Applying the Method of Simulated Moments
to Estimate a Small Agent-Based Asset Pricing Model:
Extended Version

Reiner Franke\textsuperscript{a,}\textsuperscript{*}

November 2008
\textsuperscript{a}University of Kiel, Germany

Abstract
The paper takes a recent agent-based asset pricing model by Manzan and Westerhoff from the literature and applies the method of simulated moments to estimate its six parameters. In selecting the moments, the focus is on the fat tails and autocorrelation patterns of the daily returns of several stock market indices and foreign exchange rates. We argue for abandoning the econometrically optimal weighting matrix in the objective function and instead invoke the moments’ \( t \)-statistics in an intuitively appealing way. This modification gives rise to estimations whose moment matching can be largely considered to be satisfactory. Also the parameter estimates across the different markets make good economic sense.

\textit{JEL classification:} C14; C15; F31.

\textit{Keywords:} Simulation-based estimation; daily returns; autocorrelation patterns; fundamentalists and speculators.

1. Introduction

Behavioural asset pricing models often include latent variables in a nonlinear framework. An estimation by maximum likelihood or the generalized method of moments becomes

\textsuperscript{*} Corresponding author.

Email address: franke@iksf.uni-bremen.de (Reiner Franke).

\textsuperscript{1} I wish to thank Thomas Lux and Frank Westerhoff for their helpful comments. Financial support from EU STREP ComplexMarkets, contract number 516446, is gratefully acknowledged.
infeasible then since the likelihood function and the moment restrictions do not have analytically tractable forms in terms of the unknown parameters. Instead of suitable approximations of these models that might simplify the objective functions, there are now several simulation-based methods available that can serve as an alternative approach. The present paper is concerned with the method of simulated moments (MSM) to test a small model of speculative dynamics from the literature. In brief, this approach seeks to identify numerical parameter values such that several summary statistics of interest, or ‘moments’, that are computed from simulations of the model come close to their empirical counterparts.

MSM is frequently criticized because of its arbitrary choice of moments in the criterion function. For this reason methods have been developed, such as indirect inference or the efficient method of moments, where in effect a set of moment conditions is endogenously defined by means of an auxiliary model. It may, however, be argued that this does not completely resolve the problem of arbitrariness, or judgement; it is only shifted from the choice of moments to the choice of the auxiliary model. Taking into account that structural models are necessarily misspecified, one may furthermore prefer MSM to the other methods because of its higher transparency. After all, a model is not meant to resemble or mimic any actual economy or financial market, but the best we can expect from it is that it matches some of the ‘stylized facts’. While MSM requires the researcher to make up his or her mind about the dimensions along which the model should be most realistic, the implications of a decision for a particular auxiliary model are usually less clear. Besides the fact that in certain cases MSM is possibly faster or more robust than other methods, we consider its explicitness and easy interpretation as the main argument to try this estimation procedure.

Regarding the estimation of agent-based asset pricing models, it seems that MSM has so far only been applied by a research group around M. Gilli and P. Winker. The latest and most elaborate paper is Winker et al. (2007). In their work the authors are, however, mainly taken up by their objective function and do not endeavour to make use of the existing econometric theory on MSM, from which one seeks to get information on the precision of the parameter estimates and the goodness-of-fit of the selected moments in

\footnote{Opinions seem to be divided over calling this method the ‘method of simulated moments’ or the ‘simulated method of moments’ (at least if there is no conflict with the acronym of Markov switching models).}

\footnote{Carrasco and Florens (2002) provide a succinct overview of MSM and the other two methods just mentioned.}

\footnote{In the sense that we prefer to try MSM first, which does by no means rule out the subsequent application of other methods. For a recent study of the estimation properties of MSM and other simulation-based methods when applied to an elementary macroeconomic Real Business Cycle model, see Ruge-Murcia (2007).}

\footnote{For a recent overview of the application of this and other estimation methods to asset pricing models see Chen et al. (2008; especially Table 3 on p. 22).}
the objective function (and possibly of other moments that were not included). It is one intention of the present paper to discuss these points in detail.

Another aspect is that the structural models to which Winker et al. apply MSM are as yet not well suited to reproduce the stylized facts of the daily returns on the financial markets (at least not with the authors’ exogenous setting of the other model parameters that were not estimated). Hence the major economic outcome of the analysis is a rejection of the models, which is so incontrovertible that presumably the same conclusion could also be reached with less effort. In contrast, with Manzan and Westerhoff (2005) we choose an agent-based model from the literature that, although it is small and highly stylized, seems far more promising concerning the properties of its returns.\(^6\)

While it will be no great surprise that this model, too, is rejected by the standards of econometric theory, the mismatch of the moments that we are examining is not overly dramatic. We then abandon the econometrically optimal weighting matrix and invoke the \(t\)-statistics of the moments to modify the estimation’s objective function in an intuitively appealing way. In consequence, the new estimations yield a moment matching that can be largely considered to be satisfactory. We have thus a reasonable basis to judge the explanatory power of the model. Accepting the limitation that our moments only include returns but not the level of the time-varying fundamental value, it turns out that the role of the fundamentalists is minimal, unless dispensable. By contrast, the switching mechanism of the group of the so-called speculators is fully confirmed. Having this established, we subsequently estimate the model on several different stock market indices and foreign exchange rates. These results will give further support to the switching mechanism since we find that the corresponding parameter estimates are fairly robust and where there are greater differences, they make economic sense.

The remainder of the paper is organized as follows. The next section recapitulates the basic econometric facts of the method of simulated moments. Section 3 presents the Manzan–Westerhoff model, selects the moment functions, simulates the model, and then re-estimates it on the artificial data. The controlled experiment gives us some hints where there are problems and what at best could be expected from empirical estimations of this model. Section 4 turns to the real-world data of daily returns. Beginning with the Standard and Poor’s 500 stock market index as our reference series, we discuss the merits of three different minimization criteria for the estimations and justify our decision for one of them which, as just indicated, is based on the \(t\)-statistics of the moments. After a brief sketch of how we cope with the problem of multiple local minima of the objective functions, we present and compare the estimation results for a few additional

\[^6\] In another paper, Manzan and Westerhoff (2007) estimate their model on monthly foreign exchange rate data. There the authors can basically get along with OLS since they specify the fundamental value by invoking the Purchasing Power Parity. With the daily changes considered in their earlier paper, however, the fundamental value is a more volatile, and unobservable, variable.
stock market indices and foreign exchange rates, and also consider the general “goodness-of-fit” of the moments. Section 5 concludes.

2. The method of simulated moments

2.1. The objective function

In confronting the model with the empirical data, MSM focusses on a set of \( n_m \) descriptive statistics.\(^7\) To begin with the empirical side, let \( L \) be the greatest lag involved in forming these statistics, \( y_t^{emp} \) a vector of (stationary) empirical variables observed in period \( t \) (whose dimension does not matter), and suppose observations are available over a span of time \( t = 1-\ldots, 1, \ldots, T \). The summary statistics derive from \( n_m \) moment functions \( m_i(\cdot) \), which are generally defined on \( L \)-stretches of a variable \( y_t \). Writing more compactly \( z_t := (y_t, y_{t-1}, \ldots, y_{t-L}) \), the \( i \)-th empirical moment is computed as the time average

\[
\hat{m}_{T,i} := (1/T) \sum_{t=1}^{T} m_i(z_t^{emp}), \quad i = 1, \ldots, n_m
\]  

The index \( T \) makes the effective length of the sample period explicit. The caret over ‘\( m \)’ distinguishes the summary statistic from a single contribution to period \( t \) and also indicates that it is only an estimate of the true unconditional moment of the real-world stochastic process (as the actual sequence \( \{y_t^{emp}\}_{t=1-L}^T \) is just one realization of it).

Turning to the structural model under study, let \( \theta \) be the \( n_p \times 1 \) vector of the parameters that are to be estimated. Given numerical values for its components, an extended simulation of the model is run, such that after discarding a sufficiently long initial history to wash out transient effects we have model-generated observations over a horizon of \( sT \) periods, where \( s \) is some number greater than one. Possibly besides several other variables that one may not be interested in or that may not be observable in the real world, for each set of parameters \( \theta \) a unique series \( \{y_t(\theta)\}_{t=1-L}^{sT} \) is produced in this way.

Since all simulations are required to employ the same sequence of pseudo-random numbers for the standardized stochastic terms in the model, the differences between two series \( \{y_t(\theta^1)\} \) and \( \{y_t(\theta^2)\} \) can be attributed to the differences in the parameters \( \theta^1 \) and \( \theta^2 \), and not to the (standardized) random noise from the stochastic part of the model.\(^8\) As the artificial data \( y_t(\theta) \), or \( z_t(\theta) \) for that matter, are the counterparts of the empirical data \( y_t^{emp} \) and \( z_t^{emp} \), it makes sense to compute the \( n_m \times 1 \) vector of the simulated moments,

\(^7\) The following exposition is based on Lee and Ingram (1991) and Duffie and Singleton (1993), which are two standard references for MSM.

\(^8\) If, for example, a term \( \varepsilon_t \) in the model is normally distributed as \( N(\mu, \sigma^2) \) and \( \mu \) and/or \( \sigma \) are among the parameters to be estimated, the expression ‘standardized’ means that for each simulation run at time \( t \) the same random number \( \tilde{\varepsilon}_t \) is drawn from the standard normal distribution \( N(0, 1) \), and \( \varepsilon_t \) is set as \( \varepsilon_t = \mu + \sigma \tilde{\varepsilon}_t \).
\[ \hat{\mu}_{sT}(\theta) := (1/sT)\sum_{t=1}^{sT} m[z_t(\theta)] \]  

(2)

Here the caret over ‘\( \mu \)’ indicates that (2) is an estimate of the expected value \( \mu(\theta) := E_\theta[m(z)] \), which are the model’s unconditional moments with respect to \( \theta \). In most structural models, especially if they contain some nonlinearities, the latter will be analytically untractable, so that one has to resort to the estimates \( \hat{\mu}_{sT} \).

Ideally, by a suitable choice of \( \theta \), the expected moments of the model come close to the moments of the stochastic process that rules the real world. As both sets of moments are unknown, a comparison has to replace them with the sample counterparts \( \hat{\mu}_{sT}(\theta) \) and \( \hat{m}_T \), respectively. Accordingly, a vector \( \theta \) has to be found that minimizes the distance between the two descriptive statistics. To formalize this idea, let \( \hat{m}_T \) be the \( n_m \times 1 \) vector of the empirical moments \( \hat{m}_{T,i} \) in (1) and define

\[ g_{s,T}(\theta) := \hat{\mu}_{sT}(\theta) - \hat{m}_T \]  

(3)

Furthermore, make use of an \( n_m \times n_m \) weighting matrix \( W \) (which is to be discussed in a moment). The minimization problem then reads,

\[ Q_{s,T}(\theta) := g_{s,T}(\theta)' W g_{s,T}(\theta) = \min_\theta \]  

(4)

and its solution yields the estimate \( \hat{\theta} = \hat{\theta}_{s,T} \) for the parameter vector (throughout, the prime denotes transposition of vectors and matrices).

### 2.2. The weighting matrix

The weighting matrix \( W \) in the objective function (4) is optimal if it provides the smallest asymptotic covariance for the estimator. Under these circumstances it is also ensured that the solution \( \hat{\theta}_{s,T} \) is a consistent estimate of the pseudo-true value \( \theta^o \) of the model’s parameters.

In the following it is presupposed that the number of moments is not less than the number of parameters, i.e. \( n_m \geq n_p \). To obtain the optimal weighting matrix we have to have recourse to an estimate of the long-run covariance matrix \( \hat{\Omega} \) of the empirical moments. It derives from the \( n_m \times n_m \) sample autocovariance matrices for lags \( j \geq 0 \),

\[ \Gamma_j := (1/T) \sum_{t=j+1}^{T} [m(z_{emp}^t) - \hat{m}_T] [m(z_{emp}^{t-j}) - \hat{m}_T]' \]

Typically, \( \hat{\Omega} \) is a kernel-weighted average of these matrices, truncated at a suitable lag \( p \). Using the Bartlett kernel with its linearly declining weights gives us the popular Newey-West estimator,

\[ \hat{\Omega} = \Gamma_0 + \sum_{j=1}^{p} \left( 1 - \frac{j}{p+1} \right) (\Gamma_j + \Gamma_j') \]  

(5)
(which is heteroskedasticity and autocorrelation consistent, HAC; cf. Davidson and MacKinnon, 2004, pp. 362ff). Several choices for the bandwidth $p$ are discussed in the literature, which are either rules of thumb or quite complicated expressions. We content ourselves with following a common practice that specifies $p$ as the smallest integer greater than or equal to $(T^{1/4})$.\(^9\)

If in addition a number of standard regularity conditions is taken for granted, an optimal choice of the weighting matrix in (4), which we denote as $W = W^\omega$, is given by\(^{10}\)

$$W^\omega = \hat{\Omega}^{-1}$$

A more naive idea of a weighting matrix is a diagonal matrix $W = W^d$.\(^{11}\) It takes direct account of the fact that some empirical moments may be less precisely estimated than others, from which it is concluded that a minimization procedure should pay less attention to the errors from the ‘imprecise’ moments. Accordingly, the squared deviations of the simulated from the empirical moments are weighted by the reciprocal of the empirical variances. The weighting matrix $W^d$ is thus given by the elements

$$w^d_{ij} = \begin{cases} 
\frac{1}{(1/T) \sum_{t=1}^{T} [m_i(z_{emp}^t) - \hat{m}_{T,i}]^2} & \text{if } i = j \\
0 & \text{if } i \neq j \end{cases}$$

(i, j = 1, \ldots, n_m)

Employing this matrix in the objective function (4) may equivalently be conceived as minimizing the sum of the squared $t$-statistics of the single moments (which will be shown shortly below).

2.3. The basic test statistics

If the moment functions are linearly independent, there are as many moments as parameters, and if also the single parameters (and linear combinations of them) shift the vector of simulated moments in different directions, then in essence the minimization

---

\(^9\) See Greene (2002, p. 267, fn10). Alternatively, the covariance matrix might be estimated by means of bootstrap techniques, as proposed by Winker et al. (2007). Given that some of the moments they use are conditional moments and that their empirical daily time series is empirical data over 10 years, they consider 10,000 block bootstrap samples with a window length of 250 to be an adequate order of magnitude.

\(^{10}\) Lee and Ingram (1991, p. 202) include the factor $(1+1/s)$ in the specification of the weighting matrix, i.e., they put $W = [(1 + 1/s) \hat{\Omega}]^{-1}$. The influence of the sampling uncertainty on the statistical fit is, however, clearer if this factor is here omitted and instead made explicit in eqs (8) and (13) below.

\(^{11}\) There are also applications of MSM that adopt a diagonal weighting matrix because a complete covariance matrix would be ill-conditioned. This problem arises when one seeks to match the model-generated impulse-response functions with those of a vector autoregression; see Boivin and Giannoni (2006, p. 453, fn 43).
problem (4) amounts to solving the system of \( n_m \) equations \( g_{s,T}(\theta) = 0 \) for its \( n_p = n_m \) unknowns \( \theta_1, \ldots, \theta_{n_p} \). On the other hand, if the number of moments exceeds the number of parameters, \( n_m > n_p \), the estimation is overidentified and some or even all of the components of the estimated moment vector \( g_{s,T}(\hat{\theta}_{s,T}) \) will be different from zero.

In this situation a set of \( n_p \) linear combinations of \( g_{s,T}(\hat{\theta}_{s,T}) \) will be equal to zero, and \((n_m-n_p)\) linearly independent orthogonality conditions remain that will not vanish in the estimation.\(^{12}\) Nevertheless, if the model is a satisfactory description of the real-world data generation process, then the latter conditions should be close to zero, too. To test these overidentifying restrictions the so-called \( J \) test can be applied. It uses the fact that asymptotically the optimal value \( Q_{s,T} \) from (4) is distributed as a chi-square random variable with \((n_m-n_p)\) degrees of freedom. Specifically,

\[
J := s^T Q_{s,T}(\hat{\theta}_{s,T}) / (1+s) \xrightarrow{as} \chi^2(n_m-n_p) \tag{8}
\]

(‘as’ stands for asymptotically, meaning for \( s \) fixed and \( T \) tending to infinity). A model is considered to be valid with regard to the chosen moments if \( J \) is less than the value of \( \chi^2(n_m-n_p) \) corresponding to, say, a 95% significance level. If the \( J \) statistic exceeds that critical value this can, but need not necessarily, be an indication that the model fails to mimic the empirical data in several dimensions. The problem that might generally arise is that the finite-sample distribution of the test statistic can happen to differ substantially from its asymptotic distribution, especially if, as in the case of the optimal weighting matrix \( W = W^0 \), a HAC covariance matrix is involved (Davidson and MacKinnon, 2004, p. 368).\(^{13}\)

To locate the potential failures of a model one may use the distribution of the moment vector, again the asymptotic one. It is given by

\[
\hat{\mu}_{s,T}(\theta^o) - \hat{\mu}_T \xrightarrow{as} N(0, U), \quad U := \frac{1 + 1/s}{T} \hat{\Omega} \tag{9}
\]

Therefore, referring to \( \hat{\mu}_{s,T}(\hat{\theta}_{s,T}) \) instead of \( \hat{\mu}_{s,T}(\theta^o) \), a good hint for an evaluation of the model’s ability to reproduce the \( i \)-th empirical moment is its \( t \)-statistic.\(^{14}\) Denoting this measure by \( t_i^o \) in order to make explicit that it is based on the covariance matrix \( \hat{\Omega} \) in

---

\(^{12}\)The first set is determined by the first-order optimality conditions for (4) which, with the matrix \( B \) defined below, read \([B'W] g_{s,T}(\hat{\theta}_{s,T}) = 0\).

\(^{13}\)Two elaborate studies on this subject are Kocherlakota (1990) and Mao (1990).

\(^{14}\)See Duffie and Singleton (1993, p. 945) for eq. (9). Collard et al. (2002, pp. 207f) state that also \( \hat{\mu}_{s,T}(\hat{\theta}_{s,T}) - \hat{\mu}_T \) is asymptotically normally distributed, having covariance matrix \( \hat{U} = (\hat{\Omega} - B(B'WB)^{-1}B')/T \) with \( B \) defined in (12). The authors fail, however, to make sure that \( \hat{U} \) is positive-definite. In addition, one wonders why the length of the simulations (represented by \( s \)) should play no role. In fact, there seems to be a little problem in their derivation since one equation in the middle of their proof is easily seen to imply that \( \hat{\theta}_{s,T} - \theta^o \sim N[0, (B'WB)^{-1}/T] \), in contradiction to eq. (13) below. To avoid these technicalities and also the more complicated covariance matrix, a rough-and-ready check like that in the text may suffice for a first evaluation.
(5) and thus compatible with the optimal weighting matrix \( W^o \), and now also omitting
the time indices \( s \) and \( T \), we have
\[
t_i^o = \frac{\hat{\mu}_i(\hat{\theta}) - \hat{m}_i}{\sqrt{u_{ii}}} ; \quad u_{ii} \text{ from (9), } i = 1, \ldots n_m
\]  
(10)
In the case of the diagonal weighting matrix \( W^d \) from (7), the matrix \( U \) in (9) is deter-
mined by
\[
U = (1 + 1/sT)(W^d)^{-1}/T
\]  
and the squared t-statistics result as
\[
(t_i)^2 = \frac{sT}{1 + s} \frac{[\hat{\mu}_{sT,i}(\hat{\theta}_{s,T}) - \hat{m}_{T,i}]^2}{(1/T) \sum_{t=1}^T [m_i(z_{emp}^t) - \hat{m}_{T,i}]^2}
\]  
(11)
For this alternative setting it is thus seen that the \( J \)-statistic in (8) coincides with \( \sum_i(t_i)^2 \).

Hence, as mentioned above, a set of parameters minimizing the sum of the squared t-statistics of (11) also minimizes the objective function \( Q_{s,T} \) in (4).

Before considering the model’s compatibility with other empirical moments, the ac-
curacy of the parameter estimates must be assessed. To this end recall that \( g_{s,T} \) is the
vector of the deviations of the simulated from the estimated moments in (3) and define
the \( n_m \times n_p \) matrix \( B \) by its partial derivatives with respect to the parameter changes,
\[
b_{ij} := \frac{\partial g_{s,T}(\hat{\theta}_{s,T})}{\partial \theta_j} \quad (1 \leq i \leq n_m, \ 1 \leq j \leq n_p)
\]  
(12)
(these values have to be computed numerically). The estimated parameters are asymptotically
normally distributed around their pseudo-true values with an \( n_p \times n_p \) covariance
matrix \( V \),
\[
\hat{\theta}_{s,T} - \theta^o \overset{as}{\sim} N(0, V) , \quad V = \frac{1 + 1/s}{T} (B'WB)^{-1}
\]  
(13)
That is, if \( T \) is sufficiently large, the expressions \( \sqrt{u_{ii}} \) provide a decent idea of the standard
error that is associated with the estimation of the parameter \( \theta_i \).

The term \( 1 + 1/s \) in the covariance matrix is a (direct) measure of the sampling un-
certainty in the simulations. Infinitely large \( s \) would correspond to a GMM estimation,
where one can work with the model’s population means \( \mu(\theta) \). The loss in efficiency that
results from estimating them with \( \hat{\mu}_{sT}(\theta) \) is, however, limited: for \( s = 5 \) or \( s = 10 \) it is
only 20% or 10% relative to the GMM benchmark.

2.4. Testing for additional moments

After deciding on a specific set of moments and obtaining an estimation of the model
from them that is regarded as satisfactory, one may wish to test if the model could
also account for additional moments. They may be examined in turn one after another,
or several of them are considered jointly. The general concern is with \( q \) empirical and
simulated moments, for which a function \( h = h(\theta) \) is set up in the same way as the
function \( g_{s,T} = g_{s,T}(\theta) \) in eq. (3) above. Certainly, its values \( h(\theta) \) are \( q \times 1 \) vectors (again,
reference to \( s \) and \( T \) is here omitted).
If the model is also compatible with these moments, then $h(\theta^o) = 0$ should prevail (asymptotically) for the pseudo-true values $\theta^o$. While the previously estimated parameter vector $\hat{\theta}$ will violate this equality, the deviations should not be too large. This can be tested with a Wald statistic $J_W$ (see, e.g., Davidson and MacKinnon, 2004, p. 422). Defining the $q \times n_p$ matrix $H$ by $h_{ij} = \partial h_i(\hat{\theta})/\partial \theta_j$ and $V$ being the covariance matrix of $\hat{\theta}$ from eq. (13), the statistic is given by

$$J_W = h(\hat{\theta})' (H V H')^{-1} h(\hat{\theta})$$

Under the null hypothesis $h(\theta^o) = 0$, $J_W$ is distributed as chi-square with $q$ degrees of freedom. Accordingly, the test advises the researcher to reject the moments represented by the function $h = h(\theta)$ if $J_W$ exceeds the critical value of $\chi^2(q)$ at the chosen significance level.\(^\text{15}\)

The Wald test is a most convenient test, in particular, since the model need not be re-estimated with the additional moments. It should, however, be used with caution (similar to the $J$ test mentioned above). One reason for this is that Wald statistics are not invariant to reformulations of the restrictions, even if they are mathematically equivalent. Thus, some formulations may lead to Wald tests that are well-behaved, whereas other may lead to severe overrejection, which seems to occur more often than underrejections (Davidson and MacKinnon, 2004, pp. 422f).\(^\text{16}\)

3. The Manzan–Westerhoff model and a first test of MSM

3.1. Formulation of the model

On a market for a risky asset, the model by Manzan and Westerhoff (2005) distinguishes fundamentalist traders and so-called speculators. The daily market orders of the former are inversely related to the percentage deviations of the price from the fundamental value. Speculators form expectations of tomorrow’s price and buy (sell) if it exceeds (falls short of) the current price. The inventive feature of the model is that these expectations are based on the daily random walk changes $\eta_t$ of the (log of the) fundamental value, such that the expected price is higher (lower) than the current price if the fundamental value has increased (decreased). Quantitatively the perceptions of the news are, however, biased. In the determination of the expected price more than $\eta_t$ is added to the current (log) price, and in this sense the speculators overreact to the news, if the historical volatility of the asset exceeds a certain threshold. Otherwise they underreact, which means less than $\eta_t$ is added to the current price. The volatility itself is specified as the mean of the

\(^{15}\) An example of this procedure in the context of a Real Business Cycle model, with one additional set of two joint moments, is Christiano and Eichenbaum (1992, pp. 439f, 444ff).

\(^{16}\) A straightforward illustration for OLS estimations is Gregory and Veall (1985).
absolute returns over a rolling sample period (‘history’) of length $H$. So at day $t$ it is given by

$$V_t(H) := \frac{1}{H} \sum_{\tau=1}^{H} |S_{t-\tau+1} - S_{t-\tau}|$$  \hspace{1cm} (15)$$

where $S_t$ denotes the log prices. A market maker is supposed to change the price in proportion to the sum of the demand of fundamentalists and speculators. In this way the dynamics of the log price can be reduced to the equation,

$$S_{t+1} = S_t + x (F_t - S_t) + \begin{cases} y \eta_t & \text{if } V_t(H) \geq K \\ z \eta_t & \text{else} \end{cases} \quad \eta_t \sim N(0, \sigma^2_\eta)$$  \hspace{1cm} (16)$$

Here $x, y, z$ are positive coefficients that are composed of the structural parameters of the model (with $y > z$), $K$ is the value of the threshold just mentioned, $F_t$ is the log of the fundamental value, and the stochastic innovations $\eta_t$ are normally distributed with standard deviation $\sigma_\eta$ (i.i.d., of course). As indicated, the latter are just the changes in the random walk of $F$,

$$F_t = F_{t-1} + \eta_t$$  \hspace{1cm} (17)$$

In sum, the Manzan–Westerhoff (MW) model is compactly described by the three equations (15)–(17). Obviously, if (represented by the parameter $x$) the direct influence of the fundamentalists is low, as it will turn out to be, then the dynamics of the returns $S_t - S_{t-1}$ from (16) is essentially a switching process between two normal distributions (akin to a reformulated GARCH-in-mean model). While this is a technical description, Manzan and Westerhoff (2005) motivate it by a little, experimentally supported story of investor psychology, which especially allows them to speak of the speculators’ responsiveness to news (the parameter $y$ and $z$), a time horizon, or memory, for the relevant returns ($H$), and a threshold ($K$) for the under- and overreactions.

3.2. Setting up the moment functions

Agent-based models are built to explain certain stylized facts of financial markets. There are four stylized facts, referring to returns, that have received the most intensive attention, namely: absence of autocorrelations, volatility clustering, long memory, and fat tails (see Chen et al., 2008, p. 19).\footnote{Note 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 long memory, which reflects long-run dependencies between returns.} The choice of the moments for our application of MSM will be quite in line with this concern. So let us express returns in per cent and denote them as

$$r_t := 100 \cdot (S_t - S_{t-1}) \, , \quad v_t := |r_t|$$  \hspace{1cm} (18)$$
Squared returns, which are also often referred to, behave very similar to the absolute returns and so can be here neglected. The first moments the model should be able to match are the mean and variance of \( r_t \) and \( v_t \). In setting up the moment functions that are to capture the autocorrelations, it is slightly more convenient to refer to the autocovariances (ACV). To this end we first compute the mean values of the empirical data \( r_{t, \text{emp}} \) and \( v_{t, \text{emp}} \) over their time horizon \( T \). Designating them \( r_{\text{emp}} \) and \( v_{\text{emp}} \), respectively, and writing \( z_t = (r_t, r_{t-1}, \ldots, r_{t-100}, v_t, v_{t-1}, \ldots, v_{t-100}) \) for the relevant \( L \)-stretch of the observations for day \( t \), we may generally consider the following functions \( m_i = m_i(z_t) \) (for suitable indices \( i \) when stacking some of them in the moment vector):

\[
\begin{align*}
q_{\text{mean}} : m_i(z_t) &= q_t \quad (q = r, v) \\
q_{\text{ACV} k} : m_i(z_t) &= (q_t - q_{\text{emp}})(q_{t-k} - q_{\text{emp}}) \quad (q = r, v; \ k = 0, \ldots, 100)
\end{align*}
\]  

(19)

Certainly, ‘\( q_{\text{ACV} 0} \)’ represents the variance of \( q = r, v \) and ‘\( q_{\text{ACV} k} \)’ the autocovariances with lag \( k \). Note that also the model-generated series of \( r_t \) and \( v_t \) will make reference to the empirical means in (19). If instead the simulated means were used, the entire time series, and not only its period-\( t \) contributions, would enter the definition of \( m_i = m_i(z_t(\theta)) \).

Though one will expect that the propositions of MSM continue to hold true in this case, this would still need to be demonstrated.\(^{18}\)

Fat tails are usually measured by the Hill estimator \( \hat{\alpha}_H \) of the absolute returns. With respect to a threshold value \( v^o \) specifying the tail (to which all \( v_t \geq v^o \) are assigned), it is given by \( \hat{\alpha}_H = 1/\hat{\gamma}_H \), where \( \hat{\gamma}_H = (1/\tau) \sum \max \{0, \ln v_t - \ln v^o\} \) and \( \tau \) is the number of entries in this tail. Hence lower values of \( \hat{\alpha}_H \) and higher values of \( \hat{\gamma}_H \) indicate a fatter tail of the data. The definition of \( \hat{\gamma}_H \) motivates us to put forward the following two additional moment functions, which are based on a given value \( v^o \) that specifies a 5% tail of the empirical returns:\(^{19}\)

\[
v_{\text{Hill} k} : m_i(z_t) = \begin{cases} 
\ln v_t - \ln v^o & \text{if } v_t > k v^o \\
0 & \text{else} 
\end{cases} \quad (k = 1, 2) 
\]

(20)

If \( k = 1 \), the function corresponds to the single terms in the sum that yields \( \hat{\gamma}_H \). In a simulation of the model, however, the moment ‘\( v_{\text{Hill} 1} \)’ results from dividing the sum of these \( m_i(z_t) \) through \( sT \), and since we do not know in advance how many \( v_t \) will exceed the predetermined value of \( v^o \), this expression cannot be suitably rescaled to compute the statistic \( \hat{\gamma}_H \) proper.

Whereas this kind of scaling does not matter if also roughly 5% of the model-generated \( v_t \) exceed \( v^o \), it is possible that the empirical and simulated moments ‘\( v_{\text{Hill} 1} \)’ are approximately equal although the model puts only, say, 3% in the tail specified by \( v^o \). This might happen if a larger part of these 3% attains more extreme values than the

\(^{18}\) At least we know of no corresponding treatment in the literature.

\(^{19}\) That is, exactly 5% of the empirical absolute returns are larger than this \( v^0 \), which is a common order of magnitude in empirical estimates of the fatness of tails.
empirical data. It is the idea of the second function ‘$v$ Hill 2’ to avert such a misleading match, since in this case the simulated moment would be noticeably larger than its empirical counterpart.  

3.3. Re-estimation of the model

Before applying MSM with the moment functions (19) – (20) to empirical data, it should first be tested if the method is reliable enough to identify the model itself. To this end we adopt the parameter values from Manzan and Westerhoff (2005) and try to re-estimate them from an artificial sample series.

Regarding the moments to be included in the estimation, the most obvious candidates are the means and variances of $r_t$ and $v_t$. A zero mean of the raw returns will be automatically brought about by the model, so it only makes sense to include ‘$v$ mean’. The functions ‘$r$ ACV 0’ and ‘$v$ ACV 0’ cannot be simultaneously included because this leads to a (nearly) singular covariance matrix $\hat{\Omega}$. Hence we only check the matching of the variance ‘$r$ ACV 0’.

As the correlograms in Manzan and Westerhoff (2005, p. 686) show, the raw returns have near-zero autocorrelations at all lags—even at the first lag, where several other models in the literature have difficulty avoiding significance. Since it can be reasonably expected that if the coefficients are close to zero at short lags they will also tend to vanish at longer lags, it will suffice to include the moment function ‘$r$ ACV 1’; functions with higher lags $k$ will then not provide any additional information.

In accordance with their empirical counterparts, the autocorrelations of the absolute returns in MW are significantly positive and quite steadily declining as the lag length increases. It thus suffices to select one short, one medium, and one longer lag, which are represented by the moments ‘$v$ ACV 1’, ‘$v$ ACV 5’, and ‘$v$ ACV 50’. Lastly, the two Hill functions in (20) are utilized, so that underlying the estimation is a total of 8 moment functions ($n_m = 8$). In the objective function $Q_{s,T} = Q_{s,T}(\theta)$ in (4) we use the optimal weighting matrix $W = W^o = \hat{\Omega}^{-1}$ from (6); the corresponding $J$ test statistic in (8) is designated $J^o$.

After these preparations we can take the parameter values from Manzan and Westerhoff (2005, p. 683), generate a sequence of (pseudo) random numbers and use them to simulate the model over more than 7000 days. We discard a transient phase of several  

---

20 Winker et al. (2007), in their MSM study of a simple GARCH-model and a stochastic volatility model, use directly an average of several Hill estimators across different tail sizes as one of their moments. Since at least this moment does not derive from pre-defined period-$t$ contributions like $m_i(z_t^{emp})$, their covariance matrix $\hat{\Omega}$ has to be determined in a different way from (5). In fact, the authors employ a bootstrap technique to estimate this matrix. In an earlier paper, Gilli and Winker (2003) also included the kurtosis as a moment, but noted later (in 2007, p. 138, fn 9) that this statistic is not very robust and so discriminates little between different parameter constellations.
hundred days and retain a return series over a period of $T = 6760$ days, which is of a similar length to some of the empirical series below. This series represents our empirical data, on which the model is now to be re-estimated. Of course, the simulations of the moments in the estimation procedure will be based on a different random number sequence (the same for each parameter set). The corresponding time horizon will be taken ten times longer than the empirical series, $s = 10$.

The true numerical parameters are a standard deviation $\sigma_\eta = 0.010$ for the innovations to the fundamental value, $H = 10$ for the length of the rolling sample period, and the values given in the first row of Table 1. In order to avoid unnecessary digits the table refers to $\phi := 100 \cdot x$ and $\rho := 100 \cdot K$ instead of $x$ and $K$ (‘$\phi$’ may help to recall that the coefficient is associated with the deviations of the current price from the fundamental value, and ‘$\rho$’ may be indicative of the overreaction threshold for the absolute returns in the price equation). The first row in Table 1 also shows that the simulation with the true parameters and the new random shocks yields a $J$ statistic of $J^o = 8.14$. Hence if the four parameters in that row were the outcome of the estimation they would not be rejected, since the critical 95% significance value is $\chi^2(n_m-n_p) = \chi^2(8-4) = 9.49$.

<table>
<thead>
<tr>
<th></th>
<th>$\phi$</th>
<th>$y$</th>
<th>$z$</th>
<th>$\rho$</th>
<th>$J^o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.100</td>
<td>1.350</td>
<td>0.650</td>
<td>0.750</td>
<td>8.14</td>
</tr>
<tr>
<td>2</td>
<td>2.138</td>
<td>1.352</td>
<td>0.650</td>
<td>0.742</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td>(2.554)</td>
<td>(0.081)</td>
<td>(0.032)</td>
<td>(0.052)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.100</td>
<td>1.356</td>
<td>0.655</td>
<td>0.745</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.047)</td>
<td>(0.020)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10.000</td>
<td>1.360</td>
<td>0.644</td>
<td>0.740</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.088)</td>
<td>(0.044)</td>
<td>(0.063)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>1.353</td>
<td>0.653</td>
<td>0.744</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.042)</td>
<td>(0.020)</td>
<td>(0.029)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Re-estimation of a model-generated return series ($T = 6760$).

Note: $J^o$ is the $J$ test statistics determined by eqs (4) and (8) with weighting matrix $W = W^o$ from (6) and $s = 10$. The parameters $\sigma_\eta = 0.010$ and $H = 10$ are fixed at their true values, the other true values are given in the first row, with $\phi := 100 \cdot x$ and $\rho := 100 \cdot K$. Numbers in parentheses are the standard errors from (13), while a long hyphen indicates that $\phi$ was exogenously fixed in that estimation.

Due to the sampling variability there will certainly be other parameter values that give rise to lower values of $J^o$ and the objective function $Q_{s,T}$, respectively. For these estimations it has first to be noted that the three parameters $y$, $z$ and $\sigma_\eta$ cannot be independently identified, because only the products $y\sigma_\eta$ and $z\sigma_\eta$ enter the determination of the price in (16). We may therefore maintain MW’s standard deviation $\sigma_\eta = 0.010$ as a
normalization, which amounts to an annual volatility in the fundamental value of 15.8% (on account of $0.158/\sqrt{250} = 0.010$). This issue will nonetheless be later reconsidered. For the present estimations we also fix the integer parameter $H$ at its correct value $H = 10$.

This being understood, the four parameters $\phi$, $y$, $z$, $\rho$ have to be estimated. In our first attempt, we choose an initial set of these parameters with, in particular, $\phi = 2$, start the Nelder–Mead simplex search algorithm from there (see Press et al., 1986, pp. 289–293), and restart it upon convergence several times until no more noteworthy improvement in the minimization occurs. As seen in the second row of Table 1, a decrease of the $J$ test statistic down to $J^o = 2.57$ is achieved in this way, with only marginal deviations of the other parameters $y$, $z$ and $\rho$ from their true values. According to the standard errors in parentheses, these estimates are rather precise. By contrast, the estimated $\phi$ remains at the initial order of magnitude of this coefficient, and its high standard error indicates that this estimate is not very reliable.

3.4. Discussion of the results

To get more information about the nature of the solution to the minimization problem, we fix all parameters except one at their estimated values and compute $J^o$ as a function of the variable parameter. For $y$, $z$ and $\rho$, this function is smooth and well-behaved with a distinct global minimum. For $\phi$, however, the graph shown in Figure 1 is obtained, which is remarkable for two reasons. First, the variations of this function are quite limited over the interval $[0, 5]$. Even for $\phi = 10$ the values of $J^o$ are still in the region of $3.50$. This feature clearly explains the standard error for the estimated $\phi$, which is in effect considerably larger than computed from the covariance matrix in (13). Second, the function is extremely jagged, so that it would be a hard job for any search algorithm to find a global minimum. We have to return to this problem further below.

From Figure 1 it can be inferred that the parameter $\phi$ could be fixed at any arbitrary value between (at least) 0 and 5: if anything, the estimation of the remaining three parameters would be only marginally inferior to the estimation in the second row of Table 1. The last three rows in the table for the true and two extreme values of $\phi$ fully confirm this. The standard errors and the value of $J^o$ for $\phi = 10$ are a first (weak) indication that the estimations would eventually deteriorate if $\phi$ were further increased, which is only natural as the stronger mean-reverting tendencies in the price thus brought about should sooner or later also destroy the coveted autocorrelation pattern of the returns.

On the other hand, although we know that with $\phi = 0.100$ the fundamentalists have a small role to play on the market, the last row in the table with $\phi = 0$ evidences that they may quite as well be completely relegated from the model. That is, the estimation of the model-generated returns can recognize the switching mechanism of the speculators in quite a precise manner, but it fails to capture an interplay between the two groups of traders. This, however, comes as no great surprise since there is simply no moment which
could notice the excessive deviations of the price from the fundamental value that would be caused by an insufficient influence of the fundamentalists. While it would be desirable to complement our set of moments by a measure of misalignment, this is not easily done because the model’s daily fluctuations of the fundamental value are unobservable (which was just one of the reasons why we opted for MSM). We will therefore have to accept that the parameter $\phi$ cannot be properly identified.

Finally the matching of the moments should be checked. The low $J$ test statistics reported in Table 1 suggest that there is no problem with the moments included in the estimations. In finer detail we can refer to the $t^*$-statistics from eq. (10). Computing them for the second estimation in Table 1, they are indeed less than one in modulus for all of the eight moments.

Regarding the matching of additional moments by this estimation, the Wald test statistic $J_W$ determined in (14) can be applied. We did this separately for each of the following 14 moments: the mean of the raw returns; ‘r ACV k’ for $k = 2, 3, 4, 5, 10, 50$; the variance of the absolute returns; and ‘v ACV k’ for $k = 2, 3, 4, 10, 20, 100$. Table 2

---

Figure 1: The function $\phi \mapsto J^o(\phi)$ (the other parameters fixed at their estimated values in row 2 of Table 1).
reports the cases where $J_W$ is larger than 2.66 (the critical value at the 95% level is $\chi^2(1) = 3.84$).

<table>
<thead>
<tr>
<th></th>
<th>raw returns $r_t$</th>
<th>absolute returns $v_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ACV 3 ACV 50 ACV 100</td>
<td></td>
</tr>
<tr>
<td>empirical</td>
<td>0.011 0.015 0.011</td>
<td>0.017</td>
</tr>
<tr>
<td>simulated</td>
<td>−0.009 0.008 −0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>$J_W$ &gt; 1000</td>
<td>41 32 67</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Wald test for selected additional moments (regarding the estimates in row 2 of Table 1).

The table illustrates that the pure Wald statistic can indeed be misleading. It might theoretically happen that $\phi = 2.138$ and the other parameters cause a mismatch of some of the moments. However, all of these moments, the empirical as well as the simulated ones, are close to zero. Under these circumstances it turns out that the Wald statistic can exaggerate the remaining differences. Incidentally, this holds whether the coefficient $\phi$ is part of the estimation and thus its imprecision is reflected in the parameter covariance matrix $V$ entering (14), or whether $\phi$ is exogenously fixed and so does not show up in $V$. In any case, one can learn from Table 2 that in evaluating high values of $J_W$ the levels of the empirical and simulated moments should be taken into account.

4. Estimation on empirical data

4.1. A comparison of three minimization criteria

After applying MSM and our specification of the moments to artificial data that was generated by the model itself, we are now ready to work with empirical data. Although Manzan and Westerhoff present their model as a behavioural exchange rate model, their story could quite as well relate to a stock market or stock market index. We will in this section consider several empirical exchange rates and stock market indices, but let us first concentrate on the Standard & Poor’s stock market index, the S&P 500. The series to be estimated comprises $T = 6864$ days from January 1980 until March 2007. However, as it is occasionally done in other studies, we purge the return series of the October 1987 crash; that is, the three days of October 19–21 are discarded. This does not by any means brighten up the estimation results; it will rather be seen later that the model performs better if the crash is included.
We begin the discussion with contrasting three sets of coefficients, which result from
the minimizations described below. They are enumerated in Table 3. Interestingly, apart
from $\phi$ (whose estimates as we know are not very precise) and the length of the rolling
sample period $H$, the parameters of set 1 are fairly close to MW’s tentative values.
Simulating the model with the three parameter sets and employing the same moments
as in the previous section (again putting $s = 10$), the statistics compiled in Table 4 are
obtained.

It is a little caprice of the data that the first-order autocovariance of the absolute
returns is lower than ‘$v$ ACV 5’, and one will not expect a model to mimic this feature.22
Indeed we have found that all model parameters of any relevance produce a
slow monotonic decay of the autocovariances of $v_t$, though possibly at different levels.
Furthermore, the three parameter sets dealt with in Table 4 keep the autocovariances
distinctly positive even at a lag of 50 days. This brief overview confirms that the model
is basically compatible with at least a certain subset of the stylized facts.

The first set of parameters originates with a minimization of $Q_{s,T}$ under the optimal
weighting matrix $W^o$ (some details of the minimization procedure across the five parame-
ters are described in the next subsection). Unsurprisingly, the associated $J$ test statistic
$J^o = 36.21$ rejects the model as the true data generation process. We may nevertheless
still ask which of our 8 moments can be reproduced by the simulations, and to what
extent. To this end the table reports the levels of the moments and the $t^o$-statistics from
(10). Clearly, an unsatisfactory match is observed for the variance of the raw returns and
the two Hill moments (as well as for the Hill estimator $\hat{\alpha}_H$ itself).

Taking the estimation with the first parameter set as a point of departure, it is remark-
able that the second parameter set, in spite of its considerably higher value of $J^o = 55.64$,

---

Table 3: Specification of three parameter sets.

<table>
<thead>
<tr>
<th></th>
<th>$\phi$</th>
<th>$y$</th>
<th>$z$</th>
<th>$\rho$</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>set 1</td>
<td>2.342</td>
<td>1.350</td>
<td>0.658</td>
<td>0.737</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.029)</td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>set 2</td>
<td>2.362</td>
<td>1.861</td>
<td>0.759</td>
<td>0.956</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>(0.193)</td>
<td>(0.028)</td>
<td>(0.069)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>set 3</td>
<td>3.297</td>
<td>1.960</td>
<td>0.777</td>
<td>0.993</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.026)</td>
<td>(0.050)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textit{Note}: Numbers in parentheses are the standard errors from the estimations on S&P 500
described in the text.

---

22 Besides, including the 1987 crash raises ‘$v$ ACV 1’ and ‘$v$ ACV 5’ to 0.102 and 0.118, respectively, which is a nonnegligible change in the time averages of such a long series.
Table 4: Moments, $t^o$-statistics and values of the objective functions for the three parameter sets in Table 3.

Note: The $t^o$-statistics are given by eq. (10), where all three parameter sets have the same matrix $U$ determined in (9) underlying. ‘Hill est.’ is the Hill estimator $\hat{\alpha}_H$.

achieves a largely superior collection of $t^o$-statistics. Unless a high priority is assigned to ‘r ACV 1’ and ‘v ACV 1’, a direct comparison of the $t^o$-statistics leads one to prefer the moment matching of parameter set 2 to that of set 1. Note also the far superior match of the Hill estimator.

As a matter of fact, the second parameter set has been obtained from employing the diagonal weighting matrix $W^d$ from (7) in the objective function (4). This is also evidenced by the minimal value of $J^d$ in bold face, which denotes the corresponding $J$ statistic (8). It may here be recalled that $J^d$ has already been noted to coincide with the sum of the squared $t_i$-statistics in (11), which likewise can thus be said to be minimized by parameter set 2. Although these $t_i$ are numerically different from the original $t^o$ of eq. (10), the two statistics convey the same information.

It follows that if one acknowledges the falsity of the model and rather aims at a close match of the selected moments that is directly recognizable, then ($J^d$ or, equivalently) the sum of the squared $t_i$-statistics is a more appropriate, or at least more intuitive, minimization criterion. We may then even take one little step further. It captures the idea that two moments having both moderate $t$-statistics may be preferred to one moment with a very low and one with a large statistic. Accordingly the $t$-statistics may count as

![Table 4](image-url)
a ‘loss’ only if they exceed a certain positive benchmark value $t^b$, which leads us to put forward the following objective function:

$$Q^{bt} = Q^{bt}_{s,T} := \sum_{i=1}^{n} \left\{ \max[0, |t_i| - t^b] \right\}^2 ; \quad t_i \text{ from (11)}$$  \hspace{1cm} (21)$$

Parameter set 3 is the outcome of a minimization of this function, with benchmark $t^b = 0.75$. While Table 4 still records the $t^a$-statistics of the moments, the effect of the modification in the objective function is still clearly visible: the two largest values 1.30 and 1.37 of the second parameter set for ‘$r$ ACV 1’ and ‘$v$ ACV 1’ are reduced to 1.22 and 1.15, respectively. In exchange, the absolute values of the $t^a$-statistics of the Hill moments, for example, increase from 0.36 and 0.11 to 0.71 and 0.19, but a deterioration of this order of magnitude has just been defined as negligible. Hence function (21), as documented in the last row of the table, yields a lower value for set 3 than for set 2, although set 3 is inferior in terms of the two other functions $J^o$ and $J^d$.

We may thus settle down on parameter set 3 as a combination that, in the sense discussed, achieves the best fit of the empirical moments. Comparing the standard errors in Table 3 with those in the second row of Table 1, which was concerned with the estimation of the model-generated return series, it is seen that the estimation of the coefficients $y$, $z$ and $\rho$ is also quite precise. The estimate of $\phi$ is, of course, as unreliable as in the artificial re-estimations of the model; its standard error is therefore omitted. There is no standard error for $H$ since it is an integer parameter which we fixed at alternative values and finally chose the one that gave rise to the best estimation result.

Although rejected by the $J$ test, the model is nevertheless able to produce moments that on the whole are not dramatically different from their empirical counterparts. That the level of the empirical first-order autocovariance of $r_t$ is seven times as high as the simulated ‘$r$ ACV 1’ (0.021 versus 0.003 in Table 4) does not matter since both of them will be readily classified as insignificant. So the most serious mismatch is in the first-order autocovariances of the absolute returns, with 0.084 versus 0.068 or a $t$-statistic of 1.58 (in place of the lower value of $t^a = 1.15$ given in the table).

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\text{'v ACV k'}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>empirical</td>
<td>0.474</td>
</tr>
<tr>
<td>simulated</td>
<td>0.423</td>
</tr>
<tr>
<td>$J_W$</td>
<td>37.6</td>
</tr>
</tbody>
</table>

| Table 5: Wald test of parameter set 3 for selected additional moments. |
Evaluating the matching of the moments as fairly satisfactory, we should finally ask for the performance of the moments not covered by the estimation. The autocovariances of the raw returns at other lags behave similarly to \( r \text{ ACV } 1 \), in that the empirical and simulated \( r \text{ ACV } k \) are all insignificant (checked up to \( k = 50 \)). The main autocovariances of the absolute returns together with their Wald statistics \( J_W \) are reported in Table 5. Surely, the deviation in the variance of the raw returns carries over to the variance of \( v_t \) (see ‘\( r \text{ ACV } 0 \)’, and ‘\( v \text{ ACV } 0 \)’ in Tables 4 and 5). The deviations in the other ‘\( v \text{ ACV } k \)’ are small or, at lag \( k = 20 \), still moderate. It can thus be concluded that the model fits the additional empirical moments quite as well as those entering the estimation itself.

4.2. A note on the minimization procedure

Figure 1 in Section 3.4 has illustrated the existence of a plethora of local minima of the objective function \( Q_{s,T} \) for \textit{ceteris paribus} variations of the parameter \( \phi \). The problem does not disappear if \( \phi \) is fixed in the estimation. A multitude of local minima still persists if the three parameters \( y, z, \rho \) vary jointly in a suitable manner. Minimization problems with these properties are often approached with the (stochastic) method of simulated annealing.\(^{23}\) Employing this search algorithm requires, however, a very large number of evaluations of the objective function, which is relatively costly in the present context. In addition, the application of the algorithm and the tuning of its parameters to the specific problem at hand requires some experience. We therefore tried another approach, which in the end turned out to work sufficiently well.

Before going on, it should be mentioned that the integer parameter \( H \) was set exogenously. For each value of \( H \) within a reasonable range an estimation was performed and finally the one with the lowest \( Q_{s,T} \) was taken. We also made some explorations in which \( H \) was included in the estimation. We even treated \( H \) as a continuous variable and interpolated the volatility function \( V_t = V_t(H) \) correspondingly, but more often than not the relationship \( H \mapsto Q_{s,T}(H) \) proved nevertheless to be ill-behaved and so this attempt was given up again.\(^{24}\)

As already mentioned, we used the Nelder–Mead simplex algorithm to find a minimum of \( Q_{s,T} \). Clearly, in the presence of multiple local minima it has to be run repeatedly. Doing this we made an important observation: if, upon restarting the simplex algorithm, an improvement is found from one local minimum to the other, the solutions for \( y, z, \rho \) change in the same direction, while the changes in \( \phi \) do not seem to follow a systematic pattern. This feature motivated the following ‘meta search’ procedure.\(^{25}\)

\(^{23}\)A good introduction to simulated annealing for continuous variables is Corona et al. (1987).

\(^{24}\)\( V_t \) can have real arguments \( H \) by computing \( V_t(H_1) \) and \( V_t(H_2) \) for \( H_1 \) the largest integer less than \( H \) and \( H_2 \) the smallest integer larger than \( H \), and then defining \( V_t(H) \) as a correspondingly weighted average of \( V_t(H_1) \) and \( V_t(H_2) \).

\(^{25}\)In other MSM estimations with less information one might try Gilli and Winker’s (2003)
Given a rolling sample period $H$, choose a parameter vector $\phi^o$ (if it is part of the estimation), $y^o$, $z^o$, $\rho^o$; this set may be the outcome of a first round of searching, or the result of an estimation with a different value of $H$. Start from this vector and iteratively search upward. That is, after the simplex algorithm starting from there has found a local minimum, maintain the optimal value of $\phi$, increase the optimal values of $y$, $z$, $\rho$ by a certain fraction of their standard errors, and restart the algorithm from there to find a new and perhaps lower local minimum. Repeat this until no further improvement is found.

Subsequently, subtract a certain fraction of the recent standard errors from the initial $y^o$, $z^o$, $\rho^o$, start the algorithm from $\phi^o$ and these values, and iteratively search downward. That is, after the simplex algorithm starting from there has found a local minimum, maintain the optimal value of $\phi$, now decrease the optimal values of $y$, $z$, $\rho$ by the same fraction of their standard errors as before, and restart the algorithm from there. Again, repeat this until no further improvement is found.

Having completed this upward and downward search, choose the better result of the two. Then take this solution as the new initial vector $\phi^o$, $y^o$, $z^o$, $\rho^o$ and repeat the entire upward and downward search procedures. The only difference is that the search is now more localized, in the sense that the temporary local minimum values of $y$, $z$, $\rho$ are modified by a lower fraction of their standard errors than before.

In principle, having obtained the minimum solution from this second iterated upward and downward search, a third round could be started with still a lower fraction of the standard errors to modify the temporary local minima. Two such rounds, however, appear to be satisfactory for our purpose.\(^{26}\)

Normally the modification of the parameters that have constituted a local minimum will lead to an increase of the objective function. Accepting such a temporary deterioration in the hope that the simplex algorithm reaches from there a more promising region of the parameter space is akin to the basic idea of simulated annealing (SA). Likewise, reducing the step size for this exploratory search bears some similarity to the gradual cooling down of the ‘temperature’ in SA. The main difference of our augmented simplex algorithm from SA is that in the latter the parameter explorations go in all directions and their step size is random, whereas we can here utilize more specific information for the re-initialization of the simplex algorithm to find the next local minimum. This gives us reason to believe that we can get along with less computational effort than SA. A systematic comparison of the two numerical search algorithms is, however, beyond the scope of this paper.

\(^{26}\)With respect to a given value of $H$ and programming it in Delphi (Pascal), this procedure takes between 5 and 8 minutes on a standard personal computer.
4.3. Estimation results for different financial markets

After we have decided to work with the objective function $Q^{bt}$ specified in eq. (21) above and standardized the minimization procedure as described in the previous subsection, we can now estimate the model on the daily returns from a few other financial markets and compare the results. First, the 1987 October crash should be included in the S&P 500. Second, we wish to check how similar an estimation of the Dow Jones Industrial Average Index may be to the S&P 500. We also consider the German stock market index DAX and the Japanese Nikkei index (in both of which the 1987 crash had less dramatic effects). Lastly, we are interested in the question if the model can differentiate between stock and foreign exchange markets. To this end, we test the model with the exchange rates of the US dollar (USD) against the Deutsche Mark (DEM), and the Japanese Yen (JPY) against the US dollar. The sample periods of these empirical series and the number of observations are given in Table 6.

<table>
<thead>
<tr>
<th>series</th>
<th>begin of series</th>
<th>observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500$^{pc}$</td>
<td>Jan 1980</td>
<td>6867</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>Jan 1980</td>
<td>6864</td>
</tr>
<tr>
<td>Dow Jones$^{pc}$</td>
<td>Jan 1980</td>
<td>6867</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>Jan 1980</td>
<td>6864</td>
</tr>
<tr>
<td>DAX</td>
<td>Nov 1990</td>
<td>4115</td>
</tr>
<tr>
<td>Nikkei</td>
<td>Jan 1984</td>
<td>5712</td>
</tr>
<tr>
<td>USD/DEM</td>
<td>Jan 1980</td>
<td>6862</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>Jan 1980</td>
<td>6857</td>
</tr>
</tbody>
</table>

Table 6: The empirical series to be estimated.

Note: All series are daily and end around March 20, 2007. Superscript ‘pc’ stands for ‘purged of crash’, that is, the returns of the three days October 19 – 21, 1987, are excluded. The stock market indices are the close price adjusted for dividends and splits. Following euro adoption, USD/DEM are pseudo rates imputed by applying the euro locking rate to the current euro exchange rate.

The feature that, as discussed above, the coefficient $\phi$ cannot be precisely identified prompts us to set it exogenously at a common value for all series. Since within a certain range one value is as good as any other, we quite arbitrarily choose $\phi = 0.50$. The estimations resulting from this are reported in the left part of Table 7.\textsuperscript{27} Alternatively, we have also included $\phi$ in the estimation, where our extensive search procedure started

\textsuperscript{27} The standard errors of the estimated parameters are of a similar order of magnitude to those in Table 3 and are so no longer discussed.
out from $\phi = 2$. Here the table, in its right part, only reports the improvement in the objective function $Q^{bt}$ that (if ever) could thus be achieved.

The common value of $\phi$ is one specification to make the results better comparable across the financial markets. An additional issue is the scaling of the standard deviation $\sigma_\eta$ of the random walk of the fundamental value, which we had so far normalized at $\sigma_\eta = 0.010$. This choice also determines the levels of $y$ and $z$. However, in accordance with their interpretation of the demand of the speculators, Manzan and Westerhoff centered the two parameters around unity (see the first row in Table 1). In Table 7 we have achieved this convenient and theoretically motivated property by a suitable rescaling of $\sigma_\eta$. That is, having estimated $y$ and $z$ on the basis of $\sigma_\eta = 0.010$, we specify new values $\tilde{\sigma}_\eta$, $\tilde{y} > 1$ and $\tilde{z} < 1$ that have to obey the relationships

$$
\tilde{y} \tilde{\sigma}_\eta = y \sigma_\eta , \quad \tilde{z} \tilde{\sigma}_\eta = z \sigma_\eta , \quad \tilde{y} - 1 = 1 - \tilde{z}
$$

The solution to these equations is

$$
\tilde{y} = 2y / (y + z) , \quad \tilde{z} = 2z / (y + z) , \quad \tilde{\sigma}_\eta = (y + z) \sigma_\eta / 2
$$

In this way the model has regained some of its structural interpretation. The values of $\tilde{y}$ and $\tilde{z}$ allow us a direct assessment of the speculators’ over- and underreactions to the news on the different markets, and $\tilde{\sigma}_\eta$ gives us an impression of the relative volatilities of the fundamental news underlying these markets.\(^{28}\)

The last two entries in the first row of Table 7 reproduce the value of $\phi$ of parameter set 3 in Table 3 and the minimal value of $Q^{bt}$ in the last column of Table 4. The first row in Table 7 shows that the constraint $\phi = 0.50$ leads to a certain deterioration in the objective function. The changes in the coefficients are, however, only marginal, which once again indicates the almost negligible role of $\phi$ (with $\sigma_\eta = 0.010$, the other coefficients are estimated as $y = 1.954$, $z = 0.776$, $\rho = 0.991$, which may be compared to parameter set 3 in Table 3).

Including the three crash days of October 1987, while still fixing the rolling sample period at $H = 13$, improves the fit of moments. A free choice of $H$, where $H = 10$ turned out to be optimal, even decreases practically all $t$-statistics below their benchmark $t^b = 0.75$. The corresponding changes in the parameter estimates remain nevertheless fairly limited.

The estimation of the Dow Jones produces results that are not much different from the first three rows in Table 7. Since the segments covered by the two stock market indices are similar, this attests to a reasonable robustness of the model.

As concerns the DAX and Nikkei, the over- and underreactions are about as strong as in the two American indices. In contrast, from the estimation of the German and Japanese

\(^{28}\)If one does not share Manzan and Westerhoff’s theoretical motivation and has an idea of a credible standard deviation $\sigma_{c\eta}$, one can easily compute the corresponding value of $y$ and $z$ from $\tilde{y}$, $\tilde{z}$, $\tilde{\sigma}_\eta$ reported in Table 7 and the relationships $\tilde{y} \tilde{\sigma}_\eta = y \sigma_{c\eta}$, $\tilde{z} \tilde{\sigma}_\eta = z \sigma_{c\eta}$.
markets we infer that their fundamental news process exhibits a higher volatility. It makes good sense that in these cases also the threshold $\rho = 100 \cdot K$ in eq. (16) for the recent history of the returns, above which the speculators overreact to the news, is higher than for the S&P 500 and Dow Jones (although the relationship between $\tilde{\sigma}$ and $\rho$ is not strictly monotonic).

The estimation of the two foreign exchange rates in the last two rows of Table 7 are remarkable for three reasons. First, the matching of the moments is inferior to the stock markets. Second, the overreaction is weaker than on the stock markets. And third, the fundamental values on the foreign exchange markets appear to be far less volatile. To this conforms a lower threshold $\rho$, which is also brought out by the estimations. From this and the results above we can conclude that the model jointly with the MSM estimation has some notable explanatory power.

Let us lastly turn to the model’s different goodness-of-fit on the stock and foreign exchange markets. For a direct comparison we take the S&P 500 index on the one hand (excluding the crash), and the two foreign exchange rates on the other hand. For each of these three series, Table 8 contrasts the simulated with the empirical moments and

---

### Table 7: Estimation of the series listed in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>$\phi$</th>
<th>$\tilde{y}$</th>
<th>$\tilde{z}$</th>
<th>$100 \cdot \tilde{\sigma}$</th>
<th>$\rho$</th>
<th>$H$</th>
<th>$Q^{bt}$</th>
<th>$Q^{bt}$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500pc</td>
<td>0.50</td>
<td>1.43</td>
<td>0.57</td>
<td>1.37</td>
<td>0.99</td>
<td>13</td>
<td>1.34</td>
<td>1.19</td>
<td>3.30</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.50</td>
<td>1.42</td>
<td>0.58</td>
<td>1.28</td>
<td>0.92</td>
<td>13</td>
<td>0.18</td>
<td>0.19</td>
<td>2.78</td>
</tr>
<tr>
<td>Dow Jonespc</td>
<td>0.50</td>
<td>1.41</td>
<td>0.59</td>
<td>1.35</td>
<td>1.00</td>
<td>13</td>
<td>2.57</td>
<td>2.25</td>
<td>2.83</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>0.50</td>
<td>1.44</td>
<td>0.56</td>
<td>1.42</td>
<td>1.02</td>
<td>13</td>
<td>0.61</td>
<td>0.62</td>
<td>2.18</td>
</tr>
<tr>
<td>DAX</td>
<td>0.50</td>
<td>1.48</td>
<td>0.52</td>
<td>1.82</td>
<td>1.24</td>
<td>13</td>
<td>1.48</td>
<td>1.47</td>
<td>2.07</td>
</tr>
<tr>
<td>Nikkei</td>
<td>0.50</td>
<td>1.42</td>
<td>0.58</td>
<td>1.69</td>
<td>1.25</td>
<td>9</td>
<td>0.07</td>
<td>0.06</td>
<td>2.79</td>
</tr>
<tr>
<td>USD/DEM</td>
<td>0.50</td>
<td>1.31</td>
<td>0.69</td>
<td>0.80</td>
<td>0.63</td>
<td>16</td>
<td>4.23</td>
<td>3.55</td>
<td>2.88</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>0.50</td>
<td>1.38</td>
<td>0.62</td>
<td>0.85</td>
<td>0.66</td>
<td>11</td>
<td>7.95</td>
<td>7.66</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Note: The estimations are based on the function $Q^{bt}$ as it is defined in (21), with $t^\beta = 0.75$. The values of $\tilde{y}$, $\tilde{z}$, $\tilde{\sigma}$ are derived from $\sigma = 0.01$ and the estimates of $y$, $z$ as stated in eq. (23). In the left part of the table, $\phi$ was exogenously set at $\phi = 0.50$, the right part reports the estimated value of $\phi$ and the value of the objective function thus brought about.

---

29 It is interesting to compare the latter two points with the distinguishing feature of stock and foreign exchange markets that was revealed by an empirical analysis of a totally different model by Alfaro and Franke (2007). There the stock markets were characterized by a consistently higher population share of noise traders vis-à-vis the stabilizing fundamentalists.
The empirical moments of S&P 500pc are reproduced from Table 4. The differences in the simulated moments of parameter set 3 in the same table (where $\phi$ was included in the estimation) and in Table 8 for the S&P 500pc (where $\phi$ was fixed at 0.50) are hardly worth mentioning, although the difference in the minimal value of $Q^{bt}$ with 1.19 in the first case and 1.34 in the second seems somewhat larger. Note that the $t$-statistics of the moments are mostly slightly higher in modulus than their $t^2$-statistics (see the last column in Table 4).

Regarding the two exchange rates it is basically four moments that are responsible for their inferior performance: the mean of the absolute returns, their autocovariance at lag 50, and the two Hill moments. However, as the last row of Table 8 demonstrates, the matching of the Hill estimator itself is quite acceptable, so that in this case the high $t$-statistics of the Hill moments are somewhat misleading. Since they contribute roughly one half to the final value of $Q^{bt}$ (the sum of the corresponding two terms $(|t_i| - t^b)^2$ is 1.91 for USD/DEM and 3.89 for JPY/USD), the fit of the exchange rates, although being worse than the stock market indices, is still not so bad after all. Hence the model can also be said to mimic the daily returns of the two foreign exchange markets to a satisfactory degree.
5. Conclusion

Any structural model is false and therefore, trivially, cannot be expected to reproduce reality in all facets. Given that a model is meant for estimation at all, it will fulfill its purpose if it captures reality along certain dimensions, for which it was also more or less designed. Besides the treatment of unobservable variables and nonlinearities in the model dynamics, it is one advantage of the method of simulated moments that it obligates the researcher to make these dimensions explicit, in the form of moments. Our estimations have emphasized this point in that, after some explorations, we decided to abandon the econometrically optimal weighting of the moments in the objective function. We instead postulated directly the minimization of the $t$-statistics of the moments, or more precisely, of the sum of their squared deviations from a benchmark value. As discussed at the beginning of Section 4, this gives us a more intuitive measure of how well the model-generated moments are able to match the empirical moments.

Applying this approach to a minimal agent-based asset pricing model by Manzan and Westerhoff (2005), it was found that the model performs reasonably well in terms of our criterion, where the selected moments themselves relate to the daily raw and absolute returns and specify the stylized facts of their autocorrelation patterns and fat tails. All parameters of the model could thus be identified except the one that reflects the relative importance of the fundamentalists. This failure was seen to be due to the unobservability of the daily fluctuations of the fundamental value, so that we had no empirical moment (readily) available that could take account of a possible misalignment if the influence of the fundamentalists is too weak. Dealing with this problem could be one challenge for future research.

On the other hand, the coefficients characterizing the behaviour of the group of speculators come out meaningful and are practically not affected by the unidentified parameter. They are furthermore quite robust across similar markets. In particular, the estimations offer us an interpretation of why stock and foreign exchange markets are different: on the stock markets the model’s speculators are more responsive to the arrival of news, and the fundamental value itself or its news process is more volatile. Given its simplicity, the estimation results lend the model already a remarkable explanatory power. Nevertheless, there is certainly scope for structural improvement, and for applying similar versions of MSM to these enhanced models.

References


