–WHP has by far the lowest of all p-values reported in this paper.[24]

---

[24] That the joint MCR of DCA–WHP coincides with the p-value of this model in Table 2 is, of course, purely accidental.

To sum up the results for the maximization of the joint MCR criterion and proclaim what we consider to be the final winner of our model contest: it is the discrete choice approach combined with the mechanisms of herding, a predisposition towards chartism, and a correction in the agents' demand for price misalignment. DCA–HPM has a clear, though not dramatic, edge over the same mechanisms within the framework of the transition probability approach, and over the DCA versions that incorporate the effects of differential wealth. The performance of the last three models is about equal. More generally, the moment coverage ratios in Table 8 provide a striking benchmark against which we would like to measure other models of speculative asset price dynamics.

5. Conclusion

In recent years, agent-based asset pricing models have made remarkable progress in replicating a diverse set of central statistical features of financial market time series. In this endeavour, the present paper has focussed on a new class of models which is based on the concept of structural stochastic volatility. This framework lets the market participants choose between a noisy technical and a noisy fundamental trading strategy, where generally the choice is governed by four socio-economic principles. In short, they include differential wealth, herding, a behavioral predisposition towards one of the strategies, and a misalignment correction mechanism, i.e. a propensity to withdraw from chartism as the gap between prices and the fundamental value widens. A switching index into which a combination of these components is aggregated is then the driving variable in, alternatively, the discrete choice approach or the transition probability approach, which are two convenient devices from the literature to determine the market shares of the two groups of traders. Simulations show that the resulting variations in the overall noise level and in the stabilizing and destabilizing tendencies of the strategies can lead to irregular switches between rather volatile and rather calm market regimes. More specifically, phenomena such as volatility clustering, long memory effects and fat tails may thus come about.

Having thus a promising collection of different models at hand, we put them to more formal tests. To this end, we employed the method of simulated moments. This is a powerful estimation procedure that is easy to apply to any agent-based asset pricing model (not only the model family considered in the present study). It is also a very transparent method as it requires the researcher to make up his or her mind about the stylized facts that a model should be able to reproduce, and to set up the precise summary statistics (the 'moments') by which he or she wants to quantify them.

We actually made use of two variants that seek to bring the model-implied summary statistics as close as possible to their empirical counterparts. In the first case, 'closeness' is represented by the distance between these moments as it is measured by a quadratic loss function. For the second case we introduced the concept of a (joint) moment coverage

ratio (MCR). Here the models are repeatedly simulated over the empirical time horizon and we count the number of times where all of the simulated moments are contained in the empirical confidence intervals. MCR is therefore another, apparently new way of summarizing how often the data from a model and the real market could not be told apart.

Accepting our (or a similar) choice of moments, both approaches provide us with a transparent and not too technical criterion to assess the overall fit of a model. This is indeed something that the burgeoning field of agent-based financial market models is currently in need of. So many models have been put forward for which we are now facing the "wilderness of bounded rationality problem" (Hommes, 2011, p. 2), meaning that there are too many degrees of freedom to set up good models. At this stage some guidance is required to judge which models can mimic the stylized facts, say, "fairly well" and which are even "very good" in this way. Thus, to sum up, a basic methodological message of our paper is that the method of simulated moments, with the specifications here proposed, is a superb tool to serve this purpose.

The results from applying this method to our family of models suggests a general call for model contest. On the one hand, we have identified mechanisms that, despite their elementary design, are already quite efficient in matching the desired moments. Specifically, regarding the ranking of the models, the clear winner of the model contest in this paper is DCA–HPM: the discrete choice approach that incorporates herding with a certain predisposition towards chartism, which is however tamed by the misalignment correction mechanism.

On the other hand, concerning the quantitative performance of this model version, it may be recalled that we bootstrapped a (moment-specific) p-value of 32.6% for it, and that according to the MCR criterion no less than 26.6 per cent of the simulations are indistinguishable from the empirical data. Our view is that these statistics set a remarkably high standard for model validation and model selection, and we hope that this benchmark may be accepted as a challenge to existing models and to future model building. [25]

Appendix A1: Standard errors of the empirical moments

Let us begin with the Hill moment, which we base on a conventional tail size of five per cent. Arranging the absolute returns in descending order, $v_{(1)} \geq v_{(2)} \geq \ldots \geq v_{(T)}$, and putting $k = 0.05 \cdot T$ (rounded to an integer number), the Hill estimator is equal to the reciprocal of

---

[25] This is not to say that goodness-of-fit along these or similar lines is the only criterion to evaluate competing models. In future research it should, however, be clearly worked out what other features might compensate for a possible inferiority in this respect.

$$\hat{\gamma} \ := \ (1/k) \sum_{i=1}^{k} \ln v_{(i)} - \ln v_{(k)}$$

The fact that this estimator is necessarily biased need not concern us here since it applies to the empirical and model-generated data alike. Its asymptotic normality has been shown by, e.g., Goldie and Smith (1987) or Hall (1990), for which they obtain the variance

$$\mathrm{Var}(\hat{\gamma}) \ = \ \gamma^2 / k$$

Certainly, in applications the 'true' value γ is replaced with the estimate $\hat{\gamma}$. The moment coverage ratios relate to $\hat{\gamma}$ rather than the Hill estimator $1/\hat{\gamma}$ itself.

The computation of the standard errors of the other moments makes use of the delta method. We first give the general argument and then apply it for our special purpose. To this end, we refer to $\{z_t\}_{t=1}^{T}$ as a time series of a state vector z_t of arbitrary dimension, $z_t \in \mathbb{R}^n$. In particular, to deal with the autocovariances the components of z_t may be v_t, v_{t-1}, etc. Let there be k 'raw' moments, a suitable combination of which yield our (composed) moments of the autocovariances or autocorrelations. Furthermore, let $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ be a continuous function, and $\widehat{\mu} \in \mathbb{R}^k$ the corresponding vector of the estimated (unconditional) raw moments,

$$\widehat{\mu} \ = \ \frac{1}{T} \sum_{t=1}^{T} f(z_t) \tag{A14}$$

Since (A14) is a consistent and asymptotically normal estimator of $\mu = E[f(x_t)]$, we have with respect to its covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$ and the true raw moments μ^o,

$$\sqrt{T}\,(\widehat{\mu} - \mu^o) \ \overset{a}{\sim} \ N(0, \Sigma)$$

An HAC estimator of Σ is the following Newey-West matrix for some suitable lag length p,

$$\widehat{\Sigma} \ = \ \Gamma(0) \ + \ \sum_{h=1}^{p} (1 - \frac{h}{p+1}) \, [\widehat{\Gamma}(h) + \widehat{\Gamma}(h)']$$

$$\widehat{\Gamma}(h) \ = \ \frac{1}{T} \sum_{t=h+1}^{T} [f(z_t) - \widehat{\mu}] \, [f(z_{t-h}) - \widehat{\mu}]' \,, \qquad h = 0, 1, \dots, p$$

We follow a usual practice and set p equal to the smallest integer greater than or equal to $T^{1/4}$ (Greene, 2002, p. 267, fn10).

Next, let $\theta \in \mathbb{R}$ be a summary statistic of z_t which is determined by the raw moments μ *via* a continuously differentiable and monotonic real function $g : \mathbb{R}^k \rightarrow \mathbb{R}$. That is, θ is estimated as

$$\widehat{\theta} \ = \ g(\widehat{\mu})$$

Putting $\theta^o = g(\mu^o)$ and employing the delta method (*cf.* Davidson and MacKinnon, 2004, pp. 207f), we know that

$$\sqrt{T}\left(\widehat{\theta} - \theta^o\right) \stackrel{a}{\sim} N(0, \sigma^2)$$

and that the asymptotic variance σ^2 can be estimated as

$$\widehat{\sigma}^2 = \frac{\partial g(\widehat{\mu})}{\partial \mu'} \; \widehat{\Sigma} \; \frac{\partial g(\widehat{\mu})}{\partial \mu} \tag{A15}$$

To obtain the standard error of the estimates of θ, it remains to divide $\widehat{\sigma}$ by $\sqrt{T}$.

We can then turn to the application of (A15) to the correlation coefficient ρ between $\{x_t\}$ and $\{y_t\}$. We only have to set

$$\mu = (\mu_x, \, \mu_y, \, \mu_{xx}, \, \mu_{yy}, \, \mu_{xy})' = [E(x_t), \, E(y_t), \, E(x_t^2), \, E(y_t^2), \, E(x_t y_t)]'$$

and express ρ as

$$\rho = g(\mu) = \frac{\mu_{xy} - \mu_x \, \mu_y}{\sqrt{\mu_{xx} - \mu_x^2} \, \sqrt{\mu_{yy} - \mu_y^2}}$$

The specification of the corresponding functions $f_i = f_i(x_t, y_t)$, $i = 1, \ldots, 5$, is obvious ($f_1(x_t, y_t) = x_t$, etc.) and with $\sigma_x = \sqrt{\mu_{xx} - \mu_x^2}$, $\sigma_y = \sqrt{\mu_{yy} - \mu_y^2}$, the vector of the partial derivatives of g is computed as

$$\frac{\partial g}{\partial \mu'} = \left(\frac{\rho \, \mu_x}{\sigma_x^2} - \frac{\mu_y}{\sigma_x \sigma_y}, \;\; \frac{\rho \, \mu_y}{\sigma_y^2} - \frac{\mu_x}{\sigma_x \sigma_y}, \;\; \frac{-\rho}{2\sigma_x^2}, \;\; \frac{-\rho}{2\sigma_y^2}, \;\; \frac{1}{\sigma_x \sigma_y} \right)$$

The lower and upper boundaries of the confidence intervals around the empirical summary statistics that we thus obtain are documented in Table A1 (the autocorrelations of the absolute returns being smoothed as described in footnote 9).

	rAC–1	1/Hill	vMean	vAC 1	vAC 5	vAC 10	vAC 25	vAC 50	vAC 100
measured :	−0.008	0.301	0.713	0.193	0.187	0.159	0.128	0.112	0.074
lower bound :	−0.042	0.269	0.672	0.106	0.142	0.118	0.095	0.074	0.041
upper bound :	0.027	0.334	0.754	0.280	0.231	0.200	0.162	0.148	0.105

Table A1: The empirical moments and their 95% confidence intervals.

Appendix A2: Bootstrapping the covariance matrix of the moments

When employing a bootstrap procedure to obtain an estimate $\widehat{\Sigma}$ of the covariance matrix of the moments, a block bootstrap is appropriate due to the long-range dependence in the return series. An immediate idea for choosing the block length would be one year, i.e. 250 days. Accordingly, we reduce the empirical return series of S&P 500 (which

comprises 6866 data points) to $T' = 6750$ observations, subdivide it into 27 fixed and non-overlapping blocks of 250 days, and construct a new series, block by block, from 27 random draws (with replacement). All of the desired moments can then be computed from this string of the blocks of returns. [26]

Repeating this procedure 5000 times, a frequency distribution for each of the moments is obtained. Ideally, the empirical moment should be more or less in the centre of the distribution. This is also essentially what we get, except for the ACF of the absolute returns at the longer lags $\tau \geq 10$, for which the bootstrapped coefficients show a non-negligible tendency towards lower values. Apparently, breaking up the dependence in the returns every 250 days is here too early. To mitigate the bias, we tried longer blocks of 500 and 750 days, which led to a noticeable (though still not yet perfect) improvement. [27] On the basis of these explorations we chose a block length of 250 days for the first five "short memory" moments (first-order autocorrelation of the raw returns and with respect to the absolute returns, the Hill estimator, mean volatility and the autocorrelations at lags $\tau = 1$ and $\tau = 5$), and a block length of 750 days for the last four "long memory" moments (the ACF of $v_t = |r_t|$ at lags $\tau = 10, 25, 50, 100$). [28]

It is the frequency distributions of these bootstrapped moments from which the desired estimate $\widehat{\Sigma}$ of the covariance matrix derives. Formally, let $b = 1, \ldots, B$ identify the bootstrap samples ($B = 5000$ presently), let $m^b = (m_1^b, \ldots, m_9^b)'$ 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

$$\widehat{\Sigma} \quad := \quad \frac{1}{B} \sum_{b=1}^{B} (m^b - \overline{m})(m^b - \overline{m})' \tag{A16}$$

To sum up, the estimations of the models in Section 3 are based on the minimization in eq. (3), where the objective function J entering there is defined by eqs (2), (4), (A16), and of course the set of the nine moments described in the text.

The minimizations themselves are carried out by means of the Nelder-Mead simplex algorithm (see Press et al., 1986, pp. 289–293). To avoid getting prematurely trapped in a local minimum, the procedure is several times restarted with a relatively large new initial simplex which hopefully reaches beyond the "hills" surrounding a local "valley", until no further noteworthy improvement in the minimization occurs. If in the course of the search some parameter 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. As

---

[26] Available evidence indicates that the increase in efficiency that one may hope to gain from the resampling of overlapping blocks is small, if it exists at all (Hall et al., 1995).

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

[28] Diagrams of a couple of these bootstrap distributions can be found in Franke and Westerhoff (2011, Section 5).

there are no corner solutions, the parameters are then driven back into the interior of the admissible set if the proportionality factor is high enough (like one thousand, say). Taking all this into account, we are fairly confident for each of the estimations that, given the random number seed, the search algorithm has led us sufficiently close to a global minimum.

References

AMILON, H. (2008): Estimation of an adaptive stock market model with heterogeneous agents. *Journal of Empirical Finance*, 15, 342–362.

ALFARANO, S., LUX, T. AND WAGNER, F. (2005), "Estimation of agent-based models: the case of an asymmetric herding model", *Computational Economics*, 26, 19–49.

BOSWIJK, P., HOMMES, C. AND MANZAN, S. (2007): Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31, 1938–1970.

BROCK, W.A. AND DE FONTNOUVELLE, P. (2000): Expectational diversity in monetary economics. *Journal of Economic Dynamics and Control*, 24, 725–759.

BROCK, W. AND HOMMES, C. (1997): A rational route to randomness. *Econometrica*, 65, 1059–1095.

CAEIRO, F., GOMES, M.I., AND PASTANA, D. (2005): Direct reduction of bias of the classical Hill estimator. *REVSTAT—Statistical Journal*, 3, 113–136.

CHAN, W.H. AND MAHEU, J.M (2002): Conditional jump dynamics in stock market returns. *Journal of Business and Economic Statistics*, 20, 377–389.

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`).

CHIARELLA, C., DIECI, R. AND HE, X.-Z. (2009): Heterogeneity, market mechanisms, and asset price dynamics. In: Hens, T. and Schenk-Hoppé, K.R. (eds): *Handbook of Financial Markets: Dynamics and Evolution*. North-Holland: Amsterdam; pp. 277–344.

CHIARELLA, C., HE, X.-Z., HUANG, W. AND ZHENG, H. (2011): Estimating behavioural heterogeneity under regime switching. Working paper, School of Finance and Economics, University of Technology, Sydney.

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

ERAKER, B. (2004): Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59, 1367–1403.

FARMER, J.D. AND JOSHI, S. (2002): The price dynamics of common trading strategies. *Journal of Economic Behavior and Organization*, 49, 149–171.

FRANKE, R. (2008): A microfounded herding model and its estimation on German survey expectations. *Intervention: European Journal of Economics and Economic Policies*, 5, 305–332.

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).

FRANKE, R. AND WESTERHOFF, F. (2011): Estimation of a structural stochastic volatility model of asset pricing. *Computational Economics* (in press); DOI 10.1007/s10614-010-9238-7.

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

GOLDIE, C.M. AND SMITH, R.L. (1987): Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics, Oxford 2nd Series*, 38, 45–71.

GREENE, W.H. (2002), *Econometric Analysis*: Prentice-Hall: Upper Saddle River, N.J. (5th edn).

HALL, P. (1990): Using the bootstrap to estimate mean square error and select smoothing parameters in non-parametric problems. *Journal of Multivariate Analysis*, 32, 177–203.

HALL, P., HOROWITZ, J.L. AND JING, B.-Y. (1995): On the blocking rules for the bootstrap with dependent data. *Biometrika*, 82, 561–574.

HE, X.-Z. AND LI, Y. (2007): Power-law behaviour, heterogeneity, and trend chasing. *Journal of Economic Dynamics and Control*, 31, 3396–3426.

HEVIA, C. (2008): Standard errors using the delta method and GMM. *mimeo*.

HOMMES, C. (2006): Heterogeneous agent models in economics and finance. In: Tesfatsion, L. and Judd, K. (eds): *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*. North-Holland: Amsterdam; pp. 1107–1186.

HOMMES, C. (2011): The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, 35, 1–24.

HUISMAN, R., KOEDIJK, K., KOOL, C. AND PALM, F. (2002): The tail-fatness of FX returns reconsidered. *De Economist*, 150, 299–312.

KINDLEBERGER, C.P. (2000): *Manias, Panics, and Crashes: A History of Financial Crisis*. New York: John Wiley & Sons (4th edn).

KOCHERLAKOTA, N.R. (1990): On tests of representative consumer asset pricing models. *Journal of Monetary Economics*, 26, 285–304.

LEBARON, B. (2006): Agent-based computational finance. In: Tesfatsion, L. and Judd, K. (eds): *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*. North-Holland: Amsterdam; pp. 1187–1233.

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

LI, Y., DONKERS, B. AND MELENBERG, B. (2010): Econometric analysis of microscopic simulation models. *Quantitative Finance*, 10, 1187–1201.

LUX, T. (1995): Herd behaviour, bubbles and crashes. *Economic Journal*, 105, 881–889.

LUX, T. (2009): 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. (2007), "Heterogeneous expectations, exchange rate dynamics and predictability", *Journal of Economic Behavior and Organization*, 64, 111–128.

Mao, C.S. (1990): Hypothesis testing and finite sample properties of generalized method of moments estimators: A Monte Carlo study. Federal Reserve Bank of Richmond, Working Paper 90–12.

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

Shiller, R. (2005): *Irrational Exuberance*. Princeton University Press: Princeton (2nd edn).

Weidlich, W. and Haag, G. (1983): *Concepts and Models of a Quantitative Sociology: The Dynamics of Interacting Populations*. Springer: Berlin.

Westerhoff, F. (2009): Exchange rate dynamics: A nonlinear survey. In: Rosser, J.B., Jr. (ed.): *Handbook of Research on Complexity*. Edward Elgar: Cheltenham; pp. 287–325.

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.