

Lux, Thomas

Working Paper

Can heterogeneous agent models explain the alleged mispricing of the S&P 500?

Economics Working Paper, No. 2020-03

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Lux, Thomas (2020) : Can heterogeneous agent models explain the alleged mispricing of the S&P 500?, Economics Working Paper, No. 2020-03, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/217226>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

C | A | U

Christian-Albrechts-Universität zu Kiel

Department of Economics

Economics Working Paper
No 2020-03

Can Heterogeneous Agent Models Explain the Alleged Mispricing of the S&P 500?

by Thomas Lux

issn 2193-2476



Can Heterogeneous Agent Models Explain the Alleged Mispricing of the S&P 500? *

Thomas Lux [†]

April 21, 2020

Abstract

Tests of excessive volatility along the lines of Shiller (1981) and Leroy and Porter (1981) count among the most convincing pieces of evidence against the validity of the time-honored efficient market hypothesis. Recently, using Shiller's distinction between ex-ante rational (fundamental) price and ex-post rational price, Schmitt and Westerhoff (2017) have demonstrated that the difference between S&P 500 market prices and their ex-post counterparts exhibits a bi-modal distribution speaking for the prevalence of long periods of either undervaluation or overvaluation. Schmitt and Westerhoff (2017) also show that this new stylized fact is shared by a large set of nonlinear behavioral models of speculative interactions between heterogeneous market participants. Most of these models allow some form of chartist or fundamentalist strategy, and the

*Helpful comments by Reiner Franke, Torsten Trimborn, Frank Westerhoff and by seminar participants at the University of Kiel and RWTH Aachen are gratefully acknowledged.

[†]Author's Address: Department of Economics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany, Email: lux@economics.uni-kiel.de.

more recent members of this family of models also allow for agents switching between both alternatives according to some fitness criterion. Here I go one step further exploring which (if any) of this legacy of behavioral models fits best the data. I discuss econometric issues in the estimation of these highly complex nonlinear models, and estimate the parameters of different versions of seven canonical models. As it turns out, most of these models perform not better than a linear chartist-fundamentalist model, and often their fit is worse than the fit of this benchmark. Among the models considered here, the one proposed by Franke and Westerhoff (2012) is the only exception. Estimation of the model confidence set indicates that this model is not outperformed by other candidates, and depending on the setting and the confidence level, it is often found to be the single member of the model confidence set.

Keywords: Stock market dynamics, bubbles and crashes, nonlinear dynamics, chartists and fundamentalists, model confidence set

JEL classification: G12, G14, G17

1 Introduction

Tests for excessive volatility of asset prices compared to the change of their fundamental values count among the most convincing circumstantial evidence against the efficient market hypothesis. Pioneered by Shiller (1981) and Leroy and Porter (1981), these tests are based on a comparison of market prices with ex-post rational prices computed on the base of realized dividends. If prices would reflect rational forecasts of future earnings (i.e., if they were ex-ante rational), they should be characterized by a lower volatility than their ex-post counterparts. The mere juxtaposition of both series for the S&P 500 in Shiller (1981) shows that if anything the opposite is the case: Rather than being less volatile, the realized price movements exhibit fluctuations that are much larger than those of ex-post rational prices. Indeed, recorded S&P 500 prices appear to fluctuate in wide swings around the ex-post rational values giving the impression of prolonged periods of overvaluation and undervaluation (cf. Fig. 1). Subsequent literature has somewhat toned down the extend of violation of Shiller’s null hypothesis (e.g. Flavin (1983)), by pointing out small sample biases and other statistical problems. However, the prevalence of excess volatility has never been convincingly refuted and it appears to be accepted within the profession that standard models of expected returns cannot explain the documented level of stock price volatility (e.g. West (1988)) and that either models with noise traders, or naive investors, or systematic time variation of discount rates would be called for to explain this “anomaly”.

In a recent paper, Schmitt and Westerhoff (2017) have added an interesting twist to this long-lasting discussion. Schmitt and Westerhoff show that the price distortion, i.e., the differences between the historical prices and their ex-post rational counter-

parts is characterized by a bimodal distribution. Using appropriate statistical tests they are indeed able to show that the visual impression of a distribution with two peaks, to the left and the right of the origin, is confirmed by a highly significant rejection of the null of unimodality of the distribution of distortions. This previously overlooked feature speaks even more in favor of some elements of fads, fashions or speculative excesses playing an important role in the price formation process. Schmitt and Westerhoff point out that bimodality could only come about if there were non-linear forces in the price formation process. In contrast, a linear model could only give rise to a unimodal shape of the resulting distribution.

Models of fads, fashions and speculative overreaction typically entail nonlinear components as important building blocks, and so, this statistical property corresponds to systematic changes between periods of overvaluation or undervaluation, rather than random deviations from the ex-post rational benchmark. Schmitt and Westerhoff also show that the bimodality of price distortions is shared by a large collection of models with heterogeneous agents pursuing trading strategies like technical or chartist trading rules, following the majority of the other market participants or obeying other behavioral traits that would be incompatible with perfectly rational formation of prices. It is interesting to note that while there exists a long legacy of models with non-rational expectations that come along with bimodal price distortions, I am not aware of any model of time-varying discount rates that could explain this phenomenon. While in recent literature, time variation of discount rates or risk premiums has become the “de facto explanation for excess volatility” (Giglio and Kelly, 2018) the new stylized fact discovered by Schmitt and Westerhoff suggests that heterogeneity of expectations and learning may have been unduly neglected in this area. Since variation of discount rates and of expectations affect either the nominator

or denominator of a standard asset pricing equation, theories of the time change of one of these entities could be transformed into theories of the other (Cochrane, 2011). However, such mathematical equivalence does not imply that an equivalent theory based on discount rates would be economically plausible. Giglio and Kelly (2018) indeed show that violation of internal consistency requirements for the term structure of various short and long term claims allows to rule out variation of discount rates as the source of observed excess volatility. They also show that their findings are consistent with a recent model of extrapolative formation of expectations. In this paper, I built upon the work by Schmitt and Westerhoff (2017) and go one step further: The aim here is to estimate the parameters of the collection of behavioural models reviewed by Schmitt and Westerhoff and to evaluate their performance and explanatory power for the S&P 500 distortions.

In recent literature, there is one other contribution with a closely related approach: Majewski et al. (2020) estimate an extended version of the model of Chiarella (1992), and show that the estimated parameters indeed give rise to bi-modality for certain financial data (e.g, the US stock market). Majewski et al., however, do not compare the performance of this model with other candidate models, and they use a state-space approach to identify the fundamental value as a hidden variable (while here we use Shiller's ex-post fundamental value as a benchmark).

The rest of the paper proceeds along the following lines: in sec. 2, we outline our general approach to estimation of a large set of behavioral asset pricing models, and we discuss the justification for a straight maximum likelihood estimation of complex nonlinear models like the ones we are interested in. Sec. 3 then starts the sequence of our estimation exercises with some simple benchmark models that use polynomials up to cubic order as a simple formalisation of the nonlinear structure of the price for-

mation process. Secs. 4 to 6 consider various subsets of the list of models considered by Schmitt and Westerhoff (2017). We start with Zeeman’s (1974) “catastrophe” approach (sec. 4), and move on to “first generation” models of speculative interaction of the early 90ies (sec. 5), followed by more complex models with strategy switching by the agents (sec. 6). Sec. 7 compares all models via the estimation of the model confidence set, i.e., it determines those models that cannot be outperformed by others at a certain level of significance. Depending on the significance level, we find either a dominance of one single model (Franke and Westerhoff, 2012), or a large sample of non-rejected models that also includes a linear model with chartists and fundamentalists, but excludes the random walk (hence pointing to non-random structure). Sec. 8 finally concludes.

2 General Approach

As demonstrated in Shiller’s (1981) derivation of his excess volatility test, rationality of agents’ expectations entering their computation of the alleged arbitrage-free price or fundamental value $p_{f,t}$ requires that there should be no systematic deviation between the expected dividends and the realised dividends, and predictions of dividends and prediction errors should be uncorrelated. In his design of a test for excessive volatility, Shiller introduces the concept of the ex-post rational price p_t^* which is the price one would obtain under perfect foresight of all future dividends (i.e., the price that could be computed *ex-post* when the realization of dividends is known). The difference between p_t^* and the ex-ante rational price or fundamental value computed on the base of information at time t , $p_{f,t}$, then consists of the discounted sum of prediction errors for future dividends which all should be uncorrelated random errors under

rational formation of dividend expectations.

Denoting this sum by ε_t , we have the following simple relationship: $p_{f,t} = p_t^* + \varepsilon_t$. Since most of the models we attempt to estimate have some form of fundamentalist baseline strategy as one of their building blocks, we can substitute $p_t^* + \varepsilon_t$ for $p_{f,t}$ using the computation of the ex-post rational price by Shiller. The prediction error ε_t then enters as the unknown stochastic component. Estimating various models with behavioural features allows an assessment of how much of the mispricing is explained by their behavioural ingredients and how much remains as unexplained prediction error.

Most models will also be estimated in an alternative form in which noise enters as an additional factor of excess demand besides the excess demand functions of various groups of traders, rather than as the prediction error of dividends. While for some models, both alternatives are observationally equivalent, for others they constitute alternative formalisations which might differ in their explanatory power. Most models also allow a formalization either in market prices or alternatively in their logarithms. The latter formalization is quite natural if one assumes that fundamentalists, for instance, react on the relative mispricing rather than absolute deviations between market price and assumed fundamental value. Fig. 1 shows that the mispricing of the S&P500 appears visually different in both representations. Only in the logarithmic representation does the distortion appear to be stationary with long periods of overvaluation interrupted by relatively short drawdowns. In the original data format, the impression is rather of a non-stationary process with an increasing amplitude of deviations from fundamental values ¹. We might, thus, conjecture that the models

¹A standard Dickey-Fuller test leads to non-rejection of its null hypothesis at the 5 percent level of significance for both time series. The pertinent values of the test statistics are -2.402 for the original data and -2.814 for the logarithmic mispricing. The latter is very close to the 5 percent

should rather be used in their log forms, but we will nevertheless apply them to both alternatives to see what difference the data format makes. In the following we group our estimation exercises into four categories: We start with some simple general functional forms, from a linear model to a cubic specification. We then move on to the simplest model covered in Schmitt and Westerhoff, the “catastrophe model” of Zeeman (1974) followed by the “first generation” of simple models with heterogeneous agents with given strategies. In the next step we estimate more involved models that allow agents to switch between groups, often governed by some success criteria.

A few remarks are in order on the theoretical justification of our use of a standard maximum likelihood approach for estimation of the above models. First, it should be noted that in all models, identification of all parameters should be generically guaranteed because of their nonlinearity (cf. McManus, 1992). Nevertheless, near-collinearity of certain parameters can often be an issue as we will see below. Another issue is whether maximum likelihood estimation is consistent and obeys some convenient limit laws in the present framework. While there do not appear to be any results that can strictly be applied to our settings, recent research suggest that maximum likelihood might indeed be consistent. The basic theoretical problem here is that most models of speculative dynamics have a nonlinear system as their deterministic kernel that undergoes a number of bifurcations under variation of its parameters leading from asymptotic stability of its steady state via cyclical dynamics to chaotic motion.

Interestingly, with their combination of centripetal and centrifugal forces, all models appear to satisfy a weak notion of stability: The system stays within a certain

boundary (2.86). The fact that a unit root is not more strongly rejected, could actually be consistent with nonlinear bubble-burst dynamics as it has been frequently reported that such dynamics are hard to distinguish from unit-root processes under the Dickey-Fuller test even though they might give rise to stationary dynamics, cf. De Grauwe et al. (1993) and Lux and Marchesi (2000).

region around its unique stationary state when converging to a limit cycle or chaotic attractor, but does not generate divergent trajectories. However, notwithstanding such limiting behavior, standard asymptotic theory is not routinely applicable in the presence of chaotic attractors because of the long-term dependence chaotic dynamics implies by its very definition. Only very recently have first results been published on the consistency of maximum likelihood estimation for certain classes of chaotic dynamics (McGoff et al., 2015a). While it is unclear whether their technical conditions for consistency would be applicable in our case, their setting is also not exactly the one we have adopted here. Namely, what they investigate is chaotic systems with observational noise only. In our above models, in contrast, noise enters in the price formation, and, therefore, the dynamic process itself. Such a framework would rather be categorized as an *observation-driven dynamic model*. This formal structure is well-known from the family of GARCH models. The general problem of maximum likelihood estimation in such a setting has been investigated by Douc et al. (2013) and Douc et al. (2015) who provide sets of regularity conditions under which maximum likelihood estimation is consistent for observation-driven models. They also show that the conditions for consistency of maximum likelihood are fulfilled for certain more complex members of the GARCH family. A further generalisation is achieved by Sim et al. (2019) who derive conditions for consistency of the maximum likelihood estimator for observation-driven systems with higher-order lag structures. This more general framework nests, for instance, the model by De Grauwe et al. (1993) as can be inferred from eq. (16).

Since consistency heavily relies on ergodicity of the process, we might expect that if the chaotic attractors of the above models are ergodic, maximum likelihood should work as expected. However, a proof of ergodicity is outside the scope of the

present paper. If the parameter would fall into a region with asymptotic convergence or cyclical dynamics, existence of a stationary distribution should be guaranteed. However, along a typical bifurcation route variation of one parameter would lead to sudden switches between higher-order cycles and chaotic attractors which could lead to discontinuous behavior of the likelihood. Furthermore, different attractors could co-exist for a given set of parameters for certain nonlinear systems. Many examples of such behavior can be found in the dynamic system literature, and indeed the present author found multiple co-existing attractors for certain parameter values in the De Grauwe model when preparing classroom presentations of the latter. Since each attractor has its own basin of attraction and different starting values could lead to different distributions of the state variable(s) along the resulting trajectories, such a scenario would obviously be a non-ergodic one.

Fortunately, in our setting, we are not dealing with deterministic systems, and adding noise could dramatically change the properties of such a system. As demonstrated via selected examples in Anishchenko et al. (2000), random disturbances of a dynamic system with multiple attractors could lead to the emergence of a single stationary distribution that connects the different attractors of its deterministic counterpart. Intuitively, adding noise to a deterministic system would lead to repeated jumps between the basins of attraction of different attractors so that a joint probability distribution would be obtained that merges the different basins. Since we assume noise with unbounded support in our applications, this effect should prevail in all our models preventing the existence of multiple separate attractors in the resulting stochastic system.

All in all, in light of the recent findings cited above, we have some reason to be confident that maximum likelihood estimation should be consistent in our set-

ting. We note, however, that the technical conditions for consistency have not been demonstrated to hold for the models under scrutiny. It is also worthwhile to note that consistency of parameter estimates would not necessarily imply consistency of the reconstructed trajectories of the underlying hidden state(s) of the system. Judd (2007) has demonstrated that maximum likelihood often is unable to reconstruct the particular chaotic path underlying the observations of a dynamic process from a chaotic attractor with observational noise. This seems plausible since the chaoticity of the process might provide many alternative trajectories that have a smaller error variance, and, therefore, seemingly appear to be more accurate reconstructions. Hence, overfitting in this sense seems intrinsic to chaotic processes. However, McGoff et al. (2015b) show that for an entire class of chaotic models, this inconsistency of reconstructed trajectories does not imply inconsistency of parameter estimates. Keeping all these caveats in mind, we move on to our empirical findings.

3 Simple Benchmark Models

All the behavioral models considered by Schmitt and Westerhoff (2017) as candidate explanations of the deviation between the S&P 500 and its ex-post rational price are strongly nonlinear specifications. While their nonlinearities are motivated by a variety of economic arguments, we can also use a more generic specification of the potential nonlinear forces at work.

In order to establish a simple benchmark, I use the basic ingredients of all models: the presence of chartist and fundamentalist investors in the market, and formulate its most basic specification using a general polynomial functional form with entries up to cubic order. Following the standard set-up, fundamentalists are assumed to react

on deviations between the fundamental value (here the ex-post rational price) and the current market valuation, i.e. their excess demand will be expressed as:

$$ED_{f,t} = a_1(p_{f,t} - p_t) + a_2(p_{f,t} - p_t)^2 + a_3(p_{f,t} - p_t)^3. \quad (1)$$

Chartists, in contrast, react on price changes in the hope of their continuation at least in the short run:

$$ED_{c,t} = b_1(p_t - p_{t-1}) + b_2(p_t - p_{t-1})^2 + b_3(p_t - p_{t-1})^3. \quad (2)$$

Prices are assumed to adjust in the presence of overall excess demand:

$$p_{t+1} = p_t + \beta ED_t \quad (3)$$

with

$$ED_t = ED_{f,t} + ED_{c,t} + \epsilon_t. \quad (4)$$

The parameters to be estimated will be the six parameters of the two excess demand functions plus σ_ϵ , the standard deviation of the noise factor which here and in all other models will be assumed to follow a Normal distribution. The price adjustment speed β will be assumed to be equal to one as it cannot be estimated independently of the other parameters. Here it is assumed that noise enters as a noise trading factor in excess demand. One could alternatively assume that it enters in the determination of the fundamental value following the argument in the preceding section which would lead to $ED_{f,t} = \sum_{i=1}^B a_i(p_{f,t} + \epsilon_t - p_t)^i$. However, since the noise factor then appears in different powers of the polynomial form of fundamentalists'

excess demand this version cannot be easily estimated and is not pursued in this section.

What results would one expect? All behavioral models include some form of centrifugal and centripetal forces: the destabilizing tendency of chartist behavior and the stabilizing forces exerted by fundamentalists. In a linear model, one of both would dominate over the other so that the fundamental equilibrium would either be asymptotically stable or explosively unstable. Appropriate nonlinearities could prevent the scenario of complete instability: with changing elasticity of one or both of the two excess demand functions further away from the equilibrium local instability could lead to cyclic or chaotic dynamics rather than explosive movement away from the fundamental equilibrium. The visual appearance of particularly the log mispricing lends support to such an approach. What is, therefore, needed to end up in such a scenario is that chartists' excess demand flattens out for larger values of the current price trend, or that fundamentalists' reaction becomes relatively stronger for larger deviations of the price from its fundamental value. The polynomial terms up to third order provide for these possibilities.

Table 1 and Fig. 2 show the results. In both the raw data and the logarithmic mispricing, we find similar parameter estimates in a linear version of the chartist/fundamentalist model. In both cases, the chartist effect appears more pronounced (as judged by its t-statistics). Adding only a quadratic or cubic term leads to relatively small improvements of the goodness-of-fit, and additional parameters are often insignificant. A larger improvement of the fit is observed if both the quadratic and cubic term are added simultaneously. For the raw data, in the full model all parameters appear to be significant. The resulting functional forms in Fig. 2 are, however, not always as expected. For the chartists, the full model indicates an over-

proportional reaction to larger price changes. For the fundamentalists, we also find a convex shape in the positive quadrant which might indeed be stronger than the convexity of chartists' excess demand. For negative misalignments, excess demand by fundamentalists appears non-monotonic and is hard to interpret. One may note that a cubic term alone without the second-order contribution would deliver 'nice' shapes of both excess demand functions in agreement with the overall predictions of the behavioral models. Unfortunately, without the quadratic term, the improvement compared to the linear benchmark appears negligible. For the log data, the results are different insofar as in the full model only the linear term is significant in chartists' demand and only higher order terms are significant in fundamentalists' demand function. While the latter still has a non-monotonic shape for negative misalignments, far away from the fundamental value, fundamentalists' reaction is on both sides characterized by a strongly convex shape. Together with a linear demand function of chartists and a relatively flat inelastic reaction of the fundamentalists in the vicinity of the fundamental value, these results are in broad support of the scenario of stronger stabilizing forces that become activated at a certain distance from the fundamental equilibrium.

Our approach in this section is very close to that of Majewski et al. (2020) who also estimate nonlinear functions up to cubic order for chartists' and fundamentalists' excess demand functions. While their setting is different (they use a state-space formalization of an extended version of Chiarella's model), the results are very close in their overall support of the relevance of higher-order terms.

However, one difference to their results is that a closer inspection of the dynamics reveals that all estimated models are still locally asymptotically stable in the neighborhood of the fundamental equilibrium $p_t = p_{f,t}$ of the deterministic skeleton of the

models. We also note that despite significance of various behavioral parameters, all the models investigated are characterized by about the same estimate of the error variance, σ_ϵ . Hence, this setting would not explain the bimodality of the S&P500 mispricing. This leaves the possibility that this feature of the data is the result of some more intricate non-linear mechanism that might be covered by some of the models reviewed by Schmitt and Westerhoff.

4 The Stochastic Cusp Catastrophe Model

Christopher Zeeman, one of the founding fathers of mathematical catastrophe theory, has proposed as one of a variety of potential applications of “catastrophic” singularities a model of interacting investors in Zeeman (1974). While the original paper also offers a profound economic intuition for the behavioral interactions that might lead to mispricing followed by catastrophic corrections, in applications of this model researchers have typically abstracted from these behavioural foundations and have estimated a nonlinear stochastic differential equation that forms the dynamic counterpart of the geometric representation of the so-called cusp catastrophe (Baruník and Vošvrda, 2009; Diks and Wang, 2016). The cusp catastrophe can be described by the geometrical features of the function

$$V = \frac{1}{4}x^4 - \frac{1}{2}\beta x^2 - \alpha x \quad (5)$$

in which α and β are control parameters. As a polynomial function in x , eg. (5) will either have three or only one solution. When varying the control parameters α and

β and adding dynamics of the form:

$$\frac{d}{dt}x_t = \frac{-\partial V}{\partial x_t} = -x_t^3 + \beta x_t + \alpha \quad (6)$$

one can identify the extrema of (5) as equilibria of the dynamic system (6), and certain changes of the control parameters will lead to abrupt switches from one stable equilibrium (that ceases to exist) to the other. Performing the parameter variation in reverse one can also observe hysteresis effects, i.e., the system will remain in the second stable equilibrium even for some time when the first one reappears. This dynamic version motivates the estimation of a nonlinear stochastic differential equation with drift term equal to the derivative of the standard form of the cusp catastrophe function:

$$dz_t = (-z_t^3 + \beta z_t + \alpha)dt + \sigma_z dW_t \quad (7)$$

where W_t is standard Brownian motion. Since linear transformations leave the topological features of the model unchanged, one can adopt this normal form to the particular scale of the data by setting $z_t = \frac{y_t - \lambda}{\sigma}$ so that for the data format at hand, y_t , the following stochastic differential equation is obtained:

$$dy_t = \left(-\left(\frac{y_t - \lambda}{\sigma}\right)^3 + \beta \frac{y_t - \lambda}{\sigma} + \alpha\right)dt + \sigma \sigma_z dW_t \quad (8)$$

Our data here is the price distortion, i.e., $y_t = p_t - p_t^*$ when using the original data, or $y_t = \ln(p_t) - \ln(p_t^*)$ when considering the relative distortion.

The potential explanatory power for bimodal price distortions is readily visible from eqs. (7) and (8). If the so-called Cardan's discriminant, $\delta = 27\alpha^2 - 4\beta^3$, is

negative, there will exist three equilibria, with the two stable ones corresponding to states of overvaluation or undervaluation under a bimodal stationary distribution for y_t in eg. (8). If the discriminant is positive only one equilibrium exists for which y_t should obey a uni-modal distribution with its maximum in the vicinity of $p_t = p_t^*$.

We estimate the cusp catastrophe model using a maximum likelihood approach². Previous applications of this model have used the stationary distribution of the stochastic differential equation (8) or Gaussian approximations (Baruník and Vošvrda, 2009; Diks and Wang, 2016). While there is no closed-form solution to the transient density, we can, however, characterize its time evolution by a so-called Fokker-Planck or forward Kolmogorov equation (e.g. Gardiner, 2009). The latter is a partial differential equation that can be solved, for instance, using standard finite difference methods. Poulsen (1999), Lindström (2007) and Lux (2009) demonstrate how this approximation can be used within a maximum likelihood framework. Proceeding in this way, we do not need to use any approximation of the crucial nonlinearity, and with a sufficiently fine mesh in the finite difference equation, we can arrive at estimation results arbitrarily close to analytical maximum likelihood. To find the maximum of the likelihood function, we can apply standard gradient-based methods. Since the nonlinearity of the drift function might lead to multimodality of the likelihood function, the optimization has been run with different starting points, but was found always to converge to roughly the same numbers for the significant parameters.

Table 2 exhibits the results for both the raw data and the logarithmic prices. Unfortunately, this model does not what Schmitt and Westerhoff or the present author himself would have expected it to do in the current framework: In both cases, the

²Since in this section the estimated model is a well-behaved diffusion process, the usual asymptotics should apply.

slope parameters α and β are insignificant. While Cardans's discriminant δ is negative for the raw data, Brownian motion even leads to a better fit than the cusp model³. The model actually does not appear to be able to make much sense of the data as all parameters turn out to be insignificant. As Fig. 3 shows, the estimated cusp model would remain in one of its two modes over the entire horizon of the data, and the fit is apparently worse than that of a simple Brownian motion.

For the log data, there is at least the expected hierarchy of goodness-of-fit with the cusp model achieving a higher likelihood than a model with $\alpha = \beta = 0$ and the latter again superior to Brownian motion. However, neither are the differences in the likelihoods significant nor are the estimates of α and β . Compared to the standard deviation $\sigma_z \cdot \sigma$ parameters α and β have no visible relevance for the resulting dynamics (cf. Fig. 3). We, thus, see that not every model with the potential of generating a bi-modal distribution provides a good fit to the bi-modal distortion of the S&P500. The close proximity of both the raw and log distortion to Brownian motion also indicates that there is a high noise-to-signal ratio in the data set which will make it hard for any behavioural model to pick up the remaining systematic patterns in the data (if any such non-linear factors exist).

5 Simple Nonlinear Behavioral Models

The literature on behavioral asset pricing models started to become a popular area around the early 90s. The seminal contributions of this time have all been formalized as nonlinear dynamic models of the interaction of chartist and fundamentalist investors. An early inspiration has been the paper by Beja and Goldman (1980).

³The Brownian motion benchmark is not nested in the cusp model since even for $\alpha = \beta = \lambda = 0$ the cubic term would be preserved.

Since the latter had formulated only linear excess demand and price formation rules, similar to the linear version of the model of sec. 3, it could, however, only give rise to either a stable fundamental equilibrium or unstable, explosive dynamics. Whether one or the other prevails is mainly determined by the relative strength of stabilizing fundamental arbitrage versus destabilizing technical trading. Neither convergence to some fundamental equilibrium nor everlasting speculative bubbles appear a satisfactory description of the volatile dynamics of many asset markets.

The contributions by Day and Huang (1990), Chiarella (1992) and De Grauwe et al. (1993) have all introduced some specific form of nonlinearity, which allows these authors to generate intermediate forms of dynamic evolutions such as periodic cycles or chaotic dynamics. As a consequence, alternating periods of overvaluation and undervaluation became possible.

In order to achieve this outcome different forms of nonlinear components are assumed by these authors: Day and Huang (1990) assume that fundamentalists' excess demand becomes increasingly elastic with larger deviations of the price from its fundamental value; Chiarella (1992), in contrast, assumes that chartists' demand is characterized by a decreasing elasticity in its reaction to observed momentum (price changes), while De Grauwe, Dewachter and Embrecht (1993) assume that chartists' and fundamentalists' weights in aggregate excess demand change in reaction to the actual degree of mispricing. Their model is embedded in an monetary framework of the foreign exchange market in which fundamentalists and chartists are advisors to foreign exchange dealers who combine the recommendations of these two groups of analysts using time varying weights. In the present context, the model of De Grauwe et al. (1993) has been adapted to a stock market framework and aligned with the formalizations chosen by Day and Huang (1990) and Chiarella (1992).

In more detail, the three models are formalized as follows:

As in sec. 3, all three models assume that prices follow a structural adjustment mechanism in that a market maker increases or decreases prices in response to excess demand:

$$p_{t+1} = p_t + \beta(ED_t) \quad (9)$$

where β denotes the price adjustment speed. Excess demand is composed of the demand and supply or fundamentalists ($ED_{f,t}$) and chartists ($ED_{c,t}$). In Day and Huang (1990), these assume the following forms:

$$ED_{f,t}^{DH} = a (p_{f,t} - p_t) f(p_t - p_{f,t}) \quad , \quad (10)$$

$$f(p_t - p_{f,t}) = \frac{1}{(p_t - p_{f,t} - m_1)^{d_1} (m_2 - p_t - p_{f,t})^{d_2}}, \quad (11)$$

$$ED_c^{DH} = b (p_t - p_{f,t}). \quad (12)$$

This formalization guarantees that fundamentalists would build up relatively higher counterpositions for higher levels of mispricing and given the parametrization of the model, would not allow prices to exceed from the interval $(p_{f,t} - m_1, p_{f,t} + m_2)$ since at these boundaries the fundamentalists' demand would become infinitely elastic. Chartists here are defined in a slightly different way than in other models as their excess demand is based on mispricing rather than momentum. This model has altogether six parameters that need to be estimated: a, b, m_1, m_2, d_1 and d_2 . The price adjustment speed cannot be estimated independently of the elasticities a and b , and hence it is set equal to unity.

Chiarella's model has linear fundamentalists and chartists whose excess demand

is driven by momentum with decreasing elasticity if momentum assumes very high positive or negative values:

$$ED_{f,t}^C = a (p_{f,t} - p_t), \quad (13)$$

$$ED_{c,t}^C = b \tanh(c(p_t - p_{t-1})). \quad (14)$$

This model has also been estimated by Majewski et al. (2020) who consider an extended version with nonlinearities in both the chartists' and fundamentalists' excess demand functions, and the fundamental value being treated as a hidden variable in a state space framework.

Finally, De Grauwe et al. (1993) allow for time-varying fractions of chartists and fundamentalists, w_t and $1 - w_t$, respectively. Excess demand of their fundamentalists is linear while chartists' trading signal is the change of momentum, i.e. one-period compared to two-period price changes:

$$ED_{f,t}^G = a (p_{f,t} - p_t) w_t, \quad (15)$$

$$ED_{c,t}^G = b((p_{t-1} - p_{t-2}) - (p_{t-1} - p_{t-3})) (1 - w_t). \quad (16)$$

The weight or fraction of chartists is determined by the function

$$w_t = \frac{1}{1 + d(p_{f,t-1} - p_{t-1})^2}. \quad (17)$$

Chartists will, therefore, dominate the market in the vicinity of the fundamental equilibrium while fundamentalists will have a higher influence in the presence of a large deviation of the market price from its fundamental value. Since the model of

De Grauwe et al. is originally formulated as a model of speculative activity in the foreign exchange market, its price formation mechanism is derived from a monetary macroeconomic model. In the present context, we replace this macroeconomic component by eq. (9). In all three models we again suppress the variable β as it always enters multiplicatively with the slope coefficients of the excess demand functions. The parameter set of the model of Chiarella is $\theta = \{a, b, c\}$ while for the model of De Grauwe et al. (1993) it is $\theta = \{a, b, d\}$. Although we have used the same symbols for the coefficients of excess demand components in all three models, it should be noted that their estimates are not directly comparable as they are based on different implementations of the behavior of these groups of traders.

All three models are estimated both for the raw data as well as for logarithmic prices and fundamentals. For the noise term, we either assume that it enters in excess demand: $ED_t = ED_{f,t} + ED_{c,t} + \epsilon_t$, due to an additional noise trading or microstructural component of demand, or we alternatively assume that the fundamental value used by the agents consists of: $p_{f,t} = p_t^* + \epsilon_t$, i.e. is obtained as the ex-post rational fundamental value plus the prediction error of the sum of future discounted dividends. This corresponds to the framework of the excess volatility tests introduced by Shiller (1981), and is economically plausible since the currently best computation of the fundamental value is based on limited information about future realisations of dividends and ex-post the expectation errors should enter as uncorrelated random variates. In the first version, in contrast, we assume implicitly fundamentalist agents have perfect foresight, i.e. $p_{f,t} = p_t^*$ which, of course, is an extreme and unrealistic scenario. We will refer to these two alternative scenarios as versions I and II of any model under scrutiny. The models are all not nested in the formalization of sec. 3 with polynomial terms up to third order, and so estimation of these alternative formalizations allows

us to investigate whether alternative nonlinear forms fit the data better.

Under a fully rational asset pricing framework, the behavioral components should not enter as significant explanatory factors. Note that the distinction between the two alternatives of noise in excess demand and noise in fundamental values only leads to different results in the models of Day and Huang (1990) and De Grauwe et al. (1993) whereas in Chiarella (1992) both alternatives would be observationally equivalent due to the linear formalization of excess demand of fundamentalists. All models will be estimated via maximum likelihood using the Nelder-Mead or simplex algorithm to find the maxima of the likelihood function. The Nelder-Mead algorithm still counts very much as the industry standard for estimating objective functions with non-continuous derivatives and possible multiple local minima or maxima. Indeed, gradient-based optimization routines did mostly worse than the Nelder-Mead algorithm although one would expect the likelihood function to be continuous over most of its support for our problems. Reported results are the maxima of the likelihood function obtained after hundreds of runs with varying initial conditions.

Table 3 displays results for both the raw and logarithmic data. One observes that with one exception the best fitting configurations come with likelihood values that are in the vicinity of those of the benchmark models of sec. 3. We also see that in many cases the behavioral parameters are not significant, and for the type I models, the estimates of the standard deviation of the noise component are very close to the standard deviation of the data. The later observation also suggests that the behavioral components of the models do not contribute much explanatory power.

Starting with the model of Day and Huang (1990), we find that most parameters are insignificant and the maximized likelihood is smaller than that of a linear chartists/fundamentalists model. Indeed, when inspecting the details of the outcomes

of the optimization we find parameter values displaying wide variations that all lead to basically the same likelihood. When plotting the excess demand functions of the best fitting parameter sets, we find large differences in the fundamentalists' excess demand, but when aggregating it with excess demand by chartists, practically the same shape results for all of them. As can be seen in Fig. 4 the resulting excess demand appears also almost linear over most of its support with a sharp bend at the boundaries. Given their insignificant t-statistics it seems that the four parameters for the nonlinear behavior of the fundamentalists appear superfluous.

The estimation of the model by Chiarella (1992) shows a more pronounced nonlinear shape as also illustrated in Fig. 4. While in this case the parameters are all highly significant, the model improves only marginally upon a linear model with the same groups of agents. The type I version of De Grauwe et al. (1993) is the worst in this category. Here, only σ_ϵ enters significantly, the parameters for the elasticity of the excess demand functions assume counterintuitive signs, and the likelihood is only marginally higher than that of a pure random walk. When inspecting the details of this model, it shows one particular feature that as we will see later is also shared by some more elaborate models with switching of strategies: Most of the time, we find a dominance of one group (here the chartists) while the other group (here the fundamentalists) only occasionally assumes a non-zero weight, cf. Fig. 4.

In the category of type II models, we find the pertinent version of De Grauwe et al. performing similarly like its type I version. The type II version of Day and Huang, however, performs better than all the other alternatives and also beats the cubic model of sec. 3. It appears on a first view hard to explain its more successful performance as again most parameters are insignificant. Inspecting the predictions of the price of this model vis-à-vis the four alternatives also provides little evidence of

striking differences. When computing the differences in the log likelihoods in Fig. 5, one observes that the advantage of Day and Huang rests on a better fit in the volatile periods between time 1500 and 1600 (1990 to 2004) which weights higher than some subsequent negative deviations from the other models (most pronounced in period 1650 which is 10/2008). Fig. 5 restricts the comparison to the last 200 observations of the sample since in the remainder of the series one finds almost no discernible difference between the models.

It seems that the responsible factor for the better performance of Day and Huang II is the much higher estimate of the parameters d_2 and m_2 and the lower estimate of a that leads to much less curvature in the excess demand function of the second version of the Day and Huang model. Predicting smaller price slumps in the booming years 1996 through 2004 than its competitors, it on average performs better. The relatively large number of better fits in this period also dominates over a large negative outlier in 2008 when Day and Huang II would have predicted a much smaller crash than the other models. The main drawback of the best-fitting model Day and Huang II is, however, that it comes with a six-digit standard deviation of the errors of agents' computation of the fundamental value. Taken literally, this implies that agents' expectations could deviate by several orders of magnitude from the ex-post rational value, i.e. they were basically ignorant of what a sensible range of fundamental values might be.

The remainder of Table 4 displays results for models that use the log data rather than the raw entries. Starting with the last model in our list, we again find that both versions of De Grauwe et al's formalization perform hardly better than a pure random walk. In the first version this comes despite the fact that its group dynamics shows interesting variations between the weights attached to either chartists or fundamen-

talists. As one would expect, the estimation assigns low weights to fundamentalism in times with large distortion and higher weights when the price is closer to the ex-post rational value.

However, due to the large discrepancy in the parameters a and b , the fundamentalist component of excess demand is much smaller than the one contributed by chartists. Both are dominated by the noise component of excess demand with a standard deviation approximately equal to that of the log price changes in the data. The lack of explanatory power of the behavioral components is corroborated by the lack of significance of all parameters but σ_ϵ . The same applies to De Grauwe II for which again we find a dominating weight of unity for fundamentalists just as for the case of the raw data.

The Day and Huang model is somewhat more successful although we again find hardly any significant parameter estimate (except for m_1, m_2 and σ in version I). Fig. 6 shows the nonlinearity for both versions. Overall, both alternatives perform worse than a benchmark linear model with chartists and fundamentalists. The best behavioral model is here the parsimonious parametrization of Chiarella (1992) that slightly improves upon the linear benchmark. However, its nonlinear parameter c is not significant, and as we can see from the first panel of Fig. 6, its nonlinearity is indeed irrelevant over the support of the excess demand function of chartists in our data.

6 Models with Discrete Choice of Strategies

Among the best-known behavioural asset-pricing models are those that model the selection of strategies by market participants as a discrete choice problem. This

literature got started with Brock and Hommes (1998) who consider the problem of choosing between a completely rational forecast that comes at some cost for the agents, and a costless adaptive alternative. While Brock and Hommes (1998) apply this framework to a cobweb-type model, very similar approaches have been developed in subsequent literature for asset markets. Schmitt and Westerhoff (2017) demonstrate that the bi-modality of the S&P500 distortions is shared by both the baseline Brock and Hommes (1997) model, appropriately adopted to the problem of forecasting an asset price, and by the slightly different version of Gaunersdorfer and Hommes (2007) that is geared towards speculative price dynamics in asset markets from the outset. These models are somewhat more complex than the previous ones. In particular, their starting point is a traditional asset pricing framework in which agents' demand is derived from a mean-variance utility function leading to excess demand

$$ED_{i,t} = (E_{i,t}[x_{t+1}] - (1+r)x_t)/\mu \quad (18)$$

with $x_{t+1} = p_t - p_{f,t}$ the price distortion, r the risk free interest rate, μ the product of the risk aversion coefficient and the expected volatility of the asset price (assumed to be constant), and i an index for the type of investor. Again, two groups with different expectation formation processes are assumed to exist which we again denote as fundamentalists and chartists. Fundamentalists expect the price to be equal to its fundamental value, hence $E_{f,t}[x_{t+1}] = 0$. Chartists are assumed to form extrapolative expectations: $E_{c,t}[x_{t+1}] = ax_{t-1}$. The discrete choice component enters by agents comparing the profitability of both strategies. Profits by both groups are given by:

$$\pi_{i,t} = (x_t - (1+r)x_{t-1})ED_{i,t-1} - d_i \quad (19)$$

where d_i is the cost of forming the expectation which is $d_i = d > 0$ for $i = f$ and $d_i = 0$ for $i = c$. The fraction of chartists is consequently determined as:

$$n_{c,t} = \frac{\exp(\gamma\pi_{c,t-1})}{\exp(\gamma\pi_{c,t-1}) + \exp(\gamma\pi_{f,t-1})} \quad (20)$$

with γ the so-called intensity of choice and $n_{f,t} = 1 - n_{c,t}$. Under an equilibrium asset-pricing framework with zero supply of shares, the asset-price is determined as

$$p_t = \frac{1}{1+r} (n_{c,t}E_{c,t}[x_{t+1}] + n_{f,t}E_{f,t}[x_{t+1}]) \quad (21)$$

where the second entry in the bracket on the right-hand side would be zero by definition in the baseline model. We again distinguish between two versions of this framework. Version I adds a normally distributed error term additionally on the right-hand side, while version II assumes again an error term in the perception of the true fundamental value by the fundamentalists so that $E_{f,t} = \epsilon_t$ and $p_t = \frac{1}{1+r} (n_c E_{c,t}[x_{t+1}] + n_{f,t} \epsilon_t)$ results.

Turning to the model of Gaunersdorfer and Hommes (2007), their demand functions are slightly more complex:

$$ED_{i,t} = (E_{i,t}[p_{t+1} + d_{t+1}] - (1+r)p_t)/\mu \quad (22)$$

with:

$$E_{f,t}[p_{t+1}] = p_{f,t} + v(p_t - p_{f,t}), \quad (23)$$

$$E_{c,t}[p_{t+1}] = p_{t-1} + g(p_{t-1} - p_{t-2}). \quad (24)$$

Profits are derived similarly as above, but Gaunersdorfer and Hommes take into

account also dividends, d_t , and assume that agents consider not only the most recent realization of profits but rather the entire history with exponential discounting of older observations:

$$\pi_{i,t} = (p_t + d_t - (1 + r)p_{t-1})ED_{i,t-1} + \eta\pi_{i,t-1} \quad (25)$$

Here η is the discount parameter for older observations. Fractions of both groups are, then, determined as in Brock and Hommes with an additional factor that prevents explosive instability by restricting the fraction of chartists in the presence of large deviations of the market price from its assumed fundamental value:

$$\pi_{c,t} = \frac{\exp(\gamma\pi_{c,t-1})}{\exp(\gamma\pi_{c,t-1}) + \exp(\gamma\pi_{f,t-1})} \exp\left(-\frac{(p_{t-1} - p_{f,t-1})^2}{\alpha}\right) \quad (26)$$

The equilibrium price is determined as:

$$p_t = \frac{1}{1 + r}(n_{c,t}E_{c,t}[p_{t+1}] + n_{f,t}E_{f,t}[p_{t+1}] + E[d_{t+1}]) \quad (27)$$

It is assumed that both groups of agents form expectations about dividends in the same way. Version I and II of this model are defined as for the model of Brock and Hommes.

These models requires some additional data as input: The risk-free rate r and expected dividends. For the former the rate of 7.6 percent p.a. is used as in Shiller's computation of the ex-post rational price. For expected dividends it has been assumed that these expectations are formed completely rationally (i.e., the next period's realization has been substituted for it) or that agents use the historical average of their realizations. While the later alternative in certainly assuming much less rationality

on the part of our agents, it provides better estimates so that this version is reported.

It is not obvious how to reformulate these models in terms of the log of prices. Mean-variance utility functions and profits would, of course, not preserve their intended meaning if prices would be replaced by their logarithms. The same applies to the equilibrium condition of eqs. (21) and (27).

In order to arrive at a log-version of the model, we, therefore, replace the latter equations by a market maker adjustment in the presence of excess demand:

$$\ln(p_{t+1}) = \ln(p_t) + \beta(ED_{c,t} + ED_{f,t}) \quad (28)$$

This adds the new parameter β (analogously to the models of the previous section). The excess demand functions in the right-hand, of course, still depend on prices so that all the other functions appearing in these model are preserved. Since β and μ cannot be estimated independently, we set the latter equal to 1. The third example of a model with discrete choice component is the one of Franke and Westerhoff (2012). These authors use very basic forms of the excess demand functions of the two groups but assume that these functions come with a group-specific noise factor:

$$ED_{f,t} = (a(p_{f,t} - p_t) + \epsilon_t^f)n_{f,t} \quad (29)$$

$$ED_{c,t} = (b(p_t - p_{t-1}) + \epsilon_t^c)n_{c,t} \quad (30)$$

Prices are determined via a market maker mechanism:

$$p_{t+1} = p_t + \beta(ED_{f,t} + ED_{c,t}) \quad (31)$$

and group occupation numbers follow a discrete choice setting so that

$$n_{c,t} = \frac{\exp(\gamma U_{t-1}^c)}{\exp(\gamma U_{t-1}^f) + \exp(\gamma U_{t-1}^c)} = \frac{1}{\exp[\gamma(U_{t-1}^f - U_{t-1}^c)]} \quad (32)$$

and $n_{f,t} = 1 - n_{c,t}$.

The difference in the “utilities” entering the discrete choice formula is compactly expressed as:

$$U_t^f - U_t^c = \alpha_0 + \alpha_n(n_t^f - n_t^c) + \alpha_p(p_t - p_{f,t})^2 \quad (33)$$

The choice of strategies is, thus, determined by (i) a bias α_0 which depending on its sign might express a general tendency in a certain market toward one of the two strategies, (ii) a herd effect favoring the selection of the most popular strategy, and (iii) a valuation effect similar to the one in De Grauwe et al. (1993) that favors the selection of a fundamentalist strategy if the actual market price is far away from the perceived fundamental value.

Estimation of the model of Franke and Westerhoff is not as straightforward as for the models considered before. In particular this model features two noise factors, ϵ_t^f and ϵ_t^c while all previous models had assumed only one source of noise. Following Lux (2018), we can interpret the model of Franke and Westerhoff as a dynamic model with latent states (cf. Douc et al., 2013, 2015). The observed variable in our framework is, of course, the market price, p_t , while overall excess demand including the fractions of both groups of investors would count as the unobservable, latent part. We can, then, assign the two noise factors, ϵ_t^f and ϵ_t^c , one of two roles: Either being part of the latent dynamics, or being observation noise that distorts the link from latent variables to the

observable variable. Since the assignment of the role to both noise factors is arbitrary in the present setting, we have pursued both alternatives, and found the results to coincide. In order to maximize numerically the likelihood function, a particle filter has been used to approximate the latent dynamics. In the case of the Franke and Westerhoff model, prices can either be set as the raw numbers or their logarithms since no explicit utility-maximization is assumed. The version I of this model is simply the model as it has been laid out above, while version II replaces the demand by fundamentalists by: $ED_{f,t} = a(p_{f,t} + \epsilon_f - p_t) \eta_{f,t}$.

Finally, the last model we estimate is the one proposed in Lux (1995). This approach splits chartists into optimistic and pessimistic ones (equivalent to buyers and sellers) and assumes transition rates for switches between both groups governed by a herding component and observed price changes. In detail, this model works as follows: Excess demand is given by:

$$ED_t = a(p_{f,t} - p_t) + bx_t \quad (34)$$

with the first part on the right-hand side representing the usual fundamental component, and the second part, bx_t , the chartist component depending on the relative difference between the numbers of optimists and pessimists, $n_{+,t}$ and $n_{-,t}$:

$$x = \frac{n_{+,t} - n_{-,t}}{n_{+,t} + n_{-,t}}. \quad (35)$$

In equation (35), stochasticity is implicit through the assumption of a finite population of chartists which we set to $N = n_{+,t} + n_{-,t} = 100$. Switches between both groups are determined by functions similar to discrete choice probabilities with the difference

that here they pertain to the decision to *change one's previous choice* and to join the other group. We, therefore, have to distinguish between switches from pessimistic to optimistic opinion, p_{+-} , and vice versa, p_{-+} given by:

$$p_{+-} = v \exp \left(\alpha_0 + \alpha_1 x_t + \alpha_2 \frac{dp}{dt} \right), \quad (36)$$

$$p_{-+} = v \exp \left(-\alpha_0 - \alpha_1 x_t - \alpha_2 \frac{dp}{dt} \right). \quad (37)$$

The implication of this formalization is that a positive price trend reinforces the formation of an optimistic herd and vice versa.

The opinion index x_t changes under these transitions according to the following law of motion:

$$\frac{dx_t}{dt} = \frac{1}{N} (p_{+-}n_{-,t} - p_{-+}n_{+,t}). \quad (38)$$

Price changes are formulated in continuous time, i.e.

$$dp_t = \beta ED_t. \quad (39)$$

This model can, like the one by Franke and Westerhoff, be interpreted as a nonlinear state space model with the hidden variable x_t and the observable variable p_t and will be estimated by numerical maximum likelihood using the particle filter. Because of the linear excess demand of the fundamentalists, there is only one version of this model to be estimated.

Table 4 shows the results. For the raw data, we find three out of six models that perform better than a linear model of chartists and fundamentalists. Two of

these have a performance comparable to Day and Huang II, while one beats all other models considered so far by a sizable margin (i.e. Franke and Westerhoff I).

Starting with the model by Brock and Hommes, we find that its version I has no significant parameters except for the standard deviation of innovations in excess demand. With the intensity of choice close to zero, the estimated fractions of both types of agents stay practically constant for the entire sample at 50 percent (cf. Fig. 7). With constant fractions the dynamics effectively boils down to an autoregressive process and the likelihood is approximately the same like that of such a simple process.

Brock and Hommes II also has an estimate of γ close to zero but achieves a somewhat better fit. Despite the small intensity of choice parameter it predicts nevertheless somewhat more variation of the fractions of the two strategies towards the end of the sample horizon. Until period 1500 the fraction of chartists is practically constant and equal to unity which is mainly caused by an extremely large estimate of the cost d of an accurate prediction of future prices. However, in the last years of the sample, the distortion becomes so high in absolute value that occasionally the fundamentalist strategy becomes preferable. The predicted reversal of prices towards their fundamental value helps to improve the overall likelihood.

The two versions of the Gaunersdorfer and Hommes model both perform worse than any benchmark model. Version I reports many significant parameters. However, the estimated error variance is almost the same as the variance of the data so that the contribution of the behavioral terms appears indeed negligible. One particular feature is that with a very large estimate of α , the built-in tendency towards a fundamentalist dominance in the presence of large distortions is effectively invalidated. With $\eta > 1$ we also have a cumulative effect of differences in profits⁴ that apparently leads to an

⁴The qualitative results remain the same if η is restricted to the interval $[0, 1]$ in the estimation.

increasing dominance of chartists over time. However, the extrapolative forecasts of chartists lead to a fit that is worse than that of a linear adaptive chartist rule.

Gaunersdorfer and Hommes II performs even worse. In the point estimates we find again a very small β and a counterintuitive sign of the chartist extrapolation parameter g with t -statistics all insignificant. The estimated choice of strategies shows a dominance of fundamentalism with only intermittent switches to chartism.

In contrast, the two versions of the model by Franke and Westerhoff both perform better than the linear benchmarks. As can be seen from Fig. 7, these are the first models that exhibit the type of strategy switching behavior that one would intuitively expect: A high fraction of chartists dominates phases of overvaluation or undervaluation. Inspecting the estimates for Franke/Westerhoff I, it appears, however, that the determinants of the switching function are all not significant. The same holds for γ , but the systematic variation of both fractions nevertheless indicates that there should be some explanatory value of this component. With $\alpha_0 = 1.99$ and $\alpha_n \approx -3.5$, the behavioral model mainly indicates an overall dominance of fundamentalism together with a tendency of antagonistic behavior away from the dominating strategy. Both the slope coefficients of the excess demand functions and the two standard deviations, σ_c and σ_f , are highly significant so that here indeed we find for the first time a model that benefits from its nonlinear structural components and cannot be reduced to a simpler linear framework. The estimate of the two error variances indicates a dominance of chartists as the group responsible for most of the volatility of the market.

Somewhat similar results are obtained for Franke and Westerhoff II. Despite the very small parameters for the switching between strategies, the predicted behavior of the fractions of both groups is very similar to the first version. However, there is an

important difference: When estimating the fractions of chartists and fundamentalists using the particle filter, two different possibilities can be observed: An outcome close to the one of Franke and Westerhoff I and another one in which the two groups reverse their roles (the blue and red lines in Fig. 7). This is probably due to the negative estimate of the fundamentalists' reaction parameter which effectively makes the behavior of both groups similar.

The model of Lux (1995) exhibits significant parameters estimates, and performs comparably to the linear model of sec. 3. Filtered trajectories of the hidden variable x_t show, however, only unsystematic fluctuations of chartists' sentiment (not shown in the figure).

Fig. 8 shows the performance in terms of the likelihood values of Franke/Westerhoff I against its main competitors during the decisive last 200 observations of the time series. The pattern is similar as for Day and Huang II in Fig. 5 but the gains during the years 1996 to 2004 are more pronounced and the losses during the onset of the financial crisis are smaller.

Results for the relative distortion (i.e., logarithmic data) show very similar overall tendencies. Brock and Hommes I has no significant parameter except for σ_ϵ and achieves a likelihood close to that of the random walk. Inspection of the group dynamics shows that the fractions of chartists and fundamentalists in this version exhibit permanent high frequency changes between zero and unity (not shown in Fig. 9).

Brock and Hommes II shows somewhat better performance despite similarly insignificant parameters, but it remains below the linear benchmarks. In this case, the fractions exhibit variations mostly in the range $n_c \in [0, 0.5]$ that become more pronounced over time.

The two versions of the model by Gaunersdorfer and Hommes also do not improve much over the random walk. In both cases all parameters appear to be insignificant. In version I, the coefficient of fundamentalists is negative implying they expect an increase of any observed mispricing in the future. Nevertheless, the estimates indicate that fundamentalists would completely dominate the market for almost the entire sample (with only a single event of intermittent existence of about 40 percent of chartists around time 800). The overall dominance of fundamentalists is shared by the second version of the model albeit with more frequent short bursts of chartist activity. The model of Lux (1995) appears again second best, but despite all parameters appearing significant, it again only predicts high-frequency fluctuations of its hidden variable, the opinion index x_t .

Like with the raw data, the one model that beats all benchmarks is Franke and Westerhoff I. Nevertheless, it does not deliver any significant parameter estimates at a traditional confidence level of, say, 5 percent. However, the population dynamics superimposed on the linear demand functions of the two groups of agents performs better by some margin than the linear benchmark. In terms of population dynamics, clear cycles can be observed (cf. Fig. 9) but they come with smaller variations of the fractions of both groups than for the same model estimated for the raw data. In contrast, Frank and Westerhoff II falls back behind the linear model despite many significant parameters. Interestingly, the first version of the model sees most of the fluctuations originating from the chartist side of the market, while the second attributes the lion share to fundamentalists' errors in their computation of the fundamental value.

It is also worthwhile to note that in the case of the models of Franke and Westerhoff, and Lux, the estimation is a stochastic process so the outcome for these models

shows variations depending on the initialization of the particle filter in the numerical likelihood approach. For these models we also indicate the variability of the maximized likelihood by its standard deviation in Table 4. Somewhat surprisingly, in version I of Franke and Westerhoff the estimated likelihood and the filtered fractions of both groups show practically no variation across different values, while in version II they do.

Fig. 11 shows likelihood differences of the best model vis-à-vis the others for the log data. The differences are distinctly different from those obtained with the raw data. Here it is basically one single episode (in 1932 when S&P 500 reached its lowest values in the onset of the great depression) that provides Franke/Westerhoff with its advantage.

7 Model Comparison

As the last step of our analysis, we conduct a simultaneous model comparison of all models estimated so far, including also the linear to cubic specifications of sec.3, and - as the simplest benchmark- the random walk. As a state-of-the art tool in the comparison of multiple, non-nested models we apply the model confidence set approach of Hansen et al. (2011). This approach takes as its input the relative performance or loss of a set of candidate models and seeks to determine the subset of superior models that are not outperformed by other competitors. The hypothesis tested sequentially is that a model performs equally well like the average of all (remaining) models or equally like another selected model. Upon one-sided rejection against a particular model, the latter is eliminated, and the sequence of tests is continued without it. The sequence ends if no more model is eliminated in this way.

A proper size of the test under a chosen level of confidence is established by bootstrapping the variance of the test statistics. Since two versions of the test exist based either on comparisons with the average or on pairwise comparison, we report results of both alternatives in Table 5. We show the outcome for levels of confidence (its complement being denoted the coverage probability) of 95, 90 and 80 percent. Choosing a higher level of confidence of the test means that one imposes a lower error probability when eliminating models from the confidence set. Hence, the model confidence set tends to be larger when insisting on a higher confidence level. As can be seen, for the raw data, results are uniform under both versions of the test and for all levels of significance: The only member of the model confidence set is Franke and Westerhoff I. For the logarithmic misalignments, there is more variation. Under the comparison with the average seven models and all the specifications from sec. 3 remain in the model confidence set. Note that the random walk is not included indicating that we can be confident that there is some structure, but since also the linear model is not eliminated, we cannot be sufficiently confident in the prevalence of nonlinear structure. However, under a sequence of pair-wise tests a confidence level of 90 percent or lower favors again Franke/Westerhoff I while at the 95 percent level six models including the linear to cubic specifications from sec. 3 are not eliminated.

8 Conclusion

In this paper a large selection of representative models from the behavioral literature on asset pricing with heterogeneous expectations has been estimated and compared. Typically, the models under scrutiny lead to complex nonlinear dynamic processes characterizing the asset price valuation. While no explicit results on the asymptotic

consistency and efficiency of maximum likelihood are available for any of the models as they stand, recent results in the mathematical literature suggest that maximum likelihood is more generally applicable (i.e, consistent) for nonlinear and chaotic processes than previously thought.

This broad estimation exercise was motivated by recent findings of the misalignment of the S&P 500 index in comparison with Robert Shiller's ex-post fundamental value. While all models considered here can potentially explain such bi-modality, at their estimated parameters they often display different behavior and contribute relatively little to the explanation of this new stylized fact. The major exception from this rather disappointing finding is the model proposed by Franke and Westerhoff (2012).

Estimating the model confidence set following the methodology of Hansen et al. (2011) provides the following insights:

- (i) in all settings, a simple random walk can be excluded from the model confidence set. Hence, we can safely state that at the chosen significance level the price distortion of the S&P 500 contains some non-random factors that call for at least a linear, if not non-linear structural model for its explanation. A linear model of chartist/fundamentalist interaction as a candidate explanation or certain nonlinear extensions of such a simple benchmark always survive in the model confidence set.
- (ii) if we consider the raw data, a much sharper result is obtained. In this case only the Franke/Westerhoff model with noise in the excess demand function remains in the model confidence set. Hence, we can say that it dominates all other models at the chosen level of confidence. As a consequence, we would then have

convinced ourselves that the non-random element of mispricing cannot be fully captured by a linear model.

- (iii) For the log distortion, the results are more muted and depend both on the level of statistical confidence and the design of the process of estimation of the model confidence set. With pair-wise comparisons, at all levels except the highest level of significance, we again find the Franke/Westerhoff model to be the only member of the model confidence set. However, when moving from the 90 percent to the 95 percent level of significance, a large number of additional models remain in the model confidence set, including the linear benchmark. In group comparisons a similarly large set is obtained at all levels of confidence always also including the linear model. Hence, there is a weaker tendency here towards superiority of Franke and Westerhoff.

In the large sets, only few models from our list are missing: The ones that are always eliminated from the model confidence set are the catastrophe model of Zeeman (1974) and the model by De Grauwe et al. (1993).

All in all, these results appear encouraging and should provide some incentive to estimate and assess the explanatory power of some more candidates from the large zoo of behavioral agent-based model of financial markets. As these models have various building blocks that can be combined in many different ways, there might still be alternatives that could more convincingly explain the bi-modality of price distortions. One advantage that likely contributes to the relatively better performance of the Franke/Westerhoff model is that it has two sources of noise and, hence, is closer to a stochastic volatility model than its competitors. However, since it would boil down to a random walk in the absence of two heterogeneous groups of traders, the baseline

chartist/fundamentalist dichotomy also seems an indispensable element of a successful model.

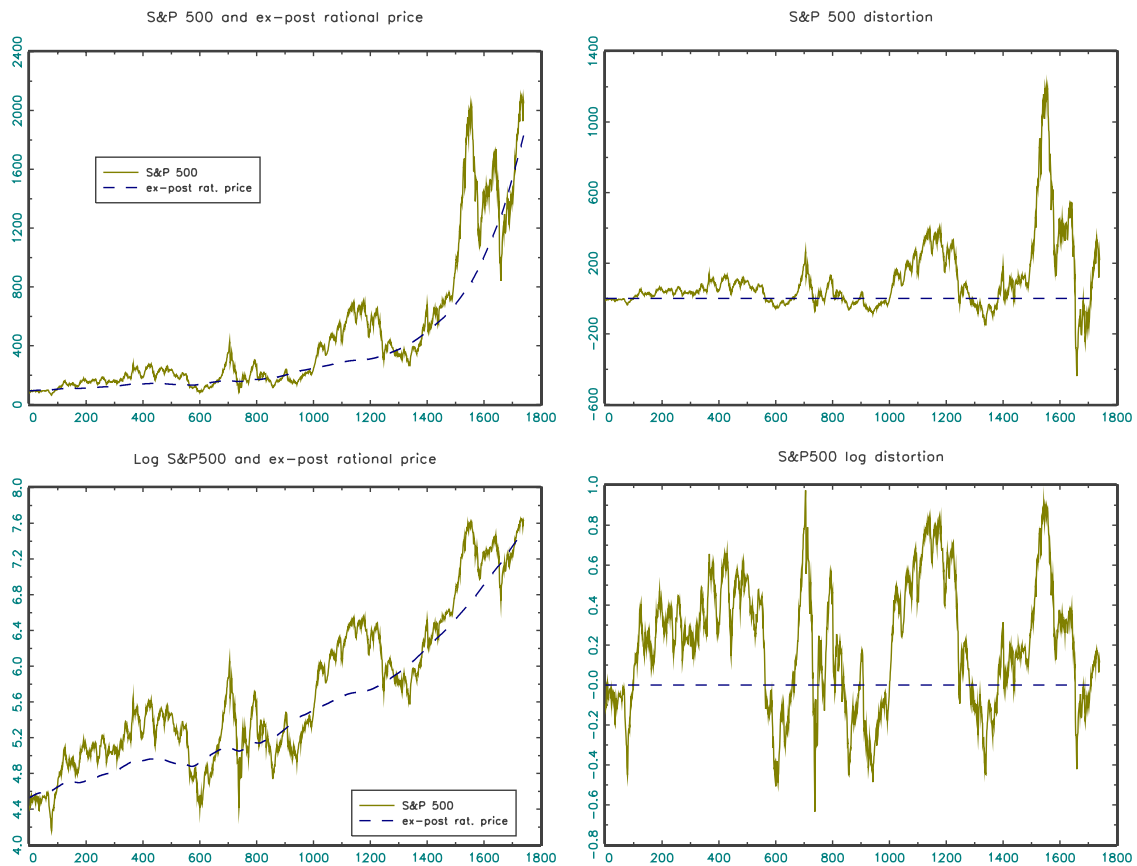


Figure 1: S&P500 and ex-post rational price (left-hand side) and its distortion (difference between market index and ex-post rational price, right-hand side). The upper panels show the original data, the lower panel depicts the log S&P500 versus the log ex-post rational price. Shiller's data set used here covers monthly records from 1871 through 2015.

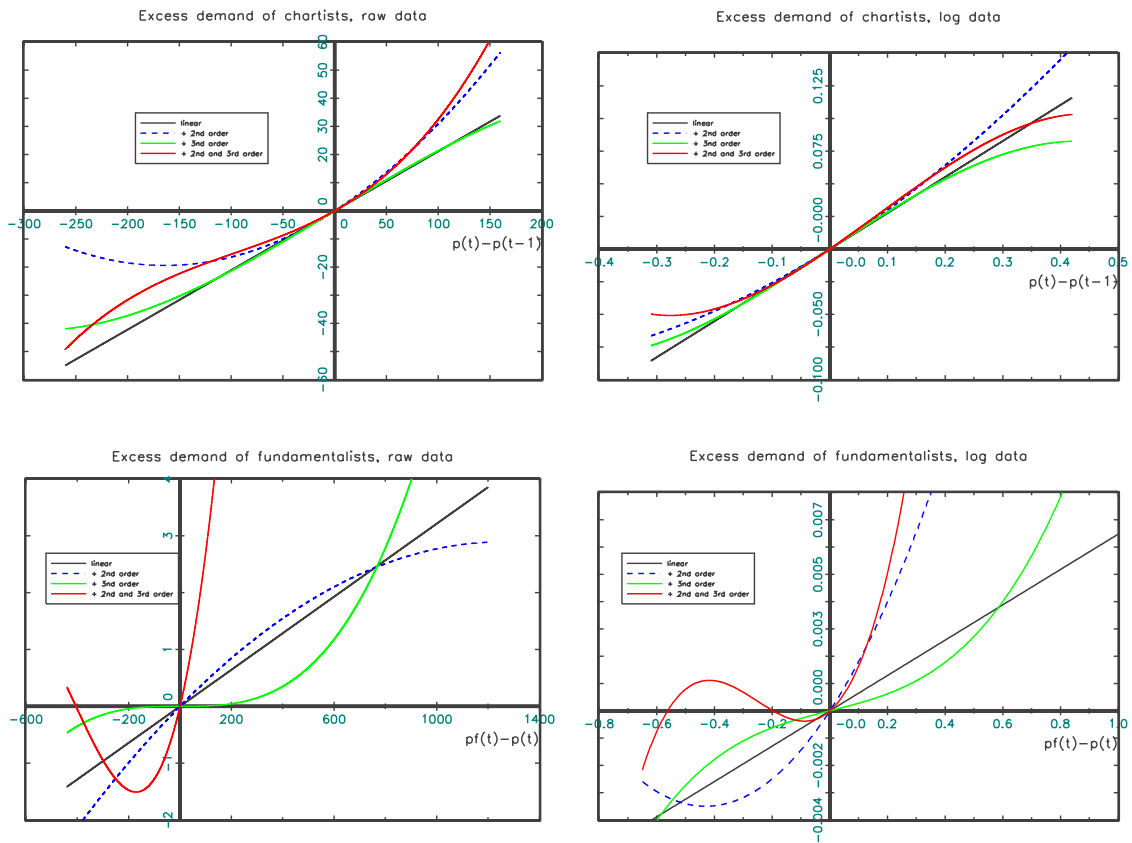


Figure 2: Shapes of the polynomial model estimates for chartists' and fundamentalists' excess demand.

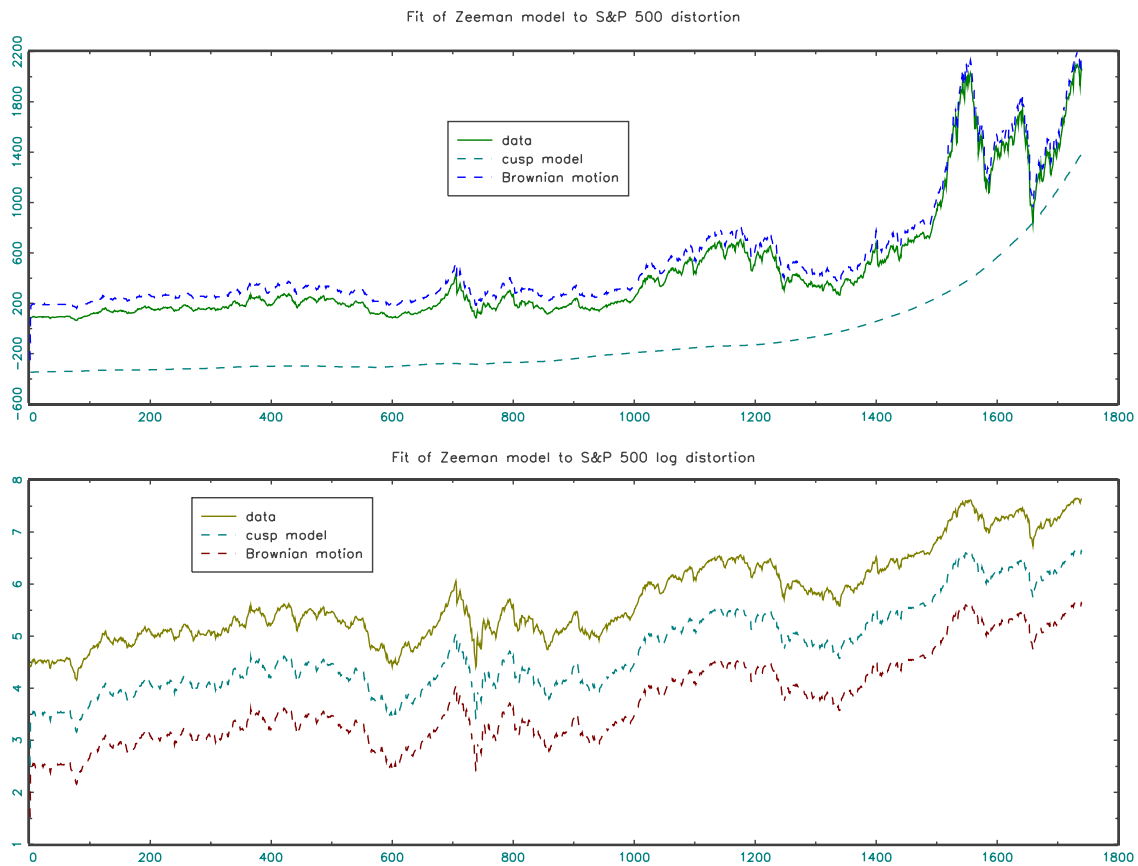


Figure 3: Estimation of Zeeman’s catastrophe model for the S&P500 distortion. The upper panel shows the estimation results for the original data, the lower panel shows results for the log distortion. For the log data, all three curves would have coincided almost perfectly. In order to enhance visibility, the estimated models have been shifted horizontally by constant factors.

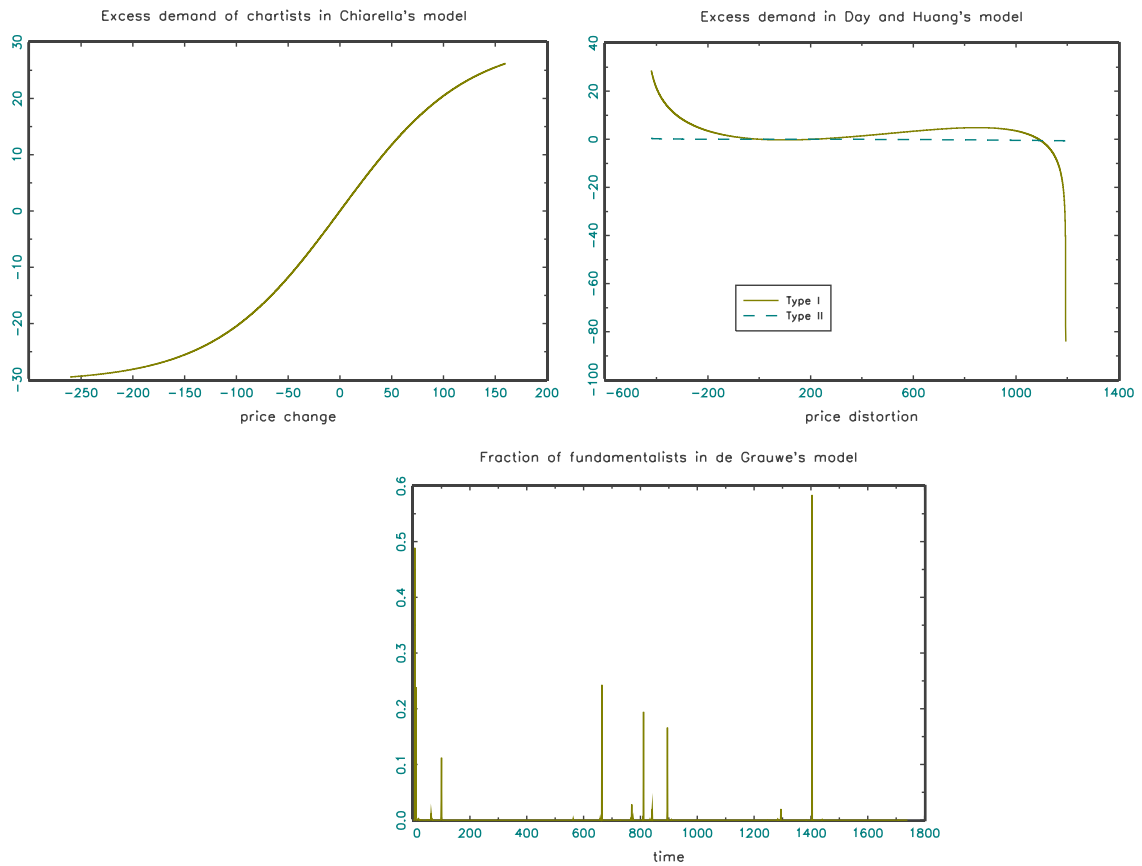


Figure 4: Estimation results for ‘first generation’ behavioural models with raw data. The upper left-hand side shows the nonlinear excess demand function of Chiarella’s model, the upper right-hand side the excess demand function in the two versions of the model of Day and Huang. The lower panel displays the fraction of fundamentalists as it evolves over the sample in De Grauwe et al (version I).

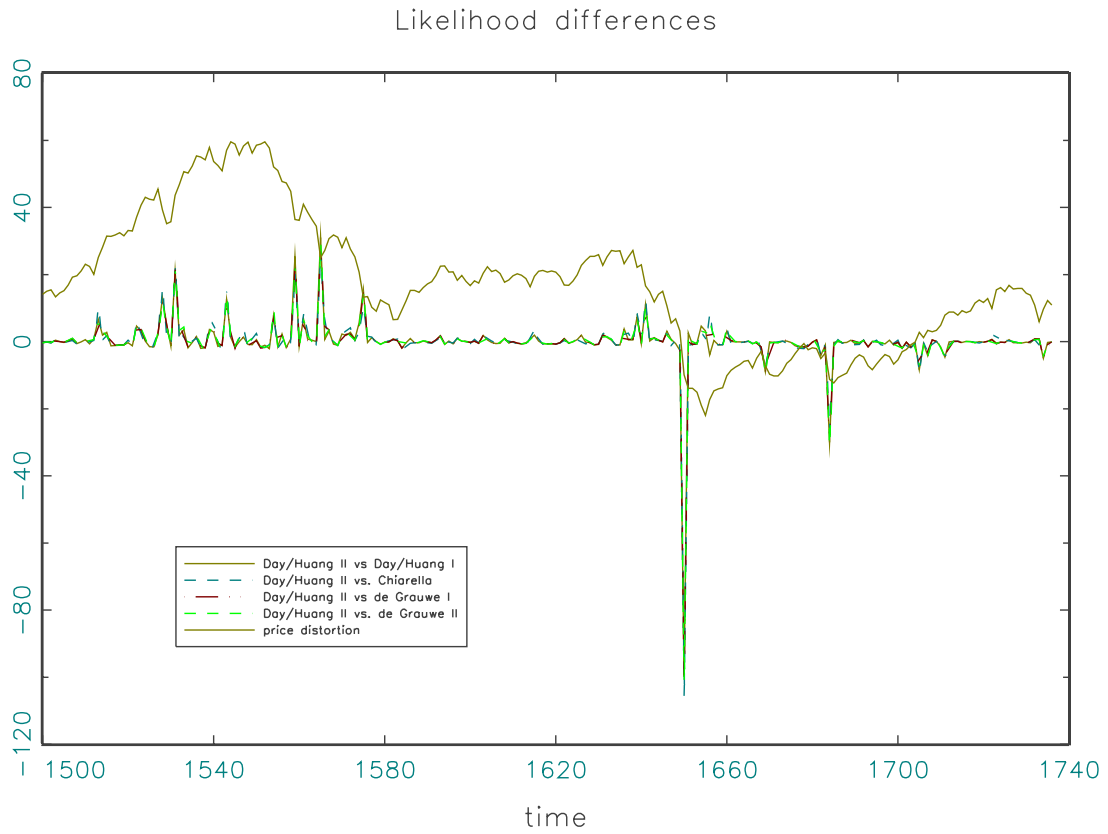


Figure 5: Relative difference in the log likelihood values of the best-performing model (Day and Huang II) against the other 'first generation' models over the last 200 observations of the sample.

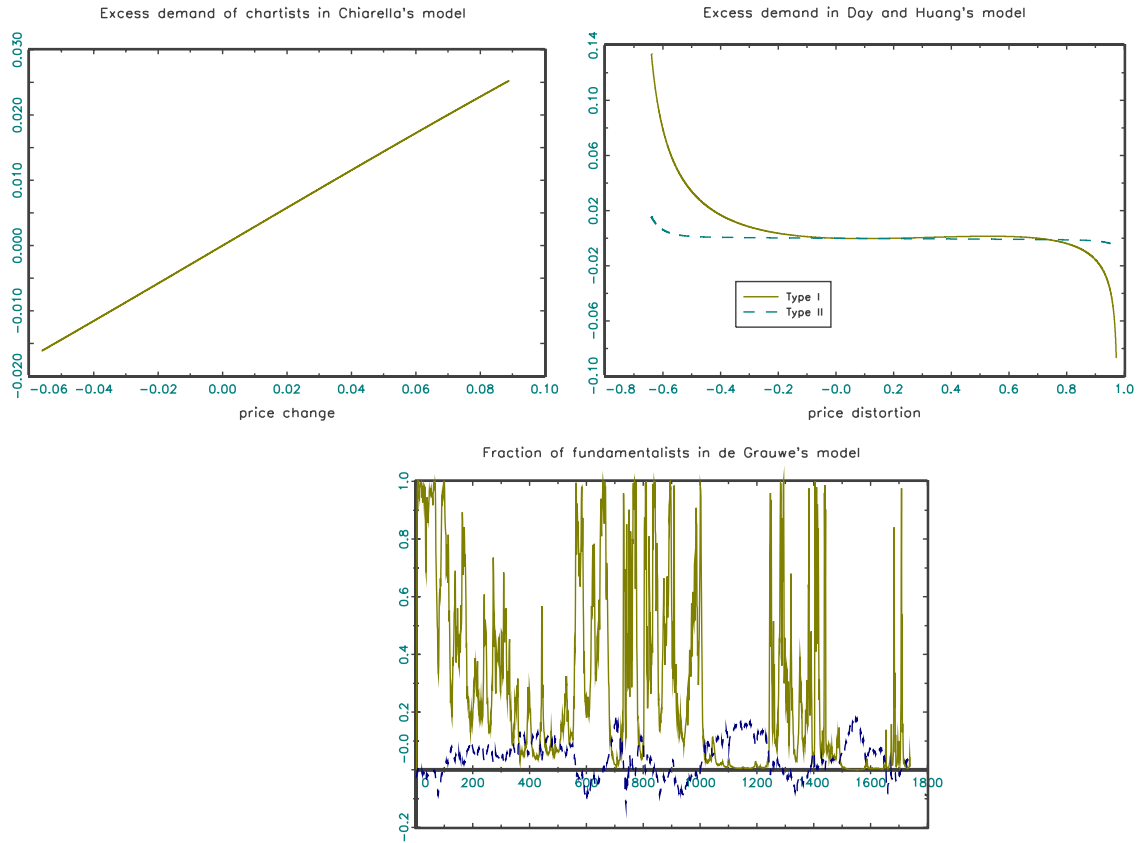


Figure 6: Estimation results for the ‘first generation’ of behavioural models with logarithmic data. Upper panels: Excess demand functions in the model of Chiarella and Day and Huang. Lower panel: Fraction of fundamentalist in De Grauwe et al, version I.

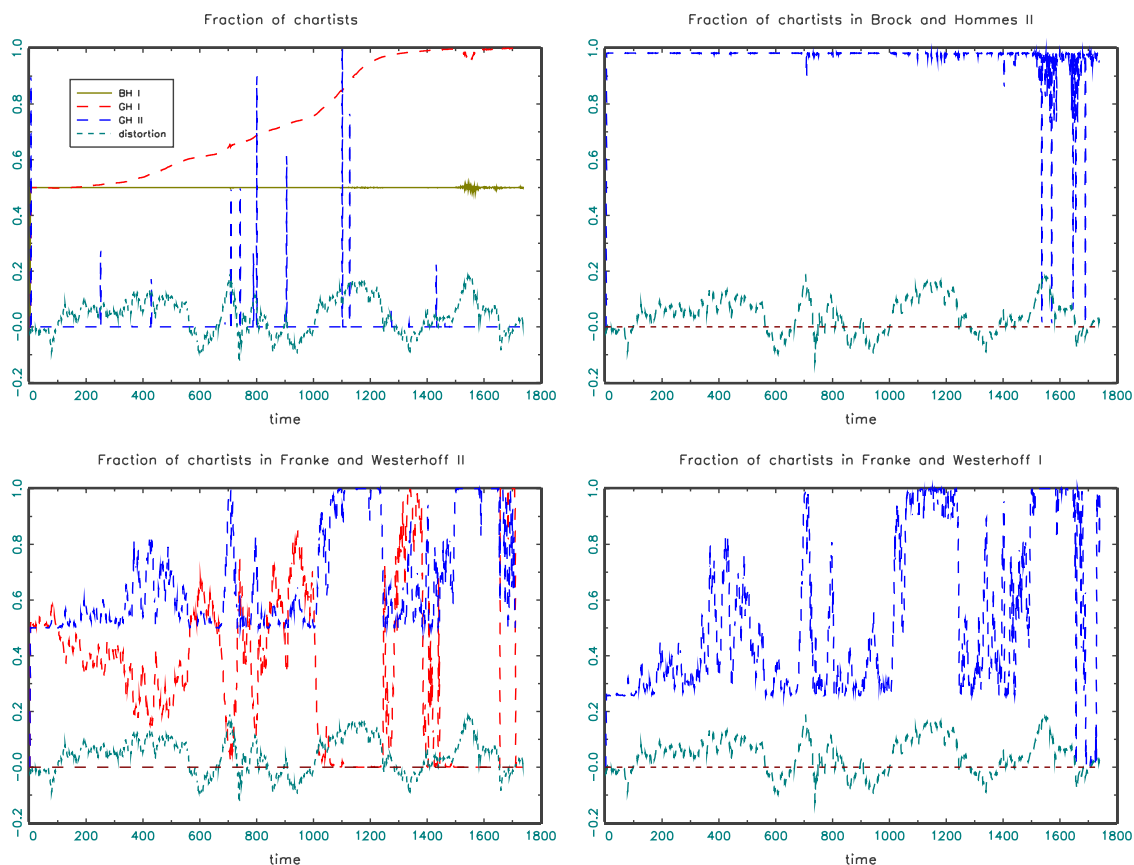


Figure 7: Estimation results for raw data of the fraction of chartists in the models of Brock and Hommes (BH), Gaunersdorfer and Hommes (GH) and Franke and Westerhoff (FW). In the lower left-hand side, two alternative paths are displayed as the filtering of the fractions of both groups can lead to different outcomes in this model. We have always added the log distortion to enable comparison of the population dynamics with historical phases of the S&P 500.

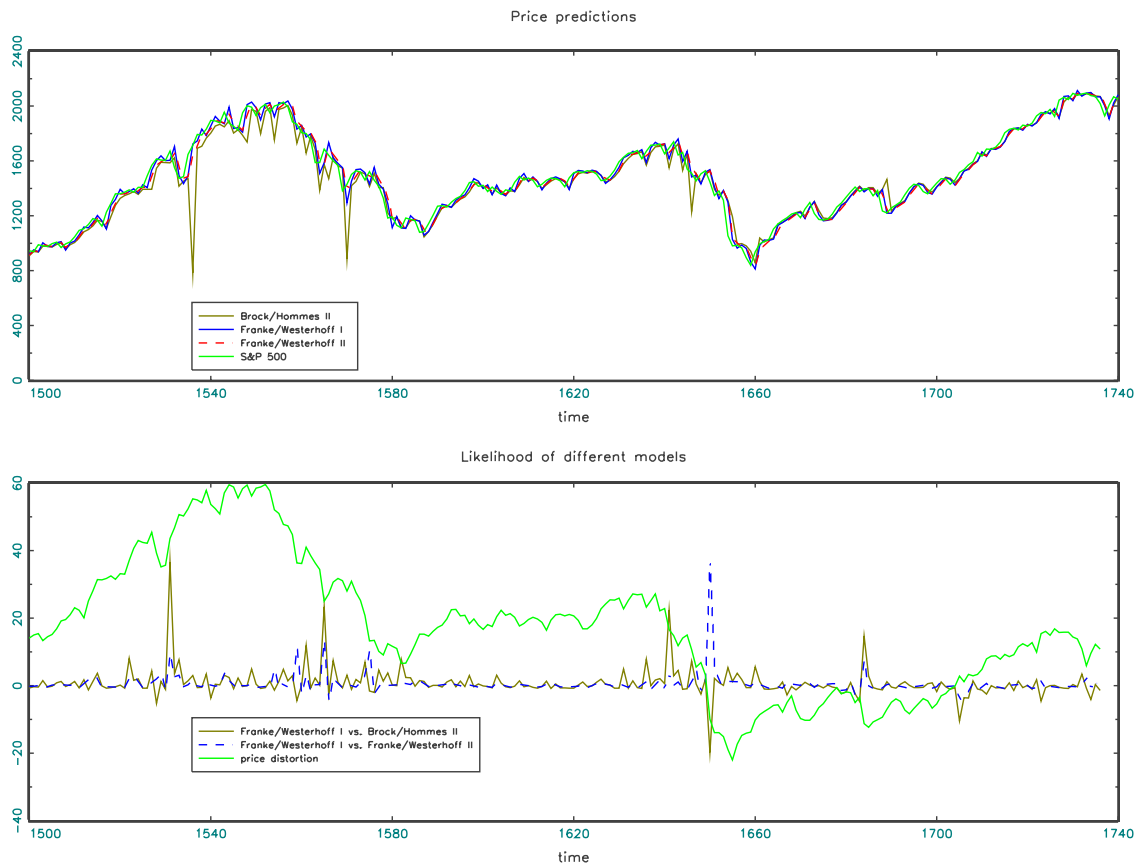


Figure 8: Price predictions and relative log-likelihood values of FW I against the alternatives FW II and BH II. Parameter estimates are obtained on the base of the original price data.

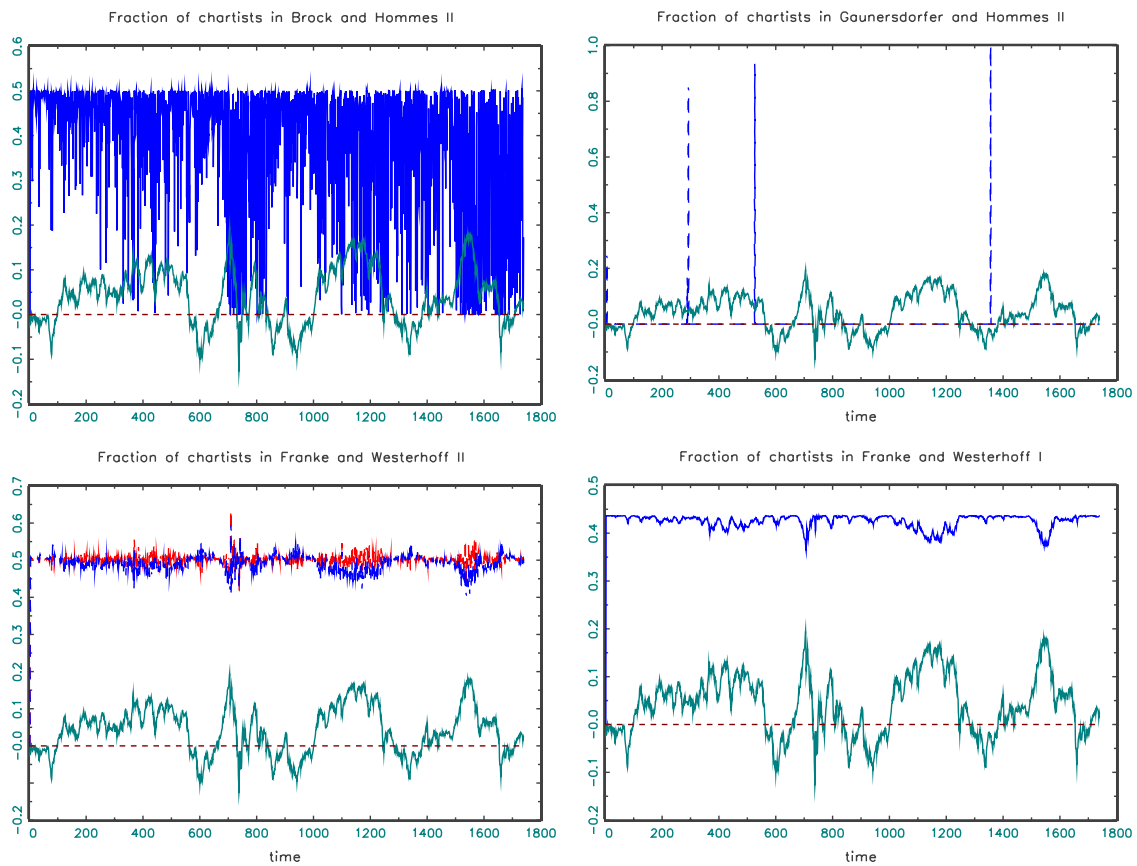


Figure 9: Estimation results for logarithmic data. Estimated fractions of chartists in the models of Brock and Hommes, Gounersdorfer and Hommes and Franke and Westerhoff. The lower left-hand side shows again two alternative filtered time series as different results are obtained in this case for different initialisations of the particle filter.

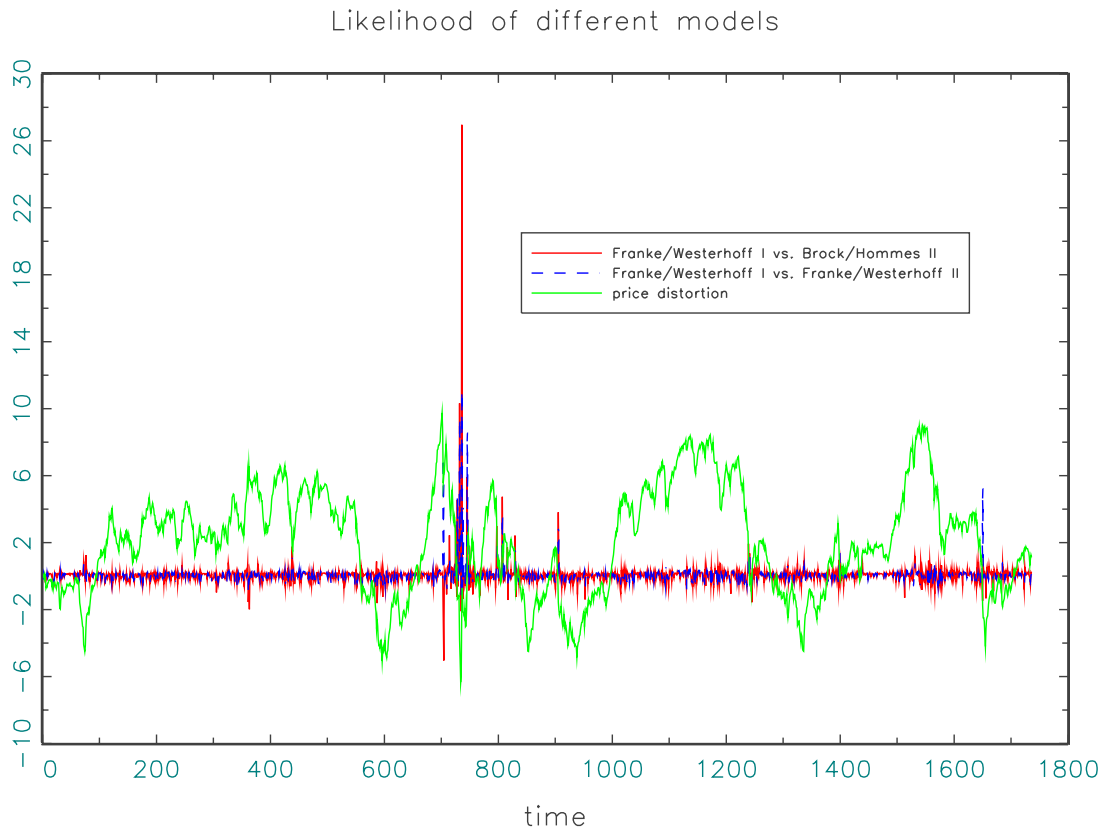


Figure 10: Log-likelihood differences of FW I against the alternative models FW II and BH II for logarithmic data.

Table 1: Simple Benchmark Models of Chartists and Fundamentalists

raw data	chartists			fundamentalists			σ_ϵ	MIL
	b_1	b_2	b_3	a_1	a_2	a_3		
Model 0 (random walk)							22.852 (58.928)	-7895.304
Model 1 (linear)	0.211 (8.915)			0.003 (1.338)			22.340 (0.00)	-7855.925
Model 2 (+ 2nd order)	0.237 (27.851)	7.202e-04 (7.198)		0.005 (1.404)	-1827e-06 (0.552)		22.280 (0.000)	-7851.245
Model 3 (+ 3rd order)	0.223 (19.086)		-9.060e-07 (0.983)	2.685e-06 (0.001)		5.442e-09 (2.411)	22.328 (0.000)	-7855.024
Model 4 (+ 2nd and 3rd order)	0.207 (17.443)	8.820e-04 (3.703)	3.114e-06 (2.198)	0.019 (6.890)	7.163e-05 (8.292)	5.895e-08 (9.748)	22.167 (0.000)	-7843.333
log data								
Model 0 (random walk)							0.041 (58.571)	3080.990
Model 1 (linear)	0.276 (11.928)			0.006 (2.533)			0.039 (0.000)	3151.027
Model 2 (+ 2nd order)	0.279 (12.089)	0.210 (1.281)		0.016 (3.508)	0.019 (2.431)		0.039 (0.000)	3155.305
Model 3 (+ 3rd order)	0.289 (11.036)		-0.522 (0.068)	0.003 (0.567)		0.011 (1.036)	0.039 (0.000)	3151.972
Model 4 (+ 2nd and 3rd order)	0.304 (11.289)	0.207 (1.251)	-0.083 (1.455)	0.009 (1.884)	0.063 (4.920)	0.083 (4.477)	0.039 (0.000)	3167.015

Note: The table shows parameter estimates with t-statistics in parentheses for the polynomial approximation of a nonlinear chartists/fundamentalists model specified in eqs. (1) to (4). Under likelihood ratio tests, all richer models do significantly outperform the restricted models that are nested in them.

Table 2: Estimation of Zeeman model

raw data			
α	-	-	-9.447 (0.000)
β	-	-	16.761 (0.000)
λ	-	244.131 (38.119)	589.776 (0.000)
σ	-	123.700 (55.445)	128.479 (0.000)
σ_z	20.643 (57.501)	10.308 (27.665)	6.809 (0.000)
lkl	-7794.632	-12174.473	-10928.893
δ	-	-	-16425.109
log data			
α	-	-	0.000 (0.088)
β	-	-	-0.002 (-0.265)
λ	-	0.211 (3.083)	0.204 (2.385)
σ	-	3.163 (8.291)	6.196 (4.456)
σ_z	0.039 (56.143)	0.039 (56.286)	0.006 (4.619)
lkl	3116.633	3120.483	3120.533
δ	-	-	$1.5 * 10^{-8}$

Note: The table presents parameter estimates for Zeeman's model and simpler models. Cardan's discriminant δ indicates whether the system exhibit multi-modality ($\delta < 0$) or uni-modality ($\delta > 0$). Number in brackets give the t-statistics of the parameter estimates.

Table 3: Estimation results for ‘first generation’ models

raw data							
Day and Huang I: Noise in demand function							
	a	b	d_1	d_2	m_1	m_2	σ_ϵ
estim.	3.417	3.229	0.006	0.002	-439.265	1193.440	22.656
t-stat	(0.001)	(0.001)	(0.001)	(0.001)	(-0.444)	(22.178)	(90.844)
lkl = -7880.276							
Day and Huang II: Noise in perception of fundamental							
estim.	6.595	0.019	0.017	0.594	-439.265	15309.191	341614.080
t-stat
lkl = -7461.622							
Chiarella							
	a	b	c	σ_ϵ			
estim.	0.003	30.332	0.008	22.330			
t-stat	(2.840)	(3.231)	(2.568)	(81.936)			
lkl = -7855.213							
De Grauwe et al. I: Noise in demand function							
	a	b	d	σ_ϵ			
estim.	-160.098	-0.004	128.344	22.847			
t-stat	(-0.046)	(-0.614)	(0.038)	(86.054)			
lkl = -7894.997							
De Grauwe et al. II: Noise in perception of fundamentals							
estim.	0.001	-2.215	0.000	18352.908			
t-stat	(0.784)	.	.	(0.783)			
lkl = -7895.184							

Table 3: Continued

log data							
Day and Huang I							
	a	b	d_1	d_2	m_1	m_2	σ_ϵ
estim.	0.074	0.082	0.459	0.197	-0.681	0.976	0.041
t-stat	(0.355)	(0.377)	(0.461)	(0.485)	(-7.332)	(8.602)	(50.996)
lkl = 3099.589							
Day and Huang II							
estim.	0.000	-0.001	9.582	7.366	-0.970	1.198	29.877
t-stat	(0.078)	(-0.327)	(0.431)	(0.421)	(-1.139)	(1.624)	(0.328)
lkl = 3142.390							
Chiarella							
	a	b	c	σ_ϵ			
estim.	0.007	0.103	2.801	0.039			
t-stat	(2.504)	(1.708)	(1.485)	(66.219)			
lkl = 3151.499							
De Grauwe I							
	a	b	c	σ_ϵ			
estim.	0.004	98.645	0.002	0.041			
t-stat	(0.785)	(0.001)	(0.001)	(58.367)			
lkl = 3083.116							
De Grauwe II							
estim.	0.004	-0.336	0.000	9.381			
t-stat	(1.244)	.	.	(1.232)			
lkl = 3082.361							

Note: The table presents parameter estimates, t-statistics and the maximized likelihood for the models by Day and Huang (1990), Chiarella (1992) and De Grauwe et al. (1993)

Table 4: Estimation results for models with strategy switching

		raw data					
Brock and Hommes I: Noise in demand function							
	a	d	γ	μ	σ_ϵ		
estim.	2.004	-1.301	1.643e-05	88.692	22.800		
t-stat	(0.057)	(0.000)	(0.000)	(0.000)	(96.026)		
lkl = -7880.182							
Brock and Hommes II: Noise in perception of fundamental							
estim.	1.025	32729.621	1.230e-04	17.639	961.875		
t-stat	(437.574)	(0.121)	(0.121)	(0.122)	(6.291)		
lkl = -7567.922							
Gaunersdorfer and Hommes I: Noise in demand function							
	ν	g_0	γ	σ_ϵ	α	η	μ
estim.	0.681	0.214	0.824	24.138	35361493.316	1.002	191070.140
t-stat	(3.609)	(21.139)	(0.000)	(82.025)	(1.252)	(739.347)	(0.000)
lkl = -7990.288							
Gaunersdorfer and Hommes II: Noise in perception of fundamental							
estim.	0.979	-1.441	0.040	1198.726	199.029	0.750	0.151
t-stat	(0.181)	(-0.002)	(0.000)	(8.357)	(0.001)	(0.001)	(0.000)
lkl = -8023.509							
Franke/Westerhoff I: Noise in demand function							
	a_0	a_n	a_p	γ	ξ	ϕ	σ_c
estim.	1.985	-3.498	0.000	2.650	0.477	0.004	39.870
t-stat	(0.943)	(-0.951)	(-0.949)	(0.965)	(17.309)	(13.934)	(32.514)
lkl = -7027.656 (0.000)							
Franke/Westerhoff II: Noise in perception of fundamental							
estim.	-0.010	-0.005	0.009	-0.010	0.008	-0.155	27.492
t-stat
lkl = -7466.624 (1.391)							
Lux							
	ν	a_0	a_1	a_2	b	a	σ_ϵ
estim.	25.735	-0.204	-6.409	-0.748	0.007	-0.550	22.165
t-stat	7.817	-9.714	-8.964	-6.926	7.000	-8.088	79.730
lkl = -7852.534 (0.128)							

Table 4: continued

		log data									
Brock and Hommes I		a	d	γ	β	σ_ϵ					σ_ϵ
estim.		2.271	-168.837	-4679109259.988	1.603e-05	0.041					
t-stat		(4.085)	(0.000)	(0.000)	(2.306)	(75.645)					
lkl = 3090.779											
Brock and Hommes II		ν	g_0	γ	σ_ϵ	α	η	β			
estim.		1.098e-04	0.078	0.169	2.533e-07	0.075					
t-stat		(1.906)	(0.065)	(3.143)	(0.035)	(9.746)					
lkl = 3123.741											
Gauersdorfer/Hommes I		ν	g_0	γ	σ_ϵ	α	η	β			
estim.		-1.098e+19	-5.963e+18	1.008e+18	1.317e+22	1.721e+19	0.056	3.099e-24			
t-stat		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(78.815)			
lkl = 3090.504											
Gauersdorfer/Hommes II		ν	g_0	γ	σ_ϵ	α	η	β			
estim.		-2.009	2208.993	-2.857e-04	16492.486	68.090	1.779e-04	8.235e-07			
t-stat		(0.459)	(0.225)	(0.063)	(0.335)	(0.770)	(0.000)	(0.302)			
lkl = 3121.693											
Franke/Westerhoff I		a_0	a_n	a_p	a_c	γ	ξ	ϕ	σ_c	σ_f	
estim.		-6.285	-15.262	0.639	0.413	0.399	-0.005	0.054	0.001	0.001	
t-stat		(-0.040)	(-0.040)	(-0.040)	(-0.040)	(1.655)	(-0.882)	(1.661)	(0.014)	(0.014)	
lkl = 3251.599 (0.000)											
Franke/Westerhoff II		ν	a_0	a_1	a_2	b	a	σ_ϵ			
estim.		-0.075	0.185	-0.203	0.156	0.427	0.017	0.002	4.499		
t-stat		(-0.069)	(1.223)	(-10.527)	(17.065)	(12.361)	(12.788)	(0.302)	(20.277)		
lkl = 3149.790 (5.964)											
Lux (1995)		ν	a_0	a_1	a_2	b	a	σ_ϵ			
estim.		1.812	0.044	-0.750	6.643	0.015	0.001	0.038			
t-stat		26.337	27.537	-24.486	28.070	36.558	33.270	51.178			
lkl = 3156.006 (0.614)											

Note: The table presents parameter estimates, t-statistics and maximized likelihood values of the models by Brock and Hommes (1998), Gauersdorfer and Hommes (2007), Franke and Westerhoff (2012) and Lux (1995). Since estimation is a stochastic process for the latter models, the maximized likelihood is given for these cases as the mean over 100 replications (with standard deviations in brackets).

Table 5: Model Confidence Sets

Test statistics t_{i_o}		
Coverage prob.	Raw data	Log data
0.2	Franke/Westerhoff I	Linear, Square, Cubic, Square+Cubic, Day/Huang II, Chiarella, Brock/Hommes II, Gau/Hommes II, Franke/Wester I, Lux
0.1	Franke/Westerhoff I	Linear, Square, Cubic, Square+Cubic, Day/Huang II, Chiarella, Brock/Hommes II, Gau/Hommes II, Franke/Wester I, Lux
0.05	Franke/Westerhoff I	Linear, Square, Cubic, Square+Cubic, Day/Huang II, Chiarella, Brock/Hommes II, Gau/Hommes II, Franke/Wester I, Franke/Wester II, Lux
Test statistics t_{i_j}		
Coverage prob.	Raw data	Log data
0.2	Franke/Westerhoff I	Franke/Westerhoff I
0.1	Franke/Westerhoff I	Franke/Westerhoff I
0.05	Franke/Westerhoff I	Linear, Square, Cubic, Square+Cubic, Chiarella, Brock/Hommes II, Gau/Hommes II, Franke/Wester I, Franke/Wester II, Lux

Note: The table shows the model confidence sets obtained under different specifications of the sequential test algorithm of Hansen et al. (2011) and under different coverage probabilities. t_{i_o} and t_{i_j} denote tests based on comparison of single models with the average of all models and pairwise comparisons, respectively.

References

- Anishchenko, V. S., A. S. Kopeikin, T. E. Vadivasova, G. I. Strelkova and J. Kurths (2000), Influence of noise on statistical properties of nonhyperbolic attractors, *Physical Review E* 62 6, 7886.
- Baruník, J. and M. Vošvrda (2009), Can a stochastic cusp catastrophe model explain stock market crashes?, *Journal of Economic Dynamics and Control* 33 10, 1824–1836.
- Beja, A. and M. B. Goldman (1980), On the dynamic behavior of prices in disequilibrium, *The Journal of Finance* 35 2, 235–248.
- Brock, W. A. and C. H. Hommes (1998), Heterogeneous beliefs and routes to chaos in a simple asset pricing model, *Journal of Economic Dynamics and Control* 22 8-9, 1235–1274.
- Chiarella, C. (1992), The dynamics of speculative behaviour, *Annals of operations research* 37 1, 101–123.
- Cochrane, J. H. (2011), Presidential address: Discount rates, *The Journal Of Finance* 66 4, 1047–1108.
- Day, R. H. and W. Huang (1990), Bulls, bears and market sheep, *Journal of Economic Behavior & Organization* 14 3, 299–329.
- De Grauwe, P., H. Dewachter and M. Embrechts (1993), Exchange Rate Theory: Chaotic Models of Foreign Exchange Markets, *Blackwell, Oxford* .

- Diks, C. and J. Wang (2016), Can a stochastic cusp catastrophe model explain housing market crashes?, *Journal of Economic Dynamics and Control* 69, 68–88.
- Douc, R., P. Doukhan and E. Moulines (2013), Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator, *Stochastic Processes and their Applications* 123 7, 2620–2647.
- Douc, R., F. Roueff and T. Sim (2015), Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models, *Lithuanian Mathematical Journal* 55 3, 367–392.
- Flavin, M. A. (1983), Excess volatility in the financial markets: A reassessment of the empirical evidence, *Journal of Political Economy* 91 6, 929–956.
- Franke, R. and F. Westerhoff (2012), Structural stochastic volatility in asset pricing dynamics: Estimation and model contest, *Journal of Economic Dynamics and Control* 36 8, 1193–1211.
- Gardiner, C. (2009), *Stochastic Methods: Handbook for the Natural and Social Sciences*, Springer, Berlin.
- Gaunersdorfer, A. and C. Hommes (2007), A nonlinear structural model for volatility clustering, in *Teysière, G. and A. Kirman, eds., Long Memory in Economics*, Springer, pages 265–288.
- Giglio, S. and B. Kelly (2018), Excess volatility: Beyond discount rates, *The Quarterly Journal of Economics* 133 1, 71–127.
- Hansen, P. R., A. Lunde and J. M. Nason (2011), The model confidence set, *Econometrica* 79 2, 453–497.

- Judd, K. (2007), Failure of maximum likelihood methods for chaotic dynamical systems, *Physical Review E* 75 3, 036210.
- Leroy, S. and R. D. Porter (1981), The present-value relation: Tests based on implied variance bounds, *Econometrica* 49 3, 555–574.
- Lindström, E. (2007), Estimating parameters in diffusion processes using an approximate maximum likelihood approach, *Annals of Operations Research* 151 1, 269–288.
- Lux, T. (1995), Herd behaviour, bubbles and crashes, *The Economic Journal* 105 431, 881–896.
- Lux, T. (2009), Rational Forecasts or Social Opinion Dynamics? Identification of Interaction Effects in a Business Climate Survey, *Journal of Economic Behavior & Organization* 72 2, 638–655.
- Lux, T. and M. Marchesi (2000), Volatility clustering in financial markets: A microsimulation of interacting agents, *International Journal of Theoretical and Applied Finance* 3 04, 675–702.
- Majewski, A. A., S. Ciliberti and J.-P. Bouchaud (2020), Co-existence of trend and value in financial markets: Estimating an extended Chiarella model, *Journal of Economic Dynamics and Control (in press)* 112.
- McGoff, K., S. Mukherjee, A. Nobel, N. Pillai et al. (2015a), Consistency of maximum likelihood estimation for some dynamical systems, *The Annals of Statistics* 43 1, 1–29.
- McGoff, K., S. Mukherjee, N. Pillai et al. (2015b), Statistical inference for dynamical systems: A review, *Statistics Surveys* 9, 209–252.

- McManus, D. A. (1992), How common is identification in parametric models?, *Journal of Econometrics* 53 1-3, 5–23.
- Poulsen, R. (1999), Approximate maximum likelihood estimation of discretely observed diffusion processes, *Working Paper, University of Aarhus* .
- Schmitt, N. and F. Westerhoff (2017), On the bimodality of the distribution of the S&P 500's distortion: Empirical evidence and theoretical explanations, *Journal of Economic Dynamics and Control* 80, 34–53.
- Shiller, R. (1981), Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71 3, 421–436.
- Sim, T., R. Douc and F. Roueff (2019), General-order observation-driven models: Ergodicity and consistency of the maximum likelihood estimator, *Working paper, Telecom Suparis* .
- West, K. D. (1988), Bubbles, fads and stock price volatility tests: a partial evaluation, *The Journal of Finance* 43 3, 639–656.
- Zeeman, E. C. (1974), On the unstable behaviour of stock exchanges, *Journal of Mathematical Economics* 1 1, 39–49.