Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Parameter Estimation of a non-Equilibrium Asset Pricing Model and Performance Analysis of the Calibration in Terms of Sloppiness

Andreas Lagerqvist

Master of Science Thesis, Spring 2014
Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden
Department of Management, Technology and Economy, Swiss Federal Institute of Technology (ETH), Chair of Entrepreneurial Risk (ER), Zurich, Switzerland
Supervisors: Prof. Filip Lindskog (KTH) and Prof. Didier Sornette (ETH)
Examiner: Prof. Boualem Djehiche (KTH)

Submission Date: May 28, 2014

Abstract

Prices of assets traded in stock markets often exhibit out of equilibrium behaviours, e.g. bubbles and recessions. Yukalov et al. have developed a model to describe these dynamics, and this Master thesis focuses on the problem of calibrating it using an Evolutionary algorithm and the Simulated Annealing method. In general, the parameter estimation performs far from desired, and a Sloppy model analysis of the deterministic system shows that the performance is linked to the sloppiness structure of the model. Accounting for sloppiness, the calibration results can be seen in a different light and the model could still be useful for predictions. Thus, the prediction performance on both synthetic and real-world data is studied, with good results in artificial markets and poor performance using real prices.

Sammanfattning

Aktier handlas ofta för priser som skiljer sig från jämviktspriser, exempelvis under finansbubblor eller i recessioner. Yukalov et al. har tagit fram en modell för att beskriva dessa beteenden, och i den här Masteruppsatsen undersöks modellkalibrering genom en Evolutionär algoritm och 'the Simulated Annealing method'. Generellt är modellparametrarna dåligt uppskattade och en Sloppy model-analys av det deterministiska systemet visar att kalibreringsresultatet är beroende av modellens sloppiness struktur. Med detta i åtanke kan kalibreringsresultatet tolkas annorlunda och modellen kan fortfarande vara användbar för prediktion. Således är prediktionsprecisionen studerad för både syntestisk och riktig data, med god precision för simulerade marknader men sämre resultat för verkliga priser.

Acknowledgements

I am truly indebted to Professor Didier Sornette at the Swiss Federal Institute of Technology for introducing me to this topic and supervising me during the work. He has inspired and supported me as well as proposed different approaches and discussed results. In addition, he has put me in contact with many interesting experts and I am very thankful for the help from Professor Vyacheslav Yukalov at the Bogolubov Laboratory of Theoretical Physics and Professor Yannick Malevergne at the Jean Monet University for reviewing the report and contributing to the discussion on Sloppy models. Furthermore, Professor Filip Lindskog at the Royal Institute of Technology has functioned as an external sounding board and helped me to improve the clarity of the report. I would also like to thank my friend Tatjana Puhan, Ph.D. student in Finance at the University of Zurich for proofreading the report and commenting on the content. Finally, I am uttermost grateful for the support from my mother, father and sister, who have been there and encouraged me during my entire studies.

Andreas Lagerqvist
Zurich, May 2014

Contents

Chapter 1

Introduction

The prices of assets traded in the stock market are, according to neoclassical theories in economics, the result of supply and demand of numerous market participants, ranging from large institutional players and investment banks to individual investors. Economists often model agents as utility maximizing individuals with assets prices as the output from an equilibrium market. Aggregation of individual preferences into market equilibrium is complex, and only recently analytical solutions for markets consisting of agents with heterogeneous utilities and income statements have been found by Christensen and Larsen [10]. However, under rather restricted assumptions. To give an example, asset prices are often modelled as Brownian Motions, an idea that dates back to 1900 and Bachelier [2]. Another commonly used approach in equilibrium theory is the standard Black and Scholes model [5], in which asset prices follow Geometric Brownian Motions.

Nevertheless, real markets tend to exhibit features that are in stark contrast to assumptions made in classical theory, such as bubbles and crashes, the equity premium puzzle, fat-tailed distributions, regime switches etc. [30]. Historical examples are the .com-Bubble and the more recent subprime mortgage crisis that lead to the most recent financial recession. To make models of asset prices more realistic, a body of literature has emerged that aims to describe regime switches, with the first ground breaking work by Hamilton in 1989 [15]. Generally, ad hoc econometric models with statistical processes are used. In the specific case of asset prices, regime switching properties are often based on measurements for periods of high and low volatility or long bull and bear markets. These models match narrative interpretations of market fundamentals, quantities that sometimes only can be interpreted ex post, but still be useful, e.g., for ex-ante real-time forecasting or portfolio choice.

Another approach, unique of its kind, is proposed by Yukalov et al. [35]. They suggest a dynamic model with competing attractors for non-equilibrium asset prices, providing a natural framework for shifts between different market phases. Furthermore, in contrast to other approaches, the model describes and quantifies the underlying characteristics of the market and its participants, including price formation delay between decision and

investment, market regulation and frictions, speculative momentum effects and mean-reversal trading strategies. Thus, if it is possible to conclude that the model has high explanatory power of asset prices, then model interpretations would provide new insights about the dynamics of financial markets.

Based on these interesting features of the model by Yukalov et al., this Master thesis analyses the performance of calibrating the model using two different parameter estimation methods, the Evolutionary algorithm and Simulated annealing. The Evolutionary algorithm is a search heuristic that implicitly tries to maximize the probability distribution of the parameters given the market prices, while the Simulated annealing method is a search algorithm that explicitly estimates this quantity. This work proceeds two earlier Master theses at the Swiss Federal Institute of Technology Zurich with similar background. In those previous works, the parameter estimation performance is rather poor. Therefore, the objective of this thesis is to obtain a deeper understanding of why the calibration problem is hard and to apply new calibration methods to investigate if these can help to improve the performance. In addition, an analysis of the asset pricing model in terms of Sloppy models is performed to enhance the understanding about why the different parameter estimation methods cannot calibrate the model satisfactory.

If it is possible to improve upon model calibration, then the model could be used to describe asset prices and to model transitions between different mispricing regimes. In particular, state dependent probabilities of regime switches for calibrated real-world applications are desirable. A more modest goal is to achieve usability of the calibrated model in forecasting, which is examined in an introductory manner in this work using distributions of simulated asset prices.

The analyses in this work show that the deterministic version of the asset pricing model by Yukalov et al. has a considerable sloppiness structure. Thus, it is not reasonable to expect that the calibration methods should be able to perform perfectly. Nevertheless, the model predictions could still be very accurate. In the context of synthetically generated data, the prediction performance is very satisfactory, with thin prediction distributions and high precision. However, the accuracy is not good on real financial data, which suggests that further studies are needed to conclude whether the asset pricing model is useful for real-world forecasting or not.

The model by Yukalov et al. is introduced in Chapter 2 and the different parameter estimation methods are presented in Chapter 3. In Chapter 4, the calibration performance of synthetically generated data is studied. The corresponding analysis of a S&P 500 time-series is conducted in Chapter 5. Finally, Chapter 6 concludes the thesis.

Chapter 2

Dynamic Asset Pricing Model

This chapter presents the continuous-time dynamic asset pricing model developed by Yukalov et al. in Section 2.1. Moreover, Section 2.2 introduces state-space methods as the main tool to approach the discrete realisations of the continuous-time model and presents necessary model assumptions used in this work.

2.1 Non-equilibrium Asset Pricing Model

In Yukalov et al. [35], the price of a single asset traded by heterogeneous agents in a market with uncertainty and regulatory constrains is derived in terms of a dynamic continuous-time model. Their most general model considers i) price formation delay between decision and investment, ii) linear and non-linear mean-reversal or speculative trading, iii) market friction, iv) uncertainty in fundamental price, v) non-linear speculative momentum effects and vi) market regulations.

The model differs from the traditional view that randomness in asset prices is due to publicly accessible information only, as used for instance in the standard Black and Scholes equilibrium model [5]. Instead, this model decomposes the consensus of the markets participants' value of the asset (the market value) in a fundamental value and a mispricing part, whereas the fundamental value should represent the *correct* value of the asset. The framework does not attempt to model the behaviour of single market participants, but instead the aggregated actions and believes of all interacting agents. The mispricing component for different market constitutions has different characteristics and the model aims to describe both the properties and the underlying causes of mispricing regimes.

2.1.1 Mathematical Description

The model is introduced to describe the market price $p(t)$ at time t of the single asset traded in the market (where the time dependence subsequently will be dropped for ease of notation), decomposed as a fundamental price p_f and mispricing x as:

$$x = \log(p) - \log(p_f). \tag{2.1}$$

Thus, x denotes the systematic deviation from the fundamental price. E.g. in times of highly overvalued asset prices (bubbles), the associated mispricing will be positive. In this model, the mispricing x is assumed to be described by a Stochastic Differential Equation (SDE) given by:

$$dx = ydt + \sigma_x dW_x, \tag{2.2}$$

where W_x is a Wiener process. In equilibrium, the drift term y is determined from the von Neumann-Morgenstern utility function of agents [16] based on consumption and dividends in the economy [9]. In the case of non-equilibrium markets, Yukalov et al. assume that y endogenously encounters these fundamentals of an economy by letting y depend on the state (x, y) through:

$$dy = f^{NL}(x, y)dt + \sigma_y dW_y, \tag{2.3}$$

where W_y is a Wiener process, possibly correlated with W_x. This structure implies a feedback mechanism from the self-consistency and reflexivity of financial markets as a consequence of the collective organization of investors, where reflexivity refers to agents believes and actions being shaped by market expectations. Thus, the model is a second order SDE where y can be interpreted as the inertia or resistance of changes in the mispricing x. Moreover, the following assumptions are made: i) there are no further external influences apart from the dynamics of the fundamental price, ii) the market is asymmetric $f^{NL}(-x, -y) = -f^{NL}(x, y)$, i.e., there are no drastic differences between rising and falling prices, and iii) f^{NL} is additive $f^{NL}(x, y) = f_1(x) + f_2(y)$. Now, using the self-similar approximation theory [34], Yukalov et al. derive the higher order formula:

$$f^{NL}(x, y) = \alpha x + \beta y + Ax^3 \exp(-x^2/\mu^2) + By^3 \exp(-y^2/\lambda^2), \tag{2.4}$$

where interpretations of the parameters are given in the next section. To completely state the model, further assumptions on the dynamics of the fundamental price need to be made, i.e., on dq_f where $q_f = \log(p_f)$.

2.1.2 Economic Interpretation of Model Parameters

The parameters in Equation (2.4) have the following economic interpretation:

- α. The individual behaviour of agents. $\alpha < 0$ suggests that individuals follow a mean-reversal strategy, while over-speculative markets are characterized by a positive α.
- β. A factor that mitigates market frictions, i.e., lowers changes in the mispricing's drift. Positive β's are associated with explosive prices and could only be transient by definition. Since only static parameters are considered, positive β's are ruled out. Theoretically, a positive β together with negative B and sufficiently large λ also yields non-explosive prices.

- A. The collective behaviour of agents. A reduces or amplifies the effect of α. The non-linearity makes A influential only for a sufficiently large degree of mispricing. $A > 0$ represents collective speculation behaviour while $A < 0$ describes collective correction behaviour.
- μ. A measure of the investors' uncertainty about the fundamental price. It is only important when $A > 0$. Small μs correspond to low uncertainty, which results in smaller mispricings, and vice versa.
- B. Represents momentum strategies. $B > 0$ is associated with trend reinforcement, while negative values on B suggest that investors follow contrarian strategies. In practice, the latter strategy is seldom used by a majority of traders and therefore only $B > 0$ is considered.
- λ. Measures market regulation. Large λs describe free markets and allow for larger mispricing, while rigid markets (small λs) are not subject to this kind of mispricing.

2.1.3 Market Types

Different choices of parameters give rise to different market characteristics, reflecting e.g., the individual and collective behaviour of agents and market properties such as information access and regulation. A thorough analysis including phase diagrams and attractor analysis of the deterministic version of the model is found in Yukalov et al., and in total there are 18 markets with different properties. A selection of these are shown in Table 2.1, illustrating, e.g., possible fixed points. For clarity, market identifiers are introduced. Each identifier refers to a number of a figure in the paper of Yukalov et al.

In the deterministic case, some of the markets have standard equilibrium characteristics with one fixed point for the mispricing $(x, y) = (0, 0)$. Other, non-equilibrium markets, have both positive and negative mispricing attractors as well as limiting cycles. It is notable, that some non-equilibrium markets exhibit completely different market phases, depending on the present state of the mispricing (or, more formally, the initialization).

Adding stochasticity does not change the fundamental structure of the market types, albeit it allows for regime switches. That is, stochasticity makes it possible for the mispricing to change from positive to negative conventions and vice versa. This could be interpreted as the asset price moving from bubble phases (over valuation) to recessions (under valuation).

Table 2.1: Overview of a selection of market characteristics for different market types and investor preferences. Note: $s > 0$ is a fixed value and $C_{(x,y)}$ is a cycle around (x, y).

Market	Characteristics	Properties of Agents	Fixed Points: $\{(x, y)\}$ Cycles: $C_{(x,y)}$
1	Certain, $\mu \to 0$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Over regulated, $\lambda \to 0$	Non-speculative, $A < 0$	
3	Certain, $\mu \to 0$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Unregulated, $\lambda \to \infty$	Non-speculative, $A < 0$	$C_{(0,0)}$
4	Uncertain, $\mu \to \infty$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Over regulated, $\lambda \to 0$	Non-speculative, $A < 0$	
7	Uncertain, $\mu \to \infty$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Unregulated, $\lambda \to \infty$	Non-speculative, $A < 0$	$C_{(0,0)}$
8	Uncertain, $\mu \to \infty$	Positive feedback, $\alpha > 0$	$\{(\pm s, 0)\}$
	Over regulated, $\lambda \to 0$	Non-speculative, $A < 0$	
10	High uncert., $\mu_c < \mu < \mu_1$	Positive feedback, $\alpha > 0$	$\{(\pm s, 0)\}$
	Soft reg., $\lambda_1 < \lambda < \infty$	Non-speculative, $A < 0$	$C_{(0,0)}$
12	Certain, $\mu \to 0$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Over regulated, $\lambda \to 0$	Speculative, $A > 0$	
14	Certain, $\mu \to 0$	Mean-reverting, $\alpha < 0$	$\{(0, 0)\}$
	Unregulated, $\lambda \to \infty$	Speculative, $A > 0$	$C_{(0,0)}$
15	Med. uncert., $\mu_c < \mu < \mu_1$	Mean-reverting, $\alpha < 0$	$\{(0, 0), (\pm s, 0)\}$
	Weak reg., $\lambda_1 < \lambda < \infty$	Speculative, $A > 0$	
17	High uncert., $\mu_c < \mu < \mu_1$	Mean-reverting, $\alpha < 0$	$\{(0, 0), (\pm s, 0)\}$
	Soft reg., $\lambda_1 < \lambda < \infty$	Speculative, $A > 0$	$C_{(0,\pm s)}$

2.1.4 Fundamental Price

As stated earlier, a complete specification of the model needs assumptions on the fundamental price. A simplified but still rich approach is to choose a deterministic fundamental price p_f with dynamics for $q_f = \log(p_f)$ given by:

$$dq_f = \mu_f dt. \tag{2.5}$$

The deterministic choice could be motivated to hold in the short run, but is primarily chosen in order to mitigate the stochasticity of the model and to provide a better overview and understanding of the already complex system dynamics. Other, more sophisticated and realistic alternatives, would be the classical Geometric Brownian Motion fundamental price or an approach with a stochastic volatility component, e.g., a GARCH process. However, increasing intricacy of the fundamental price makes it more difficult to justify the exogenous fixing of some parameters because of lower transparency and increased interdependence, cf., Section 2.2.3.

One consequence of this assumption is that the mispricing x is given directly by elementary operations. More precisely, it is given, up to an additive constant, by subtracting the fundamental price from the market price. Unless stated otherwise, the model is, in this work, translated so that the initial fundamental price is given by $p_f(0) = 1$, i.e., $q_f(0) = 0$. Thus, the mispricing x can be considered the observed variable. Note that the interpretation of the parameter μ as the investors' uncertainty about the fundamental price is not dependent on the stochasticity of the fundamental price. For instance, the investors could believe that the fundamental price follows some highly stochastic dynamics. For real-world interpretations, a deterministic fundamental price seems to be reasonable when studying asset prices ex post and the information about both the asset and the stock market is greater, retroactively.

2.2 State-Space Representation

As a direct consequence of the specification of Yukalov et al.'s model, the interesting components x and y that describe the mispricing are generally not observable. Exception is the case with deterministic fundamental price, that immediately yields the mispricing component x from the observed price. Nevertheless, the mispricing component y is still hidden from direct observations. In addition, (x, y) describes a continuous time process while observations of asset prices are given in a discrete time framework. An appropriate mathematical tool in this set-up is a state-space representation, which will be introduced in this section.

2.2.1 General Model

State-space models (SSM) refer to a class of methods that describe the probabilistic dependence between the latent state variable $\mathbf{U}_t$ at time t, representing the state of the system, and the observed measurement $\mathbf{V}_t$ at time t. The approach originates from the American space program for tracking satellites, and has been successfully applied in e.g., engineering, statistics and finance.

A state-space model provides a general framework for analysing both deterministic and stochastic dynamical systems that are observed or measured via a stochastic process. The hidden state $\mathbf{U}$ evolves according to some dynamic, and the observations $\mathbf{V}$ are outcomes of a probabilistic function of $\mathbf{U}$. The system is in many applications Markovian, and the models are therefore often called Hidden Markov Model (HMM). For instance, a first order Markovian model has the properties:

$$(\mathbf{U}_t \perp\!\!\!\perp \mathbf{U}_{1:t-2}, \mathbf{V}_{1:t-1})|\mathbf{U}_{t-1}, \tag{2.6}$$

$$(\mathbf{V}_t \perp\!\!\!\perp \mathbf{U}_{1:t-1}, \mathbf{V}_{1:t-1})|\mathbf{U}_t. \tag{2.7}$$

However, there are more general state-space models than the HMM. For example the dynamics could be non-linear and non-Gaussian. A schematic overview of a first order

Hidden Markov Model is illustrated in Figure 2.1.

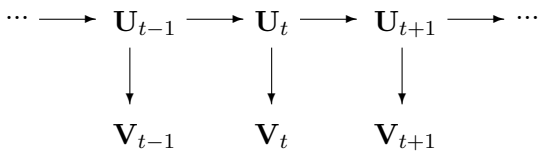

Figure 2.1: Schematic overview of a general Markovian state-space model with state process U and observation process V.

2.2.2 Discretization of the State Variables

To be able to adapt the model by Yukalov et al. to observed (discrete) asset prices, a state-space representation is needed. However, the continuous time model is not directly transferable to the state-space representation with discrete dynamics. Nonetheless, this is only a matter of discretization and the most straight forward method is to use the Euler-Maruyama approximation scheme. This translates equations 2.2 and 2.3 into:

$$X_{t+\Delta t} = X_t + Y_t \Delta t + \sigma_x \sqrt{\Delta t} \epsilon^x_{t+\Delta t}, \tag{2.8}$$

$$Y_{t+\Delta t} = Y_t + f^{NL}(X_t, Y_t)\Delta t + \sigma_y \sqrt{\Delta t} \epsilon^y_{t+\Delta t}, \tag{2.9}$$

where $\epsilon^x, \epsilon^y \sim \mathcal{N}(0,1)$ are the increments of the Wiener processes. From this point and onwards, ϵ^x and ϵ^y are for simplicity assumed to be independent, although they could be correlated. Capital letters X and Y represent the discretized version of the continuous-time model given by lower-case letters x and y. For ease of notation, $X_{t+\Delta t}$ and $Y_{t+\Delta t}$ will be written as X_{t+1} and Y_{t+1}, respectively. That is, the time unit is assumed to be Δt.

The assumptions made on the fundamental price in Section 2.1.4 imply that the interdependency between the hidden state Y and the observed process X is more delicate than in the HMM model. In particular, both the state and the observation process interact with each other. A schematic overview is in Figure 2.2. The most critical property of the dependence structure is the linkage of Y_t with X_{t-1} and X_t with X_{t-1}, i.e., the fact that the state depends on the previous observation and that observations are interdependent. Classical filtering techniques cannot usually cope with this kind of dependency, and special methods to be able to face this problem are presented in Chapter 3.

One way of implementing classical filtering methods, as previously studied in the Master thesis by Robert [28], is adding stochasticity to the modelling by a non-deterministic fundamental price. In his work, fundamental prices are modelled as Geometric Brownian Motions, making it possible to have a two-dimensional state (X, Y) and a observation process given by the log-returns of market prices. For further details, see the work of Robert.

The distinguishing attempt of this work is to investigate if there is some possibility to exploit the information structure in a less stochastic environment, focusing more on the hidden mispricing structure. A well-suited calibration approach should not be overly dependent on the complexity of the model. For instance, the need of imposing uncertainty might be a sign of trying to fit the model into a framework that in reality is not designed to contemplate the full structure of the problem. More detailed discussions on previous works and properties of the attempts in these works are found in Sections 3.1 and 3.2.

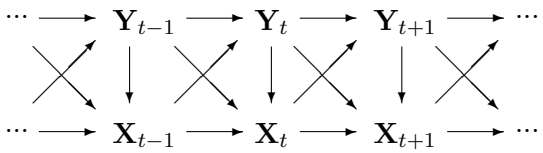

Figure 2.2: Schematic overview of the state-space representation of Yukalov et al.'s model with deterministic fundamental price, state process Y and observation process X.

2.2.3 An Example of a Simulated Process

With use of the Euler-Maruyama approximation scheme presented in Section 2.2.2, it is possible to simulate asset prices from the dynamic model. It is a well-known fact in macroeconomics that business cycles usually change on yearly time-scales. Moreover, qualitative studies of the model by Yukalov et al. indicate that also the change of attractors and cycles corresponds to a scale of a time unit in years. Using 250 trading days every year gives a time increment of $\Delta t = 1/250$, which is small enough to ensure that the Euler-Maruyama numerical scheme converges appropriately.

With a fixed time scale, it is possible to adapt the other parameters to the model. In this work, the parameters that globally describe the market price are assumed to be given exogenously to enhance the transparency of the fundamental model characteristics. That is, μ_f, σ_x and σ_y are assumed to be known. Naturally, this limits the direct usage of the model in a real-world context, since these assumptions need to be verified externally, but the relevance of the unique model properties given by the other parameters will be more easily addressed. Thus, this assumption trades a less complex parameter space for a less endogenous use of the model. However, there exist numerous studies of the global market parameters μ_f and σ_x that can be used, see below.

μ_f is interpreted as the average yearly return, and the work of Robert used the value 12.5%. For simulation purposes with deterministic fundamental price, the growth rate of the fundamental price is redundant, since it is added to the modelling of x to get the market value p. However, for application to real financial data, the choice of μ_f is crucial. In Chapter 5, the more realistic value on 4%, corresponding to the long-term GDP growth, is used [31]. Furthermore, a short analysis of the sensitivity to μ_f is presented. In turn, σ_x describes the volatility of asset prices, which empirically is observed to be about 1-2% daily. To mitigate the stochasticity of the problem, a daily volatility

of 1% is chosen. This implies a yearly volatility of $\sigma_x = 0.01\sqrt{250}$. The case of σ_y is more complicated, since its effect on the market price is not explicitly given, but influences the dynamics of the system indirectly. In accordance with the work of Robert [28], $\sigma_y = 3\sigma_x$ is chosen. This is set from testing of the filtering on Dow Jones Industrial Average (DJIA) data during the time-period 1970 to 2012. In general, the filtering does not seem to be overly dependent on the choice of σ_y, and is therefore not investigated further.

Figure 2.3 exhibits one simulated trajectory of market 15 in Table 2.1. The market has speculative, mean-reverting agents that trade in a weakly regulated market with intermediate uncertainty about the fundamental price. The simulation starts at $(x_0, y_0) = (0, 0.1)$, which is close to the zero attractor $(x, y) = (0, 0)$. It is interesting to see how the mispricing x, after about 10 years, exhibits a change of regime to the positive attractor around $(x, y) = (3, 0)$. This can also be observed for the market price p, which after 10 years clearly deviates from the exponential growth in fundamental price.

2.2.4 Summary of Model Assumptions

The closing part of this chapter provides a short list of all of the different assumption made in this work with regard to the asset pricing model by Yukalov et al.

- The model parameters α, A, β, B, μ and λ are constant, with unknown values that are to be estimated.
- The fundamental price is assumed to be deterministic and exponentially growing, implying that the market price follows a type of Ornstein-Uhlenbeck process.
- The hyperparameters μ_f, σ_x and σ_y are known constants and the model is valid on a yearly time scale. The constants are determined to keep the stochasticity of the model as low as possible. Moreover, the Wiener processes driving x and y are independent.
- The Euler-Maruyama discretization scheme has enough convergence properties to be used in simulations of the model.

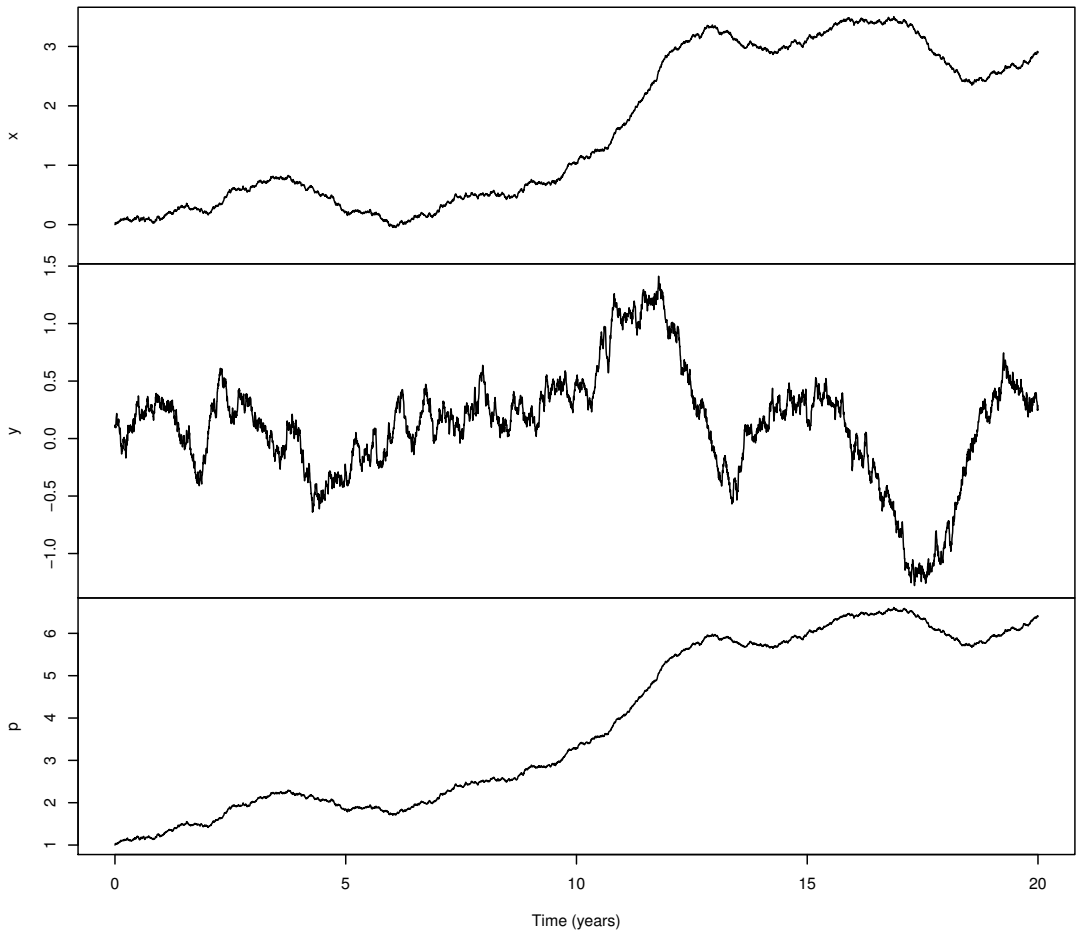

Figure 2.3: One trajectory of the mispricing (x, y) and the market price p of a simulation of the asset price in a market with parameters $\alpha = -1$, $A = 1$, $\beta = -1$, $B = 1$, $\mu = 2$, $\lambda = 1$, $\mu_f = 0.125$, $\sigma_x = 0.01\sqrt{250}$ and $\sigma_y = 3\sigma_x$. The simulation starts in $(x_0, y_0) = (0, 0.1)$.

Chapter 3

Parameter Estimation Methods

This chapter starts in Section 3.1 with an overview of previous calibration studies on the asset pricing model of Yukalov et al., their key findings and shortfalls. In addition, the improvements of the earlier studied methods that are considered in this work are summarized. In Section 3.2, the Evolutionary algorithm is presented as one new calibration approach. Moreover, Section 3.3 contains another example of a parameter estimation method that uses random walkers in the parameter space, called Simulated annealing.

3.1 Previously Studied Calibration Procedures

In this section, earlier attempts on model calibration are summarized, including both their insights and limitations. These former works are two Master theses at the Swiss Federal Institute of Technology, by Bertolace in 2009 [3] and Robert in 2012 [28]. The reader is expected to be fairly familiar with the concepts in this section, and only general discussions with no mathematical details and precise statements are included.

3.1.1 Kalman Filter Techniques

The Master thesis by Bertolace includes a method of solving for the probability distribution of the mispricing $p(x, y|\theta)$ numerically using a Fokker-Planck equation, where θ is the vector of model parameters. Thereafter, the widely used Kalman filter is implemented. Initially, a naive Linear Kalman filter is used, which is only theoretically applicable in special cases of the asset pricing model. Finally, Bertolace implements the Extended Kalman filter and attempts to calibrate the model using state augmentation.

For the filtering purpose, the Kalman filter performs rather well. But for the calibration exercise, the results are heavily dependent on the initial conditions of the parameters. Small deviations from the true parameter values result in a poor calibration, which could be a consequence of biased estimates due to the non-linearity of the model. Moreover, Bertolace mentions a concern regarding the possibility that the same attractors, and hence the same mispricing realizations, can be reached for different sets of parameters.

3.1.2 Sequential Monte Carlo Methods

Bertolace's thesis is followed by the work of Robert, applying Sequential Monte Carlo methods for calibration of the model, in particular Particle filters. For filtration, the appropriate prediction and update densities are derived using Bayesian inference, and Particle filters using systematic resampling are implemented. Most effort is put in deriving a fully adapted Particle filter. The properties of this filter are not examined, but Kong et al. [23] show that using the posteriori itself (Robert's choice) as importance density gives minimal filtration variance. Often, a problem in practice is that this density is not explicitly known. However, very fortunately, this is given by Yukalov et al.'s model. Regarding the calibration, Robert considers three major approaches:

- Maximum Likelihood estimation (ML)
- A Monte Carlo Markov Chain algorithm (MCMC)
- State augmentation

The traditional Maximum Likelihood estimation approach cannot be applied to the system in a straight forward manner, since the probability distribution of the mispricing $p(x, y|\theta)$ is not explicitly known. Thus, Monte Carlo estimations are needed to integrate the hidden states out, demanding simulations of whole paths of the system. This, in turn, increases the dimensionality of the problem leading to a very inefficient implementation without reasonable results.

The MCMC algorithm is based on sampling of both the state and the parameters using Gibbs samplers. However, it is by no means clear what an appropriate conjugate prior for the parameters should be. Studying a reduced model with only α and β unknown, a normal prior can be used as a consequence of their linear influence on the mispricing dynamics. Nevertheless, the implementation was disappointing and with no reasonable convergence, which can be explained by the state sampler mixing too slowly and high variance of the estimators of α and β. Another reason could be non-appropriate priors.

Lastly, Robert investigates filtration with state augmentation as an opportunity to achieve validity to the Sequential Monte Carlo approach after all. Both fixed and artificial dynamics of the parameters are considered, and both methods lead to a high degree of sample impoverishment. The conclusion is that one cannot expect to estimate the parameters precisely, albeit signs of parameters and over-all market characteristics can be acquired. Thereafter, Robert applies this procedure to Dow Jones Industrial Average data with some interesting results. For further reading, please see Robert's work [28].

3.1.3 General Remarks and Considered Improvements

An implication of Robert's work is that, due to the lack of knowledge about the probability distribution of the mispricing in combination with no appropriate prior on the parameters, state augmentation is the only applicable calibration method with reasonable

results. The greatest issue with the method is the wide-spread sample impoverishment, meaning that only a few particles survive as the filtering proceeds. In this thesis, a few methods to alleviate this drawback have been considered and are presented in detail in the next paragraphs.

In the literature, there are two methods that are often used in the presence of sample impoverishment; the Regularized Particle filter (RPF) and the Resample-Move algorithm. The fundamental problem with sample impoverishment is that the continuous distribution of asset prices in Yukalov et al.'s models is approximated via state-space representation into a discrete one. In practice, this means that samples drawn from the same distribution could be identical, whereas this does not occur for a continuous distribution.

The Regularized Particle filter is designed to alleviate this problem by using a kernel density estimate of the particles, rather than the particles themselves. This means that particles are sent through a kernel during the resampling step. If a systematic resampling is used, i.e., the particles are reweighted at every time-step to equal weights, Arulampalam et al. [1] derive the optimal kernel function and bandwidth minimizing the mean integrated square error between the true filtering distribution and the kernel density estimate, known as the Epanechnikov kernel. In the practical implementation for this asset pricing model, the kernel uses the empirical covariance matrix of the states in the filtering. With the state augmentation procedure and reasonably wide initial intervals for the different parameters, the kernel moves the particles too widely and allows for explosive behaviour of the filtering process. The only solution is to increase the number of particles substantially, leading to a lack of computer power and no practical use. Another problem concerns the relatively high dimensionality of the state, in turn demanding even more particles to be used. Therefore, this method is not pursued any further.

Likewise, the Resample-move algorithm, introduced by Berzuini et al. [4], moves the particles after (or before) resampling through a Markov kernel. Numerous kernels could be constructed, and one of the most commonly used choices is a Gibbs sampler. Common for all these kernels are that a proposal for the density of the vector of parameters θ_{t+1} at time $t+1$ given the vector of parameters θ_t at time t is needed, i.e., an assumption about the dynamics of the parameters. In the work of Robert, a normally distributed artificial dynamic is considered with only marginally improved results.

Another approach could be to impose the artificial dynamics by considering two possible gains, i) to couple the parameters with those of the iterative Bayes' estimation of the states and ii) to provide top-down information of the parameters via a self-consistency condition. The idea is that, given a state of the system (x, y) at time t, it is in principle (numerically from Bertolace's Fokker-Planck formulation) possible to reverse engineer, which parameters θ_t^i are those that the observations up to time t can be associated with. It would be possible to get a set $i \in \mathbb{N}$ of different suggestions, and a straight forward

dynamic for a single parameter, e.g., α could be:

$$\alpha_{t+1} = \theta_t^{\alpha,i} + \eta(\alpha_t - \theta_t^{\alpha,i}) + \text{ noise}, \tag{3.1}$$

where $0 < \eta < 1$ is a damping parameter, and in case of multiple suggestions of $\theta_t^{\alpha,i}$ a horse race could be performed to get the estimate. This approach fulfils the criterion on coupling with the state estimation, and it also provides a top-down information on suggestions for parameters.

However, even if the idea is theoretically clear, there exist practical issues. The biggest challenge is how to reverse engineer the parameters efficiently. One suggestion would be to have a set of fixed parameters to choose from, perhaps the ones corresponding to different general market types presented by Yukalov et al., and compare the corresponding probability distribution function $p(x, y)$. This would be possible to implement, but would constrain the search to a limited part of the parameter space. A more dynamic approach on the parameter set would require solving the Fokker-Planck equation for every time step at the interesting positions, increasing the complexity tremendously. This has not been implemented, and one might get more insights on the problem by putting a higher effort on this kind of approach. Nonetheless, this work chooses to consider the methods presented in the next sections.

3.2 Evolutionary Algorithms

In contrast to the earlier proposed density estimation approach and the Kalman filter technique, the Evolutionary algorithms (EAs) are search heuristics that imitate the process of Darwinism and originate back to Fraser in 1957 [13]. Today EAs are widely used in e.g. computer science, artificial intelligence and optimization. These approaches are designed to solve the problem at hand more quickly than compared to classical methods, achieved by trading optimality and precision for speed. In the following sections, the concept of Evolutionary algorithms is introduced in a way applicable to the asset pricing model in this work, focusing on computational efficiency.

3.2.1 Overview of the Algorithm in View of the Asset Pricing Model

A naive approach to calibrate the asset pricing model of Yukalov et al. would be to apply a full-space search algorithm, i.e., try every possible combination of parameters. For this model, with six unknown parameters and two unknown initial conditions to be specified on rather wide continuous intervals, the search space would be overwhelmingly large. In fact, with only 100 different values for each unknown variable the number of different trails is $100^8 = 10^{16}$, a huge number.

Evolutionary algorithms are methods invented to omit this increasingly high complexity. Probably the most popular EA is the Genetic algorithm (GA), which is based on the idea of natural selection by Darwin in 1859 [12], and has an associated terminology. In the

case of the asset pricing model by Yukalov et al., every parameter α, A, β, B, μ, λ, and each of the initial conditions (x_0, y_0) are called *genes*, while the specific values of genes are named *alleles*. Strictly speaking, GAs are defined by genes coded in binary numbers, and Evolutionary strategies (ESs) have real valued genes. Solutions, or combinations of the unknown genes are referred to as *chromosomes* [6]. The fundamental idea of EAs is to have a *population* of chromosomes and evolve *good* solutions by natural selection. Thus, a measurement is needed to distinguish *good* solutions from *bad* solutions. Both subjective and objective value functions could be used, but for programmability an objective function is often chosen. Once this is done, the algorithm evolves solutions using the following schematic steps as presented in Burke et al. [6]:

- *Initialization.* The initial population of chromosomes is usually generated randomly across the intended search space. Domain-specific knowledge is easily incorporated in the set-up, and corresponds to setting a prior on the parameters and initial conditions of the asset pricing model.
- *Evaluation.* For the initialized population or any off-spring, the fitness of the chromosomes are calculated using the value function. For the asset pricing model, this would in general correspond to comparing observed market prices with simulated prices from each chromosome.
- *Selection.* The selection step represents the *survival-of-the-fittest* mechanism on the chromosomes, where better fitted solutions, in terms of the value function, are preferred.
- *Recombination.* From the pool of selected chromosomes, a recombination of two parental solutions are formed to create new, possibly better off-springs.
- *Mutation.* While recombination operates on parental chromosomes, mutation performs a random walk in the vicinity of a candidate solution to enlarge the parameter space reached by the population.
- *Replacement.* After the off-springs are created by selection, recombination and mutation, replacement of poorly fitting parents with better suited off-springs are done.

The algorithm repeats the steps including and followed by evaluation until some terminating condition is met, e.g., a maximum number of generations created. For an overview of the algorithm, see Algoritm 1. Goldberg [14] has shown that the operators selection, recombination, mutation and replacement are individually ineffective, but performs well when combined together. For example a combination of selection and mutation is said to give continual improvement, a form of hill climbing, and selection together with recombination allows for cross-fertilization. In the next section, these operators are introduced and explained more thoroughly.

Algorithm 1: Overview of a general Evolutionary algorithm.

```
Generate a population P randomly with n chromosomes
generation := 1
while generation ≤ gen_max do
    Clear the new population P'
    Evaluate each individual in P with a fitness function f(.)
    while P' ≤ n do
        Select two parents from P
        Perform crossover with rate p_c
        Perform mutation with rate p_m
        Insert offsprings in P'
    end
    Replace P with most fitted chromosomes in P and P'
    generation = generation + 1
end
Output: population P
```

3.2.2 Comparison Evolutionary Strategy and Genetic Algorithm

From a theoretical point of view, ESs and GAs are nearly identical. Both are direct, global search methods based on recombination and mutation of chromosomes in a population. Nevertheless, the difference in representation is highly important. Using binary chromosomes, GAs are easily implemented and recombination together with mutation are performed by switching parts of the binary strings between different chromosomes. Operating on ESs' real-valued chromosomes is not as straight forward as on the GAs' solutions, but is better suited for self-adaptive mechanisms achieving a higher rate of convergence [18]. However, self-adaptive methods have also been developed for GAs [24], where e.g. crossover and mutation rates are being adapted. Apart from the traditional denominational difference between GA and ES, parts of the literature and practice seldom distinguishes between binary and real-valued genes. In addition, papers on EAs are often named GAs. This work uses the terminology GAs even for real-valued chromosomes, which are to be studied.

3.2.3 Basic Genetic Algorithm Operators

In this section, the basic EA operators, selection, recombination, mutation and replacement, are presented in a similar fashion as in Burke et al. [6]. The methods presented in this section are highly intuitive and often used as the first mode of procedure. Moreover, the objective value function to distinguish fitness among candidate solutions needs to be specified. This will be done explicitly in Chapter 4.

Initially, the highly important size of the population needs to be determined. Small populations might lead to premature convergence at suboptimal solutions, while on the

other hand large population sizes increases the computational complexity. For the most basic methods, a reasonable number of chromosomes are selected from a computational point of view. In more intricate cases, the optimal number of chromosomes can be calculated, but the computational power is still a limiting factor. This work uses a fixed number of chromosomes, set by the limits on computational power.

Selection Methods

Basic selection methods are broadly classified in two classes, fitness proportionate selection and ordinal selection. The most classical fitness proportionate selection method is the roulette-wheel selection [19], which assigns a chromosome i a roulette wheel slot p_i sized in proportion to its fitness f_i. Usually, the choice $p_i = f_i / \sum_{j=1}^{n} f_j$ is made, where n is the number of chromosomes. The cumulative distribution is formed as $F_i = \sum_{j=1}^{i} p_j$ and the selection step is performed as generating n uniformly distributed random variables r. If $r_k < F_1$, then choose the first chromosome, else select the chromosome i such that $F_{i-1} < r_k \leq F_i$ for every $k \in \{1, ..., n\}$. These chromosomes form the mating pool.

In turn, ordinal selection chooses the fittest individuals in certain groups. For instance, groups could be formed randomly and compete towards each other, or the whole population could be one group. The advantage of using ordinal selection is that it is easy to implement on parallel machines. However, the complexity is higher than the roulette-wheel selection when there are multiple groups. Another drawback is that the number of groups and how many chromosomes to be replicated from each group needs to be determined. This is especially crucial when the whole population forms only one group, since there is a trade-off between the efficiency on choosing a few, good chromosomes and a lost of chromosome diversification if the replication pool is too small. In addition, it is possible to formulate the roulette-wheel selection to depend on only one random variable, called systematic resampling in particle filter terminology. This enhanced method is given in Algorithm 2, where N_i denotes the number of times chromosome i is selected.

Algorithm 2: Enhanced roulette-wheel selection method.

Input: $\{p_i\}_{i=1}^{n}$
`// Sample` r
$r \sim \text{Unif}(0, 1/n)$
for $i \in \{1, ..., n\}$ **do**
  $r_i = r + \dfrac{i-1}{n}$
end
`// Transform by` F_n^{-1}`:`
for $i \in \{1, ..., n\}$ **do**
  $N_i = \{\#r_j : \sum_{k=1}^{i-1} p_k < r_j \leq \sum_{k=1}^{i} p_k\} = \{\#r_j : F_{i-1} < r_j \leq F_i\}$
end
Output: $\{N_i\}_{i=1}^{n}$

Recombination Operators

From the mating pool, new, hopefully more suited, individuals are created via recombination, also called crossover. Most recombination operators randomly selects two chromosomes and these are crossed over with a probability p_c, the crossover probability. In practice, recombination is determined by generating a uniform random number on $[0, 1]$ and if $r \leq p_c$ then the chromosomes are recombined, otherwise not. The value of p_c is traditionally set experimentally. As an empirical rule of thumb, the crossover probability be should about $0.5 \sim 1.0$ [24]. Heuristically, a higher crossover probability leads to quicker exploitation of local minima, while a too high p_c disrupts chromosomes before they could be exploited. In this work, the most efficient values of p_c is tested to be 0.5.

In turn, to actually perform the crossover, the literature suggests numerous of different methods and it is also possible to develop problem specific methods. The simplest and most widely applied crossover method on binary chromosomes is the k-point crossover. The method is illustrated in Figure 3.1 for $k = 1$ and $k = 2$ respectively. In the one-point crossover ($k = 1$), a crossover point is randomly selected over the chromosome, and the alleles on one side of the site are exchanged. For the k-point crossover, k crossover points are generated and the alleles are exchanged accordingly.

In the case of real-valued chromosomes, k-point recombination is not particularly suitable and is easily understood by an example. Imagine a problem of two variables (u, v) increasing along the line $u = y$ with one optimum, say $(0.25, 0.25)$, and two chromosomes are at hand, $c_1 = (u_1, v_1) = (0, 0)$ and $c_2 = (1, 1)$. By 1-point recombination, non-trivial offspring are $c_3 = (1, 0)$ and $c_4 = (0, 1)$. These are worse than their parents, and the algorithm would only depend on mutation. The solution is a linear crossover, at each step constructing two out of three possible offspring by weighting the two parental chromosomes by one of the equally probable weight pairs $\{(0.5, 0.5), (1.5, -0.5), (-0.5, 1.5)\}$.

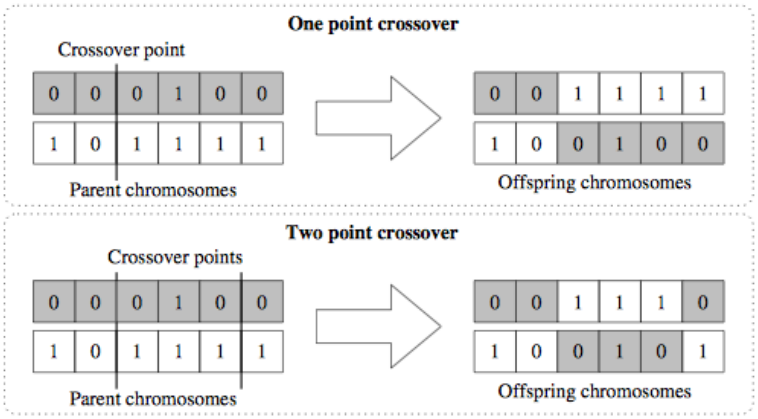

Figure 3.1: Schematic overview of the k-point recombination operator when $k = 1$ and $k = 2$ respectively.

Mutation Operators

Using the recombination operator, new generations of chromosomes are formed to get a better and better fitness. However, if parental chromosomes, or even worse the whole population, have the same allele at a given gene, then a simple k-point crossover would not change this. In other words, that gene will never change its allele. As a solution, the concept of mutation is introduced, and adds diversity to the population, which ensures exploration of a greater part of the search space.

In analogy with recombination, the mutation operation is determined by the mutation probability p_m. One of the most widely used mutations is the bit-flip mutation, where every allele is decoded as a binary string and each digit is altered with the mutation probability. Empirical studies show that p_m usually should be in the range of $0.001 \sim 0.05$ [24]. More tangible, p_m controls the speed of the GA in exploring new domains of the search space.

An improvement of the bit-flip mutation for real-valued chromosomes consist of methods for EAs based on parametric distributions [21], and examples include Gaussian, Cauchy and Adaptive Lévy mutation. The tail properties of the distributions allow for different characteristics. For instance, normal mutation has lighter tails than a Cauchy distributed mutation, leading to larger mutation, preventing premature convergence but giving a less accurate local search. In comparison, the Lévy mutation acts as an intermediary case, with both Gaussian and Cauchy mutations as explicit choices of the scaling parameter. What choice to make is a matter of taste and the properties of the fitness function. In this work, Gaussian mutation is considered since the study by Koenig [21] shows that for quadratic fitness functions (see. Chapter 4), the different mutation methods performed similarly. In addition, the mutation is of higher importance for chromosomes with real-valued genes, and it is often valid to use $p_m \sim 1$.

The Gaussian method operates on each chromosome $c_i = (\alpha_i, A_i, \beta_i, B_i, \mu_i, \lambda_i)$ for all $i \in \{1, 2, ...n\}$, i.e. $p_m = 1$, and an offspring is created by:

$$\sigma'_{i,j} = \sigma_{i,j} \exp(\tau' N_i(0,1) + \tau N_{i,j}(0,1)), \tag{3.2}$$

$$c'_i = c_i + \sigma'_i N_i(0,1), \tag{3.3}$$

where $j \in \{x_0, y_0, \alpha, A, \beta, B, \mu, \lambda\}$ and $N(0,1)$ denotes a standard Gaussian distributed random variable. The random variable $N_i(0,1)$ is common for all the genes in one chromosome. Moreover, Schwefel has suggested that [21]:

$$\tau = \left(\sqrt{2\sqrt{l_c}}\right)^{-1}, \tag{3.4}$$

$$\tau' = \left(\sqrt{2l_c}\right)^{-1}, \tag{3.5}$$

where l_c is the length of the chromosome vector, in this case $l_c = 8$. What still needs to determined is the values of $\sigma_{i,j}$, which in the most general case can differ between both

chromosomes and parameters. However, if there is no prior knowledge of the uncertainty of the different parameters it is reasonable to use a common variance for all object variables. Usually, this is set by trial-and-error, and testing in this work suggests the initial value $\sigma_{i,j} = 1$.

Replacement Methods

As the last step of each iteration of the GA, the fitness of offspring chromosomes are compared with the fitness of their parents. Even for this step, there exists a lot of different methods. A straight forward, efficient and parameter-free method is to simply choose the n most fit chromosomes as the next generation. More sophisticated approaches, such as checking for duplicates, could be used, but are all inferior regarding implementation speed.

3.3 Simulated Annealing Parameter Estimation

As seen earlier, one of the greatest concerns regarding parameter estimation is due to the fact that the probability distribution of the model parameters contained in the vector θ given the mispricing pair (x, y), denoted $p(\theta|x, y)$, is unknown. The previously studied sequential Monte Carlo methods try to estimate the distribution $\hat{p}(\theta|x, y)$ and extract information about the parameters sequentially. The Evolutionary algorithm is instead a search algorithm that uses the whole time-series and implicitly tries to find the optimum of $p(\theta|x, y)$. In this section, a new method for calibration using a search algorithm to explicitly estimate the probability distribution $p(\theta|x, y)$ is suggested, which consequentially is used for parameter estimation.

3.3.1 Intuition and Background of the Algorithm

The idea behind the Simulated annealing parameter estimation apporach is a problem formulation by the Nobel laureate de Gennes in statistical mechanics called 'Ant in a labyrinth' and concerns percolation. In this description, he used the terminology ants for random walkers diffusing through porous materials and used them to calculate the probability distribution of ants inside the material. From this set-up, an analogy can be made to construct an algorithm useful for parameter estimation. In this case, the parameter dependent landscape imposed by the model's fitness works as the porous material and the ants should diffuse over this landscape. This diffusion is to be governed by the properties of the fitness over the parameter space. Starting with many ants distributed uniformly over the parameter space, the steady-state location of ants should approximatively be distributed as $p(\theta|x, y)$. As discussed before, the observation from real financial markets are assumed directly to be the mispricing component x. Thus, this probability distribution will be denoted $p(\theta|x)$.

The name Simulated annealing was introduced by Kirkpatrick et al. [20], linking optimization and statistical mechanics for deterministic problems. Their work has later been enlarged to methods for parameter estimation of stochastic systems [17]. The key insight in the work of Kirkpatrick et al. is the introduction of mechanisms from statistical mechanics and regards the diffusion of random walkers in parameter space. Given a fitness function, a simulated time-series from a point in the parameter space and an observed mispricing time-series, the fitness could be interpreted as a potential energy. Thus, in statistical mechanics, the system (the position of the random walker) could be weighted by its Boltzmann probability factor, $\exp(-V(\theta)/k_BT)$, where $V(\theta)$ is the potential of the random walker with parameter set θ [20]. The classical criterion for changing the state is proposed by Metropolis et al. [27], the Metropolis rule, and accepts the new state with probability $\min(1, \exp(-\Delta V/k_BT))$. This means that states with lower energy are always preferred to higher energy states, but also higher energy states are possible and introduced to avoid the problem of getting stuck in a state that is only a local minimum, not the global. By repeatedly constrain random walkers with the Metropolis rule, one simulates the thermal motion of atoms in thermal contact with a heat bath with temperature T. This yields that the system evolves into the Boltzmann distribution.

The free parameter in the Metropolis rule, the temperature T, determines the sensitivity to energy increases. In the limit $T \to 0$, the system freezes and no changes occur. The annealing process consist of first optimizing the system at a high effective temperature and then lower the temperature until freezing occurs. It is important that the procedure continues long enough for the system to reach steady state, to be able to estimate $p(\theta|x)$ accurately.

3.3.2 Description of the Algorithm

The details of the diffusion of one random walker in the parameter space is specified in Algorithm 3. The crucial decisions in the implementation of the algorithm introduced in the previous section concern the normalization used in the potential, i.e., the corresponding value to k_B in the statistical mechanics analogy, and the search method for suggestion of the new position of the random walker in the parameters space.

Concerning the normalization, a quantity representative for the fitness landscape is necessary for the Boltzmann probability to be useful. If the normalization constant is too large, the random walker will walk freely in the parameters and is not stopped from climbing the potential and vice versa, cf., the intuition of the free temperature parameter T. For example, the average of the initial potential of all of the ants can be used. In this thesis, an adaptive choice of normalization constant is used via the potential of the old position of the ant. The advantage is that it will be relatively more difficult for the random walker to climb the potential when the potential is low, i.e., closer to the optimum, which makes it harder for the ant to leave this position.

For efficiency, the suggested new position of the random walker is chosen via Monte-

Carlo sampling. This is done by sampling a vector r on a hypersphere in the parameter space. The length of the radius of the hypersphere is adjusted through analyses of the performance of the diffusion, and chosen to be $r = 0.4$. Likewise, a plausible set of temperatures and number of steps K at each temperature need to be adjusted for the annealing method to be useful. These parameters are also set by hand-on tuning.

Algorithm 3: Simulated Annealing Parameter Estimation Algorithm

Input: $\{x_t\}_{t=0}^{T}$ `// Mispricing time-series`
`// Initialize position of random walker, i.e. sample` θ
$\theta \sim$ Uniformly on parameter space
`// K is the number of steps to obtain steady-state`
for $T \in \{T_1, ..., T_n\}$ **do**
  for $i \in \{1, ..., K\}$ **do**
    $V = f(\theta, \{x_t\}_{t=0}^{T})$ `// Potential (fitness) of random walker`
    `// Monte-Carlo sampling of suggested position`
    $r \sim$ Uniformly on hypersphere in parameter space
    $\theta' = \theta + r$
    $V' = f(\theta', \{x_t\}_{t=0}^{T})$
    $p = \exp(-\frac{1}{T}\frac{V' - V}{V})$ `// Boltzmann probability`
    $s \sim \text{Unif}(0, 1)$
    if $s < \min(1, p)$ **then**
      $\theta = \theta'$
    end
  end
end
Output: θ `// Position of random walker in parameter space`

3.3.3 Simulated Annealing Calibration Step

After the diffusion of ants, a set of points in the parameter space is obtained and from these the joint probability distribution $p(\theta|x)$ can be estimated. In this work, two different methods are considered, an adapted kernel estimation method as well as the k-Nearest Neighbour (kNN) density estimator. The adapted, multivariate kernel estimation is used to smooth the ants' positions to be able to estimate the probability distribution function. In detail, the kernel estimation technique in d-dimensions can be constructed as a product of d univariate kernels with independent smoothing parameters. This means that the estimate of the probability distribution $\hat{p}_0(\theta|x)$ is given by:

$$\hat{p}_0(\theta|x) = \frac{1}{n_a h_1 ... h_d} \sum_{i=1}^{n_a} \prod_{j=1}^{d} K\left(\frac{\theta_j - \theta_j^{a_i}}{h_j}\right), \tag{3.6}$$

where n_a is the number of ants, K the univariate kernel distribution function with

smoothing parameters $h_1, ..., h_d$ and for ant a_i, $\theta^{a_i} = (y_0^{a_i}, \alpha^{a_i}, A^{a_i}, \beta^{a_i}, B^{a_i}, \mu^{a_i}, \lambda^{a_i})$. A well understood and studied choice of kernel K is the standard Gaussian density, i.e.:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2). \tag{3.7}$$

Another useful choice is the Epanechnikov kernel, which (as mentioned earlier) can be proved to be optimal in a mean variance sense [33]. In this thesis, the Gaussian kernel will be used, which gives positive definiteness, infinite differentiability and is defined on an infinite support. This means that the estimate $\hat{p}_0$ is smooth and well-defined in the tails. The other part of the kernel, the bandwidth h, specifies the scale of the smoothing and is determined by the data. In the limit of normally distributed data, the optimal bandwidth in a mean integrated squared error sense is given by:

$$h_j^* = \left(\frac{4}{d+2}\right)^{1/(d+4)} \sigma_j n_a^{-1/(d+4)}, \tag{3.8}$$

where σ_j is the estimated standard deviation of the parameter j. However, for grounds of self-consistency, h_j^* should be determined by local quantities, but σ_j is a global quantity. Thus, the adapted kernel estimation method includes a second step of the estimating procedure $\hat{p}_1(\theta|x)$, with an adapted bandwidth. In addition, as proposed by Cranmer [11], the covariance structure of the data needs to be considered. This means that the diagonal covariance structure of the kernels might not match with that of the random walkers'. A linear transformation that diagonalize the covariance matrix Σ specified by the data (given by $D = A\Sigma A^T$), can be applied to mitigate this effect. The following optimal bandwidth in the Gaussian case is:

$$h_{ij}^* = \left(\frac{4}{d+2}\right)^{1/(d+4)} n_a^{-1/(d+4)} \left(\frac{\sigma_j}{\sigma}\right) \sigma^{(1-d'/d)} \hat{p}_0^{-1/d'}(\theta_i), \tag{3.9}$$

where $\sigma = \det(A\Sigma A^T)^{1/2d}$ is the geometric mean of the standard deviations of the transformed data and d' the effective dimensionality. d' could be determined by performance measurements, but is often approximately given by d, therefore $d' = d$ is chosen. Thus, with h_{ij} given by Equation (3.9), the estimate of the probability distribution function $p(\theta|x)$ is given by the analogue of Equation (3.6) for $\hat{p}_1(\theta|x)$.

In contrast to the adapted kernel estimation, the k-Nearest Neighbour method uses the local positions of the k ants closest to the considered ant (including the ant itself). Luenberger and Woehrmann [25] have shown that an unbiased estimate for the probability distribution $p(\theta|x)$ is given by:

$$\hat{p}(\theta|x) = \frac{k-1}{n_a V(r_k)}, \tag{3.10}$$

with $2 \leq k < n$ where r_k is the distance to the k-th nearest neighbour from θ and $V(r_k)$ is the volume of the hypersphere with radius r_k, given by:

$$V(r_k) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} r_k^d. \tag{3.11}$$

Furthermore, for $k > 2$, the variance of the estimate is given by:

$$\mathrm{Var}(\hat{p}(\theta|x)) = \left(\frac{(k-1)(n_a-1)}{(k-2)n_a} - 1 \right) \hat{p}(\theta|x)^2. \tag{3.12}$$

Another issue with the kNN-method is the choice of k. As can be seen in Equation (3.12), the variance of the estimate vanishes only if both $n_a \to \infty$ and $k \to \infty$. Nonetheless, if a Monte Carlo estimate of the probability density is used, the variances goes to zero and it is even possible to find the optimal number (related to computational efficiency and accuracy) of neighbours k. However, since the generation of ants in steady-state is computationally heavy, it is not reasonable to re-run the program to get a Monte Carlo estimate for which the optimal number of neighbours can be specified. Nevertheless, Luenberger and Woehrmann have shown that the most efficient value of k will be low ($\sim$ 5-15), hence values in this span are considered.

Finally, the estimates of the parameters of the asset pricing model is given by maximizing the estimated posterior probability distribution function $\hat{p}(\theta|x)$ for the adapted kernel and the kNN estimate, respectively.

Chapter 4

Calibration of Synthetic Data

This chapter begins in Section 4.1 with preliminary assumptions needed for model calibration on synthetic data. Thereafter, the asset pricing model is analysed in terms of Sloppy Models in Section 4.2, and the parameter estimation results from applying the Evolutionary algorithm and Simulated annealing methods are presented and analysed in Sections 4.3 and 4.4, respectively.

4.1 Simulation Preliminaries

In this section, necessary further assumptions on the parameter space, initial conditions and the calibration methods are made. Furthermore, market types that are selected for the purposes of this study of the asset pricing model are presented more thoroughly than in Section 2.1.3.

4.1.1 Value Function

A critical component for the implementation of Evolutionary algorithms and Simulated annealing is the choice of the value function or potential (in Evolutionary algorithm and Simulated annealing terminology respectively). As implied by the work of Koenig [21], the combination of value function and mutation operator is highly affecting the performance of the convergence of the Evolutionary algorithm. In this work, the squared L^2-norm of the difference between the model of the mispricing component x^{model} and the observed value x^{obs} will be chosen as value function and potential. To smooth the simulated mispricing x^{model}, the average at every time-point of 5 simulations is used. That is, for all considered times steps $\{0, 1, ..., T\}$, the value function for chromosome/ant i is given by:

$$f_i = \sum_{j \in \{0,1,...,T\}} |x_i^{obs}(j) - x_i^{model}(j)|^2. \tag{4.1}$$

This value function corresponds to a classical least squares measure, which is a well understood approach and assigns a reasonable weight to deviations. As a consequence

of this choice of value function, the objective of any Evolutionary algorithm is to reach as small values as possible of f for all the chromosomes in the population. Thus, to be able to use the roulette-wheel selection operator, the probability p_i corresponding to chromosome i is instead constructed by:

$$p_i = \frac{1/f_i}{\sum_{j=1}^{n} 1/f_j}. \tag{4.2}$$

The inverse-roulette-wheel operator assigns high probability to chromosomes with a better fit, i.e., lower f_i, and vice versa. In the limit, a perfectly fitting chromosome gets probability 1. Moreover, the choice of f_i as potential in Simulated annealing is straight forward and transferees naturally without any need of modifications.

4.1.2 Prior-Distributions of Parameters and Initial Conditions

To generate the initial population of chromosomes and initial positions of ants, composed of real values for the initial conditions (x_0, y_0) and the parameters α, A, β, B, μ and λ, assumptions on how to sample in the parameter space must be made. Ideally, a joint prior sampling the chromosomes/ants simultaneously would be used. However, no such prior is obvious and other approaches are needed.

Nevertheless, there is some domain-specific knowledge available, restricting the parameter space somewhat. Using the asset pricing model with constant parameters implies that e.g., β is negative and B is positive. In addition, both α and A cannot be positive, since it would create an unsustainable market condition. With this information at hand, the chosen prior-distribution is to sample most parameters independently and uniformly distributed over reasonable ranges of values. Hopefully, this will introduce enough diversity among the chromosomes to have a good starting population, as well as distribute initial positions of ants properly. In theory, the algorithms should be applicable even with poor starting populations, but with slower convergence as a consequence. The chosen ranges for the different parameters and initial conditions are shown in Table 4.1. The joint sampling of α and A is constructed by first sample between the three different valid situations $\{\{\alpha > 0, A < 0\}, \{\alpha < 0, A > 0\}, \{\alpha < 0, A < 0\}\}$ with equal probability. In addition, note that the signs of both μ and λ are irrelevant because of the symmetric dependence in f^{NL}, cf. Equation 2.4.

The simulation trajectories are sampled over a time period of 7 years, which is chosen as a trade-off between having a sufficiently long time horizon to incur changes in mispricing behaviour and computational complexity. For every point in the parameter-space and in every iteration, the fitness function f is calculated. Thus, if the time horizon is too long, the simulation of time-series and computation of the fitness will be computationally very heavy.

Table 4.1: Prior-distributions of unknown parameters and initial conditions. Note: Both α and A cannot be positive and are sampled jointly to avoid this case.

Unknown Variable	Range	Sampling Properties
x_0	$\{-3, 3\}$	Independently, uniform
y_0	$\{-3, 3\}$	Independently, uniform
α	$\{-5, 0\}$ or $\{0, 5\}$	Jointly with A, uniform
A	$\{-12, 0\}$ or $\{0, 12\}$	Jointly with α, uniform
β	$\{-5, 0\}$	Independently, uniform
B	$\{0, 5\}$	Independently, uniform
μ	$\{0, 6\}$	Independently, uniform
λ	$\{0, 6\}$	Independently, uniform

4.1.3 Studied Market Types

For the simulations in this work, a subset of the market types in Section 2.1.3 are chosen and quantified. To keep track of the different market types, identifiers are introduced. The market identifiers are in correspondence with the work by Yukalov et al. [35], where each identifier refers to a number of a figure in their paper. These graphs show the associated phase portrait for the deterministic system. Once again, studies of these are highly recommended to get a deeper understanding of the asset pricing model. For interpretation of the parameters, see Section 2.1.3.

The chosen markets are presented in Table 4.2 and could also be found in Table 2.1. These choices are made to test the calibration procedures on a range of different market complexities. Firstly, market 1 represents the most simple equilibrium market with individual and collective mean-reverting strategies, over regulation and rather certain knowledge about the fundamental price. Secondly, market 10 includes individual speculative traders balanced by a collective mean-reversal strategy in a softly regulated market with high uncertainty about fundamental price. This market has two different attractors and a limiting cycle. Lastly, market 15 is somewhat reversed with individual mean-reverting and collective speculative behaviours, representing a situation where some groups of traders can create big market changes with their speculations. These market conditions allow for three different attractors of the mispricing.

4.2 Sloppy Model Analysis

To get a deeper understanding of why the performance of the past calibration procedures (and results in this thesis) are far from perfect, the asset pricing model is analysed in the context of Sloppy models introduced by Brown et al. [7]. A Sloppy model is a multi-parameter model, whose behaviour depends only on a few *stiff* combinations of parameters, and there are many *sloppy* directions which are rather unimportant. A well-

Table 4.2: Parameter values of market types studied in this work.

Market	α	A	β	B	μ	λ	Fixed Points $\{(x,y)\}$ Cycles $C_{(x,y)}$
1	-1	-10	-1	1	1	1	$\{(0,0)\}$
10	5	-1	-1	1	5	3	$\{(\pm 3,0)\}, C_{(0,0)}$
15	-1	1	-1	1	2	1	$\{(0,0),(\pm 3,0)\}$

known problem with multiple parameters models is *ill-conditioning*, i.e., that different parameter sets can exhibit similar behaviour, which Brown et al. denote the sloppiness of the model. For a model designer, ending up with a Sloppy model is displeasing, since the interpretation of the model parameters is rather meaningless in the sloppy directions. Nevertheless, the model could still be very useful for predictions. There are numerous models that are sloppy but still very precise. One example is quantum Monte Carlo modelling of variational wave-functions for high-accuracy molecular energy calculation, and other examples are found in system biology. In this section, the asset pricing model is formulated in a way that admits a sloppiness analysis.

4.2.1 Deterministic Model Analysis

Since the introduction of Sloppy models in 2004, mostly deterministic models from system biology have been analysed. Only recently, published in November 2013, a method to investigate the sloppiness of stochastic models has been introduced [26]. However, this approach is not directly applicable to the stochastic model of Yukalov et al., since the model lacks knowledge about the probability density $p(x,y|\theta)$ and more precisely the Hessian of it. Nevertheless, it is possible to find a description of the deterministic counterpart of the asset pricing model, which allows the use of parts of the deterministic framework. Since the model-stochasticity is in form of driving Wiener-processes, an analysis of the deterministic system would be highly informative in the sense that the stochasticity alters the behaviour of the stochastic model between different phases of the underlying deterministic dynamics. Thus, understanding the properties of the deterministic model implies foundations for what to be expected from the stochastic system.

For sloppiness analysis, a cost function $\chi^2(x_t^{model}, y_t^{model}, x_t^{obs}, y_t^{obs}, \theta)$, where in the most general setting $\theta = (x_0, y_0, \alpha, A, \beta, B, \mu, \lambda)$ corresponds to the parameters to be estimated, is considered. The cost function measures the deviation of the model from observations and gives rise to sub-domains of the parameter space where the cost is constant. In the case of two unknown parameters, it is possible to visualize this by constructing contour plots, see Figure 4.2. On sufficiently small scales, the level curves around the optimal point, i.e., with minimal cost, will form ellipsoids. The sloppiness at this point in the parameter space is then measured in terms of the eigenvalues and the eigenvectors of the cost function's Hessian evaluated at this point, also denoted the Fisher Information

Matrix. The eigenvectors point along the axes of the ellipsoid and the lengths of the axes are proportional to one over the square root of the corresponding eigenvalue. Thus, eigendirections with large eigenvalues represents stiff directions, i.e., where the cost is sensitive to changes in the parameters associated with this direction. The Hessian matrix is a local quadratic approximation to the cost function, and Brown and Sethna [8] conclude from a Principal component analysis of extensive Monte Carlo simulations that the sloppiness of the Hessian is indicative for the full cost function.

One possible choice of cost function is the value function given by Equation (4.1). However, the mispricing x^{model} does not explicitly depend on the vector of parameters θ. Instead, the analogous deterministic Euler-forward discretization of Equations (2.8) and (2.9) reads as:

$$\bar{X}_t = \frac{X_{t+1} - X_t}{\Delta t} = Y_t, \tag{4.3}$$

$$\bar{Y}_t = \frac{Y_{t+1} - Y_t}{\Delta t} = \frac{\bar{X}_{t+1} - \bar{X}_t}{\Delta t} = f^{NL}(X_t, Y_t). \tag{4.4}$$

Thus, with this formulation, it is on one hand possible from observations to construct $\bar{Y}$ and on the other hand calculate it from the model via $f^{NL}(X_t, Y_t)$, which explicitly depends on the unknown parameters θ given that Y_t is observable or possible to estimate with high precision. As a first approximation, assume that Y_t is observable. This assumption will later be examined in Section 4.2.2 using a Savitsky-Golay filter. Moreover, if explicit knowledge about x_0 and y_0 is assumed, then the cost function can be chosen as:

$$\chi^2(\alpha, A, \beta, B, \mu, \lambda) = \frac{1}{2} \sum_{t \in \{0,1,...,T-1\}} (f^{NL}(X_t, Y_t) - \bar{y}_t^{obs})^2, \tag{4.5}$$

with a slight simplification of notation with only the unknown parameters explicitly stated and the time steps denoted by $\{0, 1, ..., T\}$. Consequently, the Hessian is given by:

$$H^{\chi^2}_{i,j} = \sum_{t \in \{0,1,...,T-1\}} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_i} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_j}, \tag{4.6}$$

where $i, j \in \{\alpha, A, \beta, B, \mu, \lambda\}$. The full derivation and explicit expression for the derivatives care found in Lemma A.1.1 and Section A.2, respectively. From the above equations, the sloppiness of the asset pricing model is easily calculated in terms of the eigenvalues λ^E of the Hessian for different markets with changing initial conditions. Different initial conditions are chosen to pin-point the discrepancy between in and out-of-equilibrium dynamics and in particular to match with the real financial time-series studied in Chapter 5.

However, for direct transferability of the results of the deterministic sloppiness analysis to the stochastic case, the vanishing of the second order terms in the stochastic analogue of Lemma A.1.1 needs to be motivated. If the second derivative of f^{NL} is constant, this

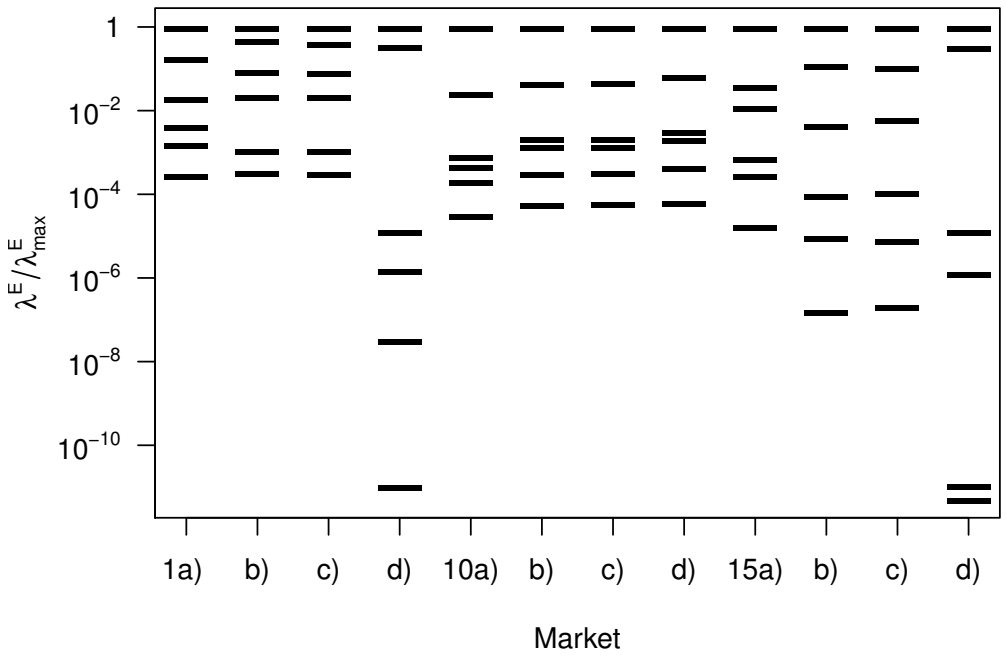

Figure 4.1: Overview of sloppiness structures for the deterministic model for market 1, market 10 and market 15 with initial conditions a) $(x_0, y_0) = (1.5, 0.2)$, b) $(x_0, y_0) = (0.8, 0.1)$, c) $(x_0, y_0) = (0.8, -0.1)$ and d) $(x_0, y_0) = (0.1, 0.1)$. Eigenvalues are normalized by the largest eigenvalue λ^E_{max} for every market and initial conditions.

term would be negligible by the law of large numbers. But, this might be too much to ask because of the strongly non-linear behaviour of f^{NL}, and further analyses should be made to determine this. Investigating the properties of f^{NL} in more detail is, however, beyond the scope of this thesis. Therefore, the deterministic sloppiness analysis is assumed to be declaratory even for the stochastic system.

As can be seen in Figure 4.1, the quotients of the eigenvalues of the Fisher Information Matrix ranges from 10^{-4} to 10^{-11}, showing that the asset pricing model has sloppy directions in the parameter space. Especially market 15 seems to be particularly sloppy, implying even more difficulties in calibrating more sophisticated markets. Another interesting observation is the dependence of the initial conditions (x_0, y_0), or more specifically the distance from equilibrium. For market 1 and market 15, with an attractor at the equilibrium point $(x, y) = (0, 0)$, the initial point d) $(x_0, y_0) = (0.1, 0.1)$ has a sloppier structure then the far from equilibrium initial point point $(x_0, y_0) = (1.5, 0.2)$. On the other hand, market 10 has no attractor at the equilibrium point $(x, y) = (0, 0)$ and exhibits not such a change in sloppiness. This implies that the calibration of the asset pricing model is highly dependent of the access to information about out-of-equilibrium dynamics. It is clear in the case of the deterministic market 1 in equilibrium, since then $f^{NL}(0, 0) = 0$ and the mispricing is always in equilibrium. That is, there is no informa-

Table 4.3: Overview of the stiffest and the sloppiest directions as well as the alignment ratio for the different bare directions for market 1, market 10 and market 15 with initial conditions a) $(x_0, y_0) = (1.5, 0.2)$, b) $(x_0, y_0) = (0.8, 0.1)$, c) $(x_0, y_0) = (0.8, -0.1)$ and d) $(x_0, y_0) = (0.1, 0.1)$. Note: When a direction is given in parenthesis, the difference in size of the corresponding eigenvalues are small.

Market	1a)	b)	c)	d)	10a)	b)	c)	d)	15a)	b)	c)	d)
Stiff	μ (β)	β (μ)	β (μ)	α	A	A	A	A	α	α	α	α
Sloppy	A	A	A	λ	μ	μ	μ	μ	λ	λ	λ	μ (λ)
I_α/P_α	0.34	0.34	0.35	0.29	0.06	0.12	0.12	0.12	0.38	0.19	0.23	0.28
I_A/P_A	0.26	0.15	0.16	0.08	0.07	0.12	0.12	0.16	0.37	0.09	0.11	0.08
I_β/P_β	0.48	0.30	0.30	0.39	0.40	0.67	0.65	0.87	0.32	0.31	0.28	0.42
I_B/P_B	0.59	0.14	0.14	0.12	0.39	0.52	0.52	0.52	0.12	0.08	0.08	0.15
I_μ/P_μ	0.69	0.25	0.26	0.11	0.12	0.16	0.17	0.17	0.22	0.12	0.14	0.10
I_λ/P_λ	0.44	0.25	0.25	0.18	0.39	0.50	0.50	0.50	0.19	0.11	0.11	0.22

tion for the calibration of the parameters. This is also shown empirically in this work using the different parameter estimation approaches.

Furthermore, after establishing that the asset pricing is sloppy, it is of highest interest to find the sloppy and stiff directions. An imprecise approach is to consider the eigenvectors corresponding to the smallest and largest eigenvalues, respectively, and to subsequently infer in which directions the model is sloppy. To enhance this analysis it is possible to define a measure of alignment by studying the ratio of the projection P_i of the hyper-ellipsoid on the bare directions (the parameter axes) and the intersection I_i of the bare direction and the ellipsoid for $i \in \{\alpha, A, \beta, B, \mu, \lambda\}$. This alignment ratio is mostly governed by the direction of the sloppiest direction, i.e., in the direction where the ellipsoid is largest. Ideally, this direction is aligned with one of the bare directions (or axis directions in the parameter space), which would imply that all the other bare directions are rather stiff and should be easier to calibrate.

The results of this analysis is shown in Table 4.3. For market 1a), the stiff direction is rather aligned with the bare direction (μ), which also can be seen in the contour plot shown in Figure 4.2. Likewise, Figure 4.3 illustrates that the stiff and the sloppy directions are not well-aligned for market 10 (note that the logarithm of the fitness function $\log(f)$ is used). Table 4.3 also suggests that the most stiff and the most sloppy directions are associated with the same bare directions apart from when the sloppiness changes drastically, i.e., near the equilibrium mispricing. Unfortunately, the sloppy directions are generally badly aligned with the corresponding bare directions, which implies less precision in the parameter calibration due to larger projections in all the other bare directions. Nonetheless, the alignment ratio is never extremely low ($< 10^{-2}$) in any of the directions, which is a good sign. This means that the sloppy direction is not too badly

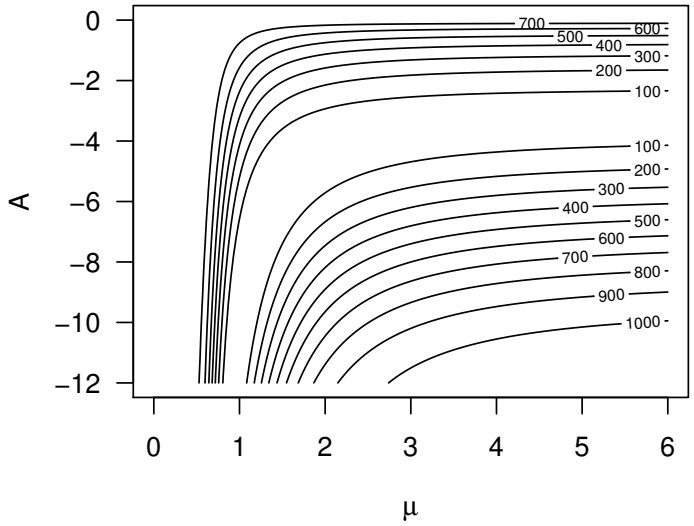

Figure 4.2: Contour plot in the sloppy A-direction and the stiff μ-bare direction around parameter optimal $(A, \mu) = (-10, 1)$ of the fitness landscape given by the function f in Equation 4.1 for market 1 with $(x_0, y_0) = (1.5, 0.2)$.

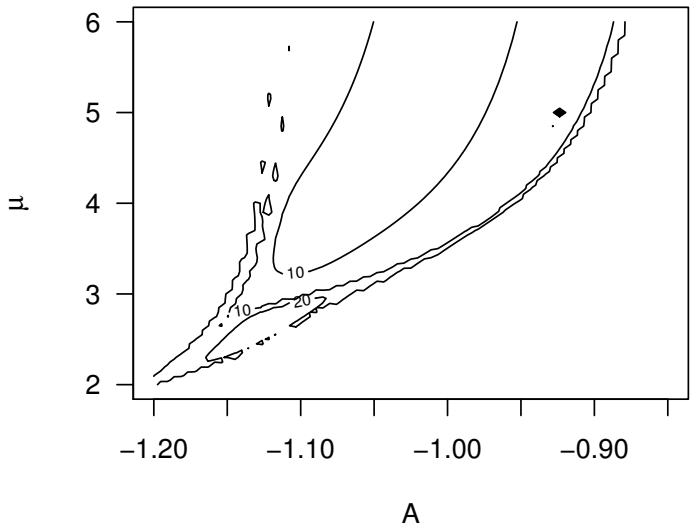

Figure 4.3: Contour plot in the sloppy μ-direction and the stiff A-bare direction around parameter optimal $(\mu, A) = (5, -1)$ of the fitness landscape given by the logarithm of the function f in Equation 4.1 for market 10 with $(x_0, y_0) = (0.8, 0.1)$.

aligned with a bare direction. At least not in comparison with the existing sloppiness in these directions. In particular, this suggests that it should be possible to estimate the parameters in the stiff directions fairly well, maybe with exception for market 10 with very small alignment ratios. These insights are investigated in the subsequent sections on parameter estimation.

In this context of identifying properties of the asset pricing model, it is interesting to pose the question on model construction with regard to Sloppy models. The work by Transtrum et al. [32] discusses this topic using differential geometry. In summary, Transtrum et al. discuss the difference in sloppiness when the parameters of a model is re-scaled. Imagine an analogue graph to Figure 4.2 with perfect ellipses around the optimal parameter point. By re-parametrization, the landscape can be transformed into circles, removing the old sloppiness structure. In consequence, the model constructor questions if this is always possible, i.e., whether sloppiness is a characteristic of the model that could be tuned away. This is briefly studied in Section 4.3.4 by considering a new parametrization. In further studies, a more thorough analysis, including a calculation of the corresponding sloppiness of the re-parametrized, system should be conducted.

Moreover, the work by Transtrum et al. studies the question whether sloppiness is intrinsic or not, by viewing the model geometrically as a function mapping from the parameter space to the model's prediction space. By varying the parameter vector, the model sweeps out a hyper-surface embedded in the prediction space, called the *model manifold*, and it has the same structure irrespective of the parametrization. The intrinsic sloppiness can now be studied by investigating geodesics, i.e., paths on the model manifold that are the closest approximations to straight lines that stay on the manifold, along the eigendirections of the Fisher Information Matrix. Thus, if the manifold's widths are large the sloppiness of the model is an intrinsic property corresponding to the physical limits of the prediction space. In this thesis, the widths of the model manifold are not studied, since the objective is to investigate the given asset pricing model by Yukalov et al. Nonetheless, a study of this type would be highly interesting for further work, since it could both indicate in what bare directions the sloppiness could be reduced by re-parametrization, and show the intrinsic sloppiness structure.

4.2.2 Filtering of the Mispricing Component y

In Equation (4.3), Y_t is given by the difference quotient of X_{t+1} and X_t. Thus, in the deterministic case, it is possible to calculate Y_t from the observable time-series of x. On the other hand, the stochastic system includes noise that does not allow for direct calculation of the time-series of y. In this section, the Savitzky-Golay filter is applied to the estimated values of $\hat{Y}_t$ to smooth the data in order to estimate the time series y. The Savitzky-Golay filter fits, by linear least squares, successive sub-sets of adjacent data points to a low-degree polynomial. There are off-the-shelves packages available, e.g., in MATLAB, with an implementation of the method, which is used in this work. For more details about the algorithm, see the original work by Savitzky and Golay [29].

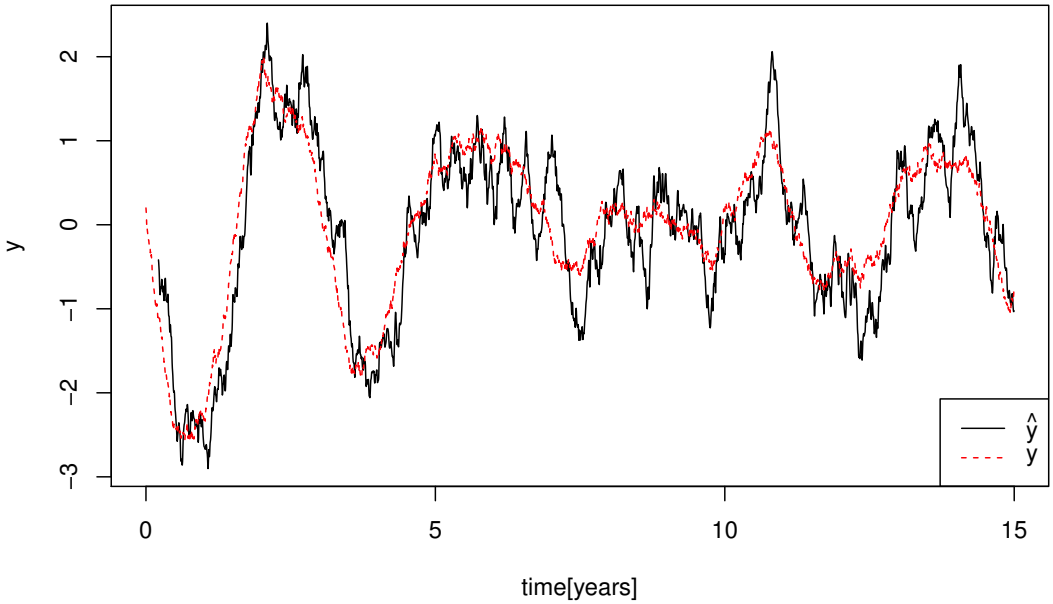

Figure 4.4: Smoothing of one estimated time-series of y, i.e. of $(X_{t+1} - X_t)/\Delta t$, for market 1 with $(x_0, y_0) = (1.5, 0.2)$ using a Savitzky-Golay filter with a polynomial of order 13 and 21 data points twice, and a moving-average smoothing between the application of the filters with 45 data points. The red line corresponds to the true values of y.

In Figure 4.4, the estimated time-series of y for market 1 with $(x_0, y_0) = (1.5, 0.2)$ is illustrated. The smoothing uses a Savitzky-Golay filter with a polynomial of order 13 and 21 data points twice, and a moving-average smoothing between the application of the filters with 45 data points. The smoothing seems to be rather good, especially when there are large changes in the mispricing x, i.e. when y deviates from 0. Closer to $y = 0$, the smoothing is poorer but still satisfactory. These preliminary results suggests that, in principle, it is possible to obtain y given x, irrespective of any knowledge about parametrization of the asset pricing model. Previous Master theses have shown good filtering results when either the parameters are estimated or known. These results are sufficient for the further use in this work.

4.3 Evolutionary Algorithm Calibration Results

In this section, the results from the Evolutionary algorithm are presented, including both the full evolution of the chromosomes and a more efficient approach concerning the final set of chromosomes only. Moreover, with background in the Sloppy model analysis, the

predictive power of the calibrated model is investigated as well as a re-parametrized version of the model.

4.3.1 Evolution of Generations

The first insight from running the Evolutionary algorithm for calibration is the dependence of the chosen starting point of the synthetic time-series, which was expected given the results from the Sloppy model analysis. The closer the initial conditions are to equilibrium, the less information about the market can be extracted and the calibration results are poorer. Thus, the starting point $(x_0, y_0) = (1.5, 0.2)$ is chosen as a reasonable, but substantial mispricing. If it is not possible to calibrate this model, it would be extremely more difficult to calibrate a market showing a less out-of-equilibrium behaviour.

In the first approach, the evolution of the chromosomes in every generation is studied, and the results are reported in Figures 4.5-4.7 and in Appendix B.1.1. In each figure, the evolution of the 10% fittest chromosomes are shown for 50 generations. Moreover, the initial condition x_0 is included in the calibration notwithstanding that, by assumption, it could be observed. Doing so, however, helps to improve the understanding about how the sensitivity to discrepancies in this initial condition reflects the calibration, and is only considered for the Evolutionary algorithm. In Chapter 5, where a real financial time-series is studied, the level of observability of x in practice is examined in more detail. From the figures it is obvious that the precision of the estimation of x_0 is far from perfect, but it is not too bad either. If x_0 would have been estimated perfectly it would most likely have corresponded to a stiff direction, which implies, that the model is highly dependent on this estimation, and the conclusion would be that the observability of the mispricing in the real market is very important. Hence, these results show that the model seems not to be super-sensitive to the ability of observing the correct mispricing x_0 in the market. In addition, the other initial condition y_0 is poorly estimated, and this is what we expect close to $y = 0$, cf., Section 4.2.2.

For market 1, Figure 4.5 illustrates a moderately satisfactory estimation of the parameters. For instance, the stiff parameter μ is not estimated correctly and the algorithm seems to diverge from the correct estimate for this parameter. The explanation might be found in Figure 4.2, where it is clear that the "stiffness" of the μ-direction (this quantity is not well-defined at points where the fitness function is not at a minima) is heavily dependent on the A-level, and since A is not perfectly estimated (in fact it is the sloppiest direction) it is not surprising that μ cannot be well-estimated. These properties of the fitness landscape make the parameter estimation problem even more difficult.

Moreover, the deterministic counterpart of market 1 illustrated in Figure 4.6 provides further information about this calibration approach. The population of chromosomes fast (after about 5 generations) turns out to be homogeneous at sub-optimal parameters. However, the mutation of the chromosomes introduces a fitter chromosome after a while (22 generations), and this chromosome accedes the population after some additional time.

Apart from showing the superiority of the mutation step in an Evolutionary algorithm to a greedy hill-climbing search, the deterministic case also indicates that the underlying deterministic system is more sensitive to the parameter values, since few chromosomes take over the whole population. This supports that the stochasticity of the model acts blurringly on the mispricing paths, as expected by the Wiener-processes driven dynamics. Furthermore, this indicates that the large size of the parameter space aggravates the calibration. Introduction of the optimal chromosome after a while changes the population to the perfect alleles quickly, i.e., the algorithm is highly dependent on the soundness of the mutation operator.

For the more sophisticated market 10 and market 15, the corresponding differences between the deterministic and stochastic systems is observed for market 10 but not for market 15. The explanation lies in the fact that the calibration of market 15 is rather good in the stiff direction (α) already for an early generation, a consequence of a lucky initialization of chromosomes. The deterministic market 15 in Figure B.3 is also rather well estimated, but does not have the same lack of diversity among chromosomes.

Turning to market 10 in Figures B.1 and B.2, the performance of the algorithm is poorer. It is somewhat puzzling that the calibration of the stiff A-direction becomes worse when the number of generations increases, but once again this could be a result of the fitness landscape. Moreover, for both the deterministic and the stochastic system, the calibration seems to push the values of A and α to opposite signs. Analysing the phase plot of the corresponding system with $A = 8$ and $\alpha = -4$ gives the same limiting cycle of the system but not the same attractors. Thus, if the time-series gets stuck in the limiting cycle it is difficult to distinguish between the two systems. In fact, this is exactly what happened with the synthetically generated mispricing time-series, and what explains the poor calibration results. More interesting is that using $(x_0, y_0) = (1.5, 0.2)$ as starting point for the deterministic system yields convergence to the positive attractor, not to the limiting cycle. Thus, this exemplifies the impact of the stochaticity of the model. In addition, the Sloppy model analysis for market 10 shows that the alignment ratio in both the α- and A-direction are very low, which implies high uncertainty.

As mentioned earlier, the introduction of nearly perfect chromosomes changes the population quickly towards the true parameters. Thus, the performance of the Evolutionary algorithm could presumably be enhanced by improving the mutation operator. Trials of tuning the operator and straight forward applications of random noise in mutation resulted, however, in no significant improvements. This problem is left for further works.

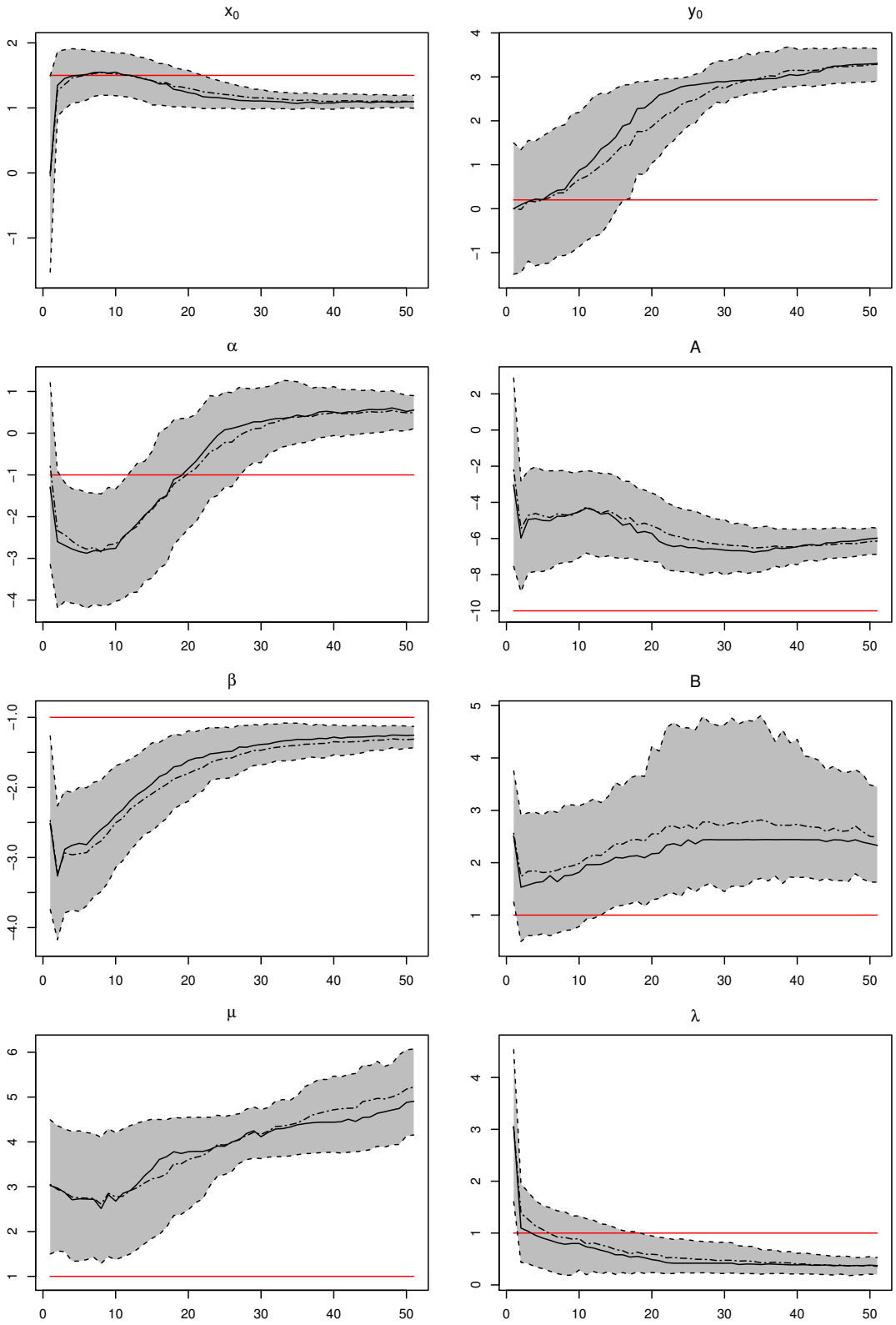

Figure 4.5: Market 1. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

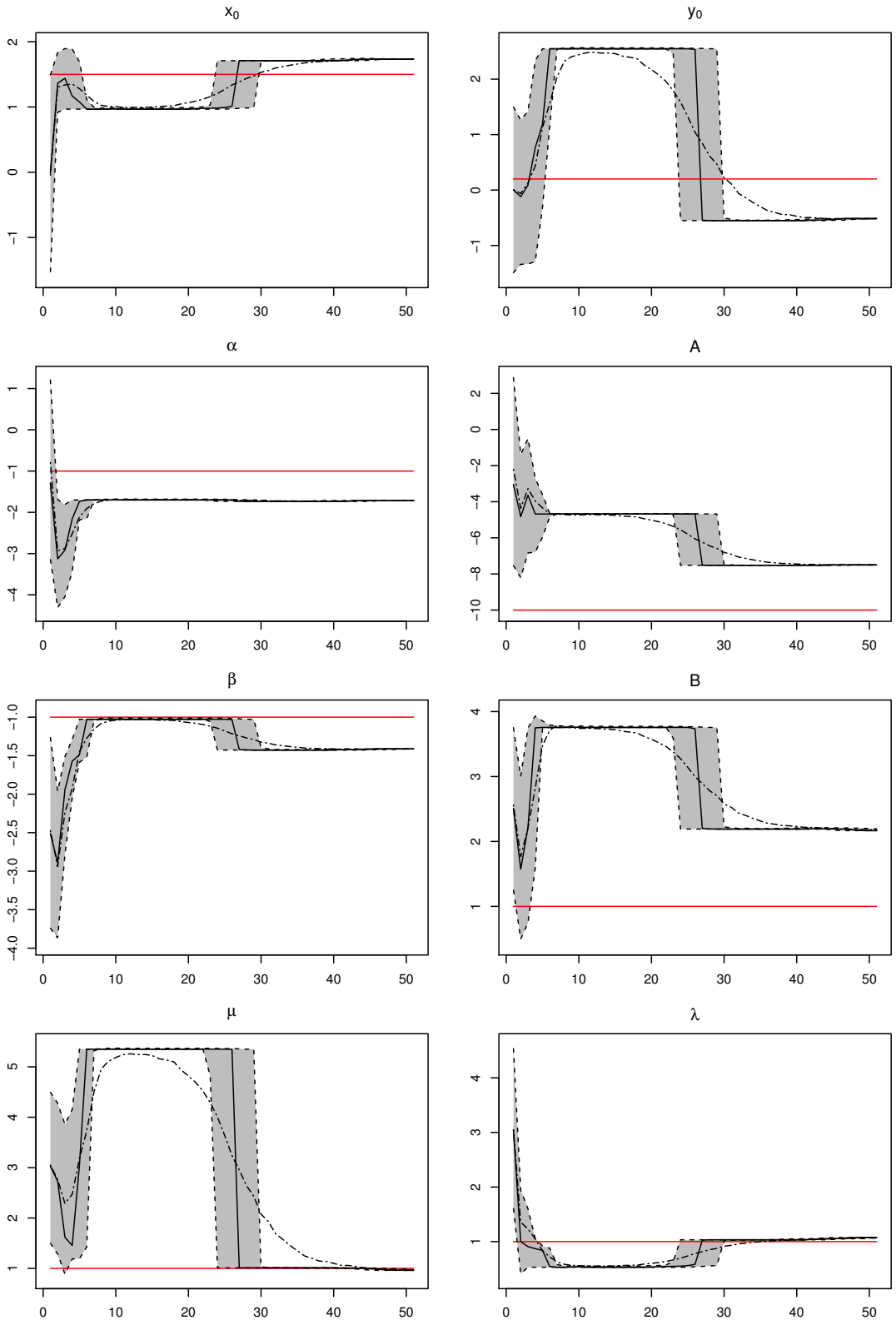

Figure 4.6: Market 1, deterministic. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

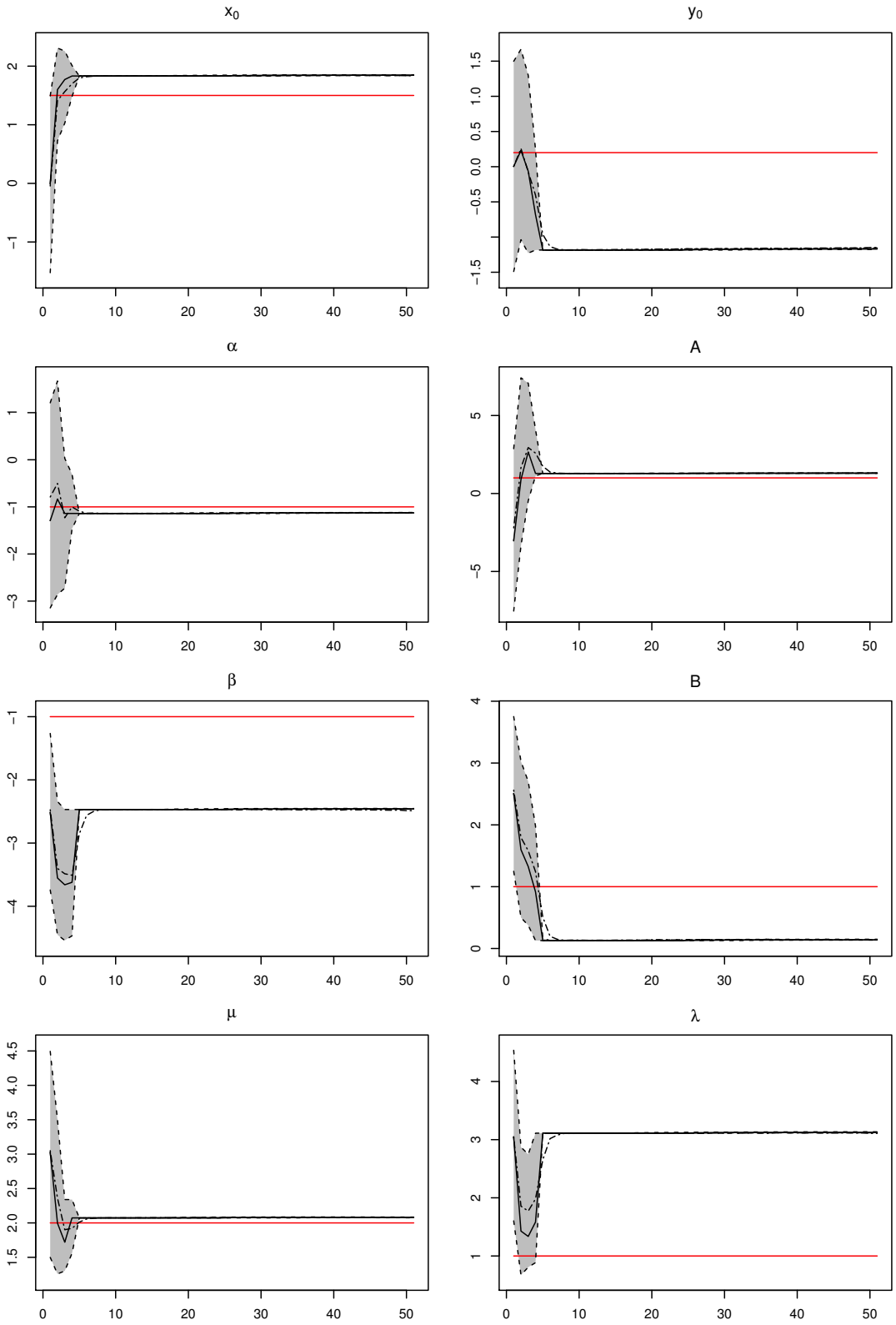

Figure 4.7: Market 15. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

4.3.2 Extensive Calibration Study

The previous section examines the performance of the Evolutionary algorithm using one initial population and by allowing evolution to take its course in the fitness landscape. Using for instance the high-performance cluster Brutus of ETH Zurich, it is possible to run this calibration algorithm in parallel to obtain a larger off-spring set and to further investigate the parameter estimation performance. Focusing only on the stochastic systems, and running the algorithm in parallel 25 times on the same time-series yields the results presented in Figure 4.8 and with descriptive statistics summarized in Table B.1. The choice of analysing the same time-series is partly due to that this is the case in reality and partly because the limiting performance of the algorithm will be examined.

Once again for market 1, the joint performance of the calibration yields satisfying results. The previously misspecified α-direction is better estimated, but not specified with enough certainty to, e.g., determine its sign. However, in 36 % of the cases all parameters are correctly specified, which is fairly good. More puzzling is still the inability to calibrate the stiff μ-direction with high alignment ratio more precisely. The explanation must once again originate from effects of the fitness landscape that are suggested in Figure 4.2. In addition, the intervals of the box plot include the true parameter values, but it is somewhat non-satisfactory that for some parameters the intervals are rather wide. Note that for μ and λ are the absolute value of the estimates used, since the calibration of these parameters is independent of signs. In conclusion, it seems possible to obtain the correct parameters for the simple market 1. The use of this calibration is further examined in Section 4.3.3.

Again, the synthetic time-series x generated for market 10 attracts to the limiting cycle. Thus, the estimates for the parameters α and A given by the EA are more or less indistinguishable form the true values of $\alpha = 5$ and $A = -1$. With this in mind, the calibration of market 10 performs seemingly well. The only issue concerns the interpretation of the parameters, i.e., the calibration changes the interpretation from a market consisting of speculative individuals and collectively correcting agents to mean-reversion individuals with a collectively speculative behaviour. Thus, the reliability of the real world interpretation from the calibration could be low, but hopefully it could still be used for forecasting.

The extended calibration of market 15 is not as good as the calibration in Figure 4.7, simply because the initialization cannot be as fortunate when running the parameter estimation procedure 25 times. Nonetheless, the median values in Table B.1 are not particularly far from the true values. However, the uncertainty bounds are rather large. This suggests that the calibration performs well, but it needs more chromosomes to generate thinner bounds.

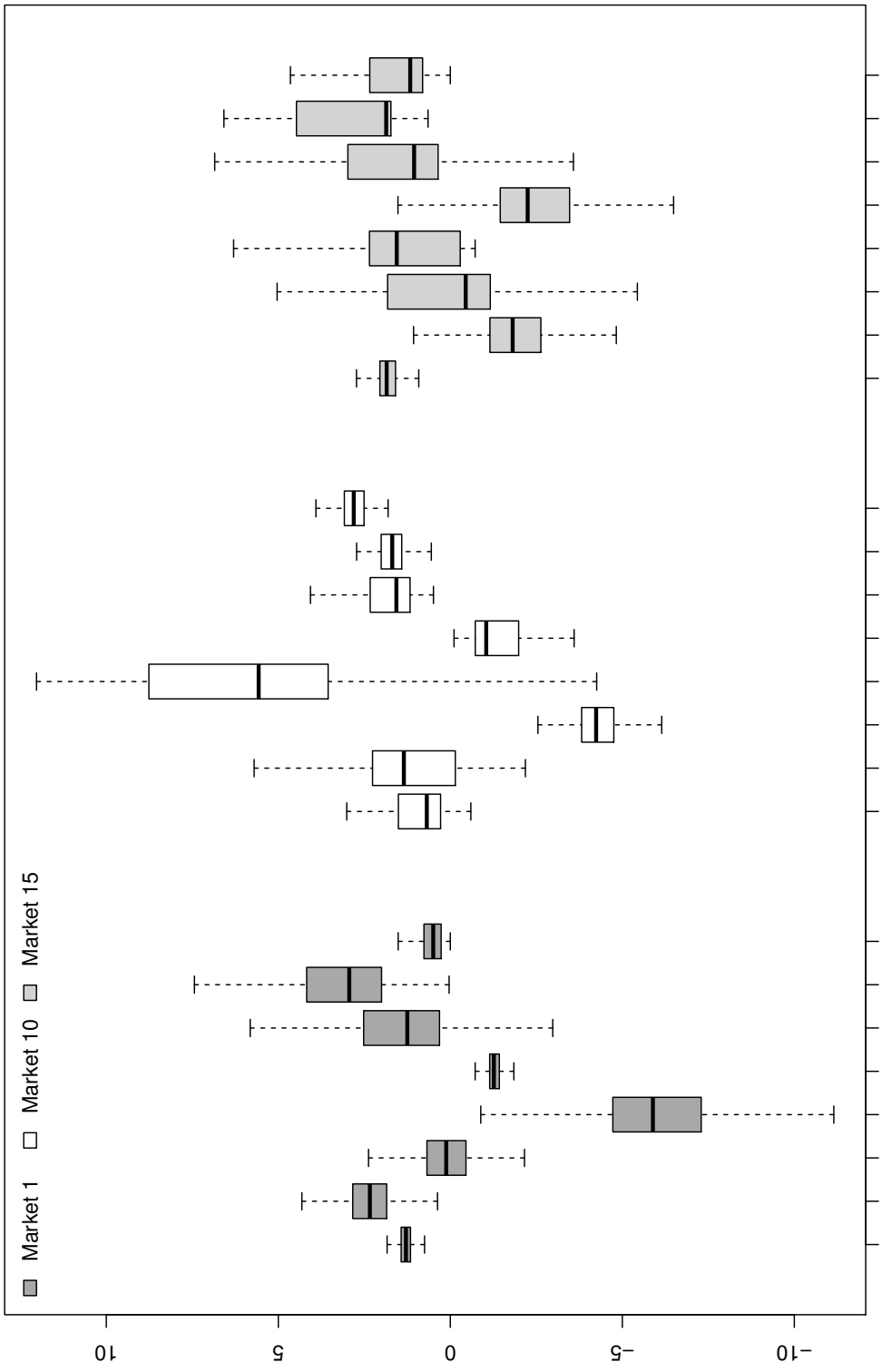

Figure 4.8: Box plot of calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using 25 parallel estimations of the 10% fittest chromosomes out of 20000 after 50 generations.

In summary, the extended Evolutionary algorithm performs rather well in calibrating the model. Some of the wrongly calibrated parameters actually allow for very alike dynamics for the mispricing. Nevertheless, the uncertainty bounds of the parameters are rather large, and decreasing these is a question of computational power. The cluster used for these simulations uses 12 cores in parallel and the computation takes about 12 hours per market, hence it is manageable to use more computer power if it is available.

4.3.3 Investigation of Prediction Power

The Sloppy model theory suggests that a Sloppy model could still be useful for prediction purposes. In this section, the prediction power of the stochastic models for market 1, market 10 and market 15, calibrated by the extended Evolutionary algorithm in Section 4.3.2, are analysed. The median values of the parameters given in Table B.1 are used as the parameter estimates, and in accordance with the observability of x and Section 4.2.2, the true values for the mispricing components x and y after 7 years are utilized. With the parameters and the initial conditions set, the distribution of the mispricing component x is simulated for time-periods of 1 week and 1 month using 10 000 different trajectories.

The calibration in the previous section also contains a generation of a corresponding out-of-sample mispricing time-series. Thus, using this time-series and the simulated distribution of mispricing, it is possible to calculate the probability of finding more extreme mispricings than the "true mispricing" x_{obs} at the end of the considered time-period. One possible measure is to consider the tail of the distribution larger than x_{obs}. If this probability is very large or very small, then it means that x_{obs} is located in one of the tails of the distribution and the forecast is poor. Ideally, the value 0.50 is obtained, i.e., half of the distribution is above and half is below x_{obs}.

Using the corresponding one-dimensional Gaussian kernel density estimator to the one presented in Section 3.3.3, the density of the mispricing x as well as the cumulative distribution of x at a future point in time can be estimated. Applying the kernel estimator yields the results presented in Table 4.4. Generally, the standard deviation of the estimated distribution is small for both time-spans, which is good if the model is to be used in real life forecasting. However, to be reliable, the prediction also needs to be consistent with real outcomes.

On a 1 week time scale, all the markets are predicted very well, the largest discrepancy is one standard deviation only. However, on the longer time scale, market 1 and market 10 are far (greater than two standard deviations) from the predicted value. It should also be mentioned that the calibration for market 10 is dependent on the mispricing following the limiting cycle. Thus, if the mispricing leaves this domain, then the forecasts would probably be way off. A speculative interpretation is that it might be more difficult to predict the mispricing close to equilibrium, since e.g., the mispricing component y is more difficult to estimate.

In summary, this initial investigation of the prediction performance indicates that the model calibrated by the Evolutionary algorithm performs rather well on a weekly time horizon. A more thorough prediction performance analysis would probably provide more insight, and could be the topic of further studies.

Table 4.4: Descriptive statistics of estimated mispricing distribution $\hat{x}$ (mean and standard deviation) and the true outcome x_{obs} after 1 week and 1 month respectively. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	1 week				1 month			
	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$
Market 1	0.29	0.28	0.01	0.15	0.31	0.25	0.02	$8.6 \cdot 10^{-4}$
Market 10	-4.09	-4.09	0.01	0.60	-4.30	-4.38	0.02	$2.4 \cdot 10^{-4}$
Market 15	3.25	3.26	0.01	0.75	3.29	3.29	0.02	0.67

4.3.4 Re-Parametrized Calibration

As mentioned in Section 4.2.1, the chosen parametrization used in the calibration of the asset pricing model might reflect the parameter estimation results. In particular, sloppy directions might be transformed into stiff directions and vice versa. Furthermore, convergence properties of an estimator can depend on the chosen parametrization, e.g., unbiasedness. One natural suggestion would be to use $1/\mu^2$ and $1/\lambda^2$ as parameters instead of μ and λ. Thus, the parameter estimation approach corresponding to the one in Section 4.3.2 is chosen and the results for the different parametrizations are compared. Since the issue of re-parametrization has been brought up relatively late, only the re-parametrization for the Evolutionary algorithm is considered without the corresponding prediction power analysis. Therefore, the results presented here are indicative and a more thorough analysis should be made in future works. In addition, further analyses should be accompanied by studies of the model manifold, cf., end of Section 4.2.1.

In Figure 4.9 and Table B.2, the values of μ and λ are transformed to the original parametrization for comparison. The performance of the re-parametrized calibration is similar to the previous results for market 1 and market 10. In particular, the new parametrization allows for the same calibration for market 1, and the small deviations are solely due to different realizations of the simulations. Thus, the differences for market 1 originates from the mutation operator, and since the differences are considerable, this might be optimized further. Nevertheless, the uncertainty for calibration of market 15 is substantially decreased, as a consequence of the new parametrization. This suggests that the sloppiness is dependent on the parametrization and that this dependence structure hinges on market characteristics. Hence, there does not seem to be an easy answer to the parametrization dependence and further investigations would be desirable.

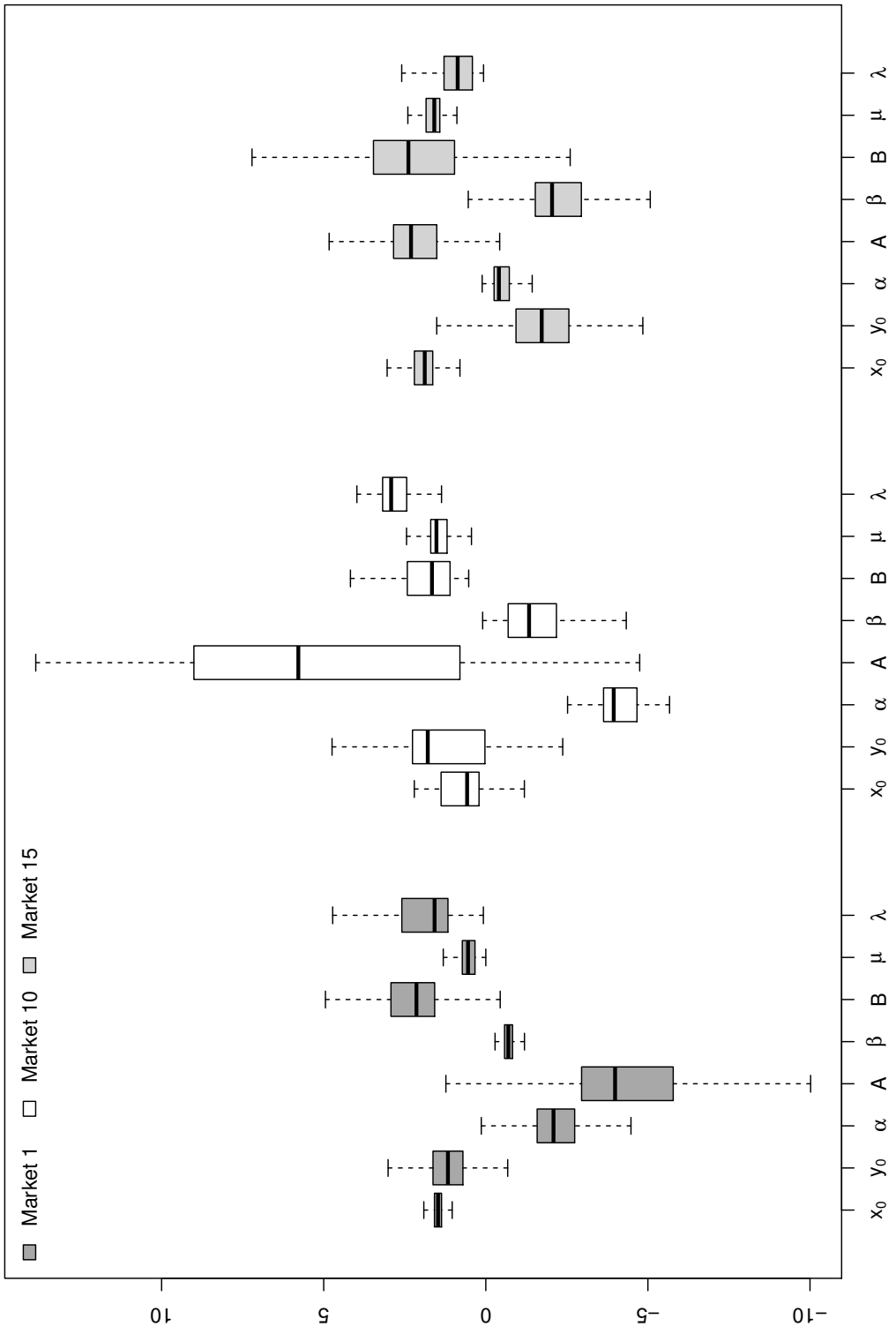

Figure 4.9: Box plot of calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ for the re-parametrized model using 25 parallel estimations of the 10% fittest chromosomes out of 20000 after 50 generations.

However, the insights that the calibration of market 10 seems to not be dependent on the parametrization and that the median values for the estimates of market 15 are similar are more convincing. Thus, the analyses made in the previous sections could be assigned more confidence, simply because the results in this section provides further support to the earlier findings. This also indicates that the sloppiness structure is intrinsic, at least partly, since the results are similar for different parametrizations.

4.4 Simulated Annealing Calibration Results

The underlying idea of introducing the Simulated annealing parameter estimation method originates from the properties of the "sloppiness structure" at different positions in the parameter space. The Simulated annealing approach with iteration over temperatures should alleviate the random walkers to explore the parameter space more broadly at high temperatures. Thus, the work to climb the potential is low initially and one could hope that the ants do not wind up in sub-optimal valleys, which is one of the greatest problems with the parameter estimation. In the example of Figure 4.2, this suggests that an ant starting at, e.g., $(\mu, A) = (5, -7)$, at high temperatures should easily walk along the potential towards $\mu = 1$, which would make it easier to calibrate the stiff μ-direction. In this section, the results from the Simulated annealing calibration approach is presented together with a prediction study.

4.4.1 Parameter Estimation

As a consolidation of the intuitive idea of the Simulated annealing method, the diminished market 1 with only μ and A unknown is analysed graphically. Figure 4.10 illustrates the initial positions of ants as well as the positions at steady-state for different temperatures. The steady-state is tested to occur after about 1000 walks. Relevant temperatures are, also after testing, chosen to be $T \in \{0.001, 0.01, 0.1, 1\}$ for the application of the algorithm on the full parameter space. As expected, the ants are spread for high temperatures, while for lower temperatures they cluster around positions in the parameter space where the potential is low, cf., Figure 4.2. Since no restrictions on the parameters are used, some of the ants walk out of the domain used for the initialization. Thus, the interpretation of the ants' positions has to be treated with caution.

The presentation of the results from the Simulated annealing method applied to the full model with a multidimensional parameter space is not as easy as the two-dimensional case in Figure 4.10. Instead, the results are presented quantitatively in Tables 4.5 and 4.6 for the adapted kernel estimation and the kNN method, respectively. A highly important issue regarding the Bayesian maximum likelihood estimation under the posterior probability distribution to calibrate the asset pricing model is to consider if there exists multiple local maxima in different parts of the parameter space. Consequently, descriptive statistics of the 10 points with highest probability for each market can be found in Tables B.3 and B.4 for the two different methods, respectively.

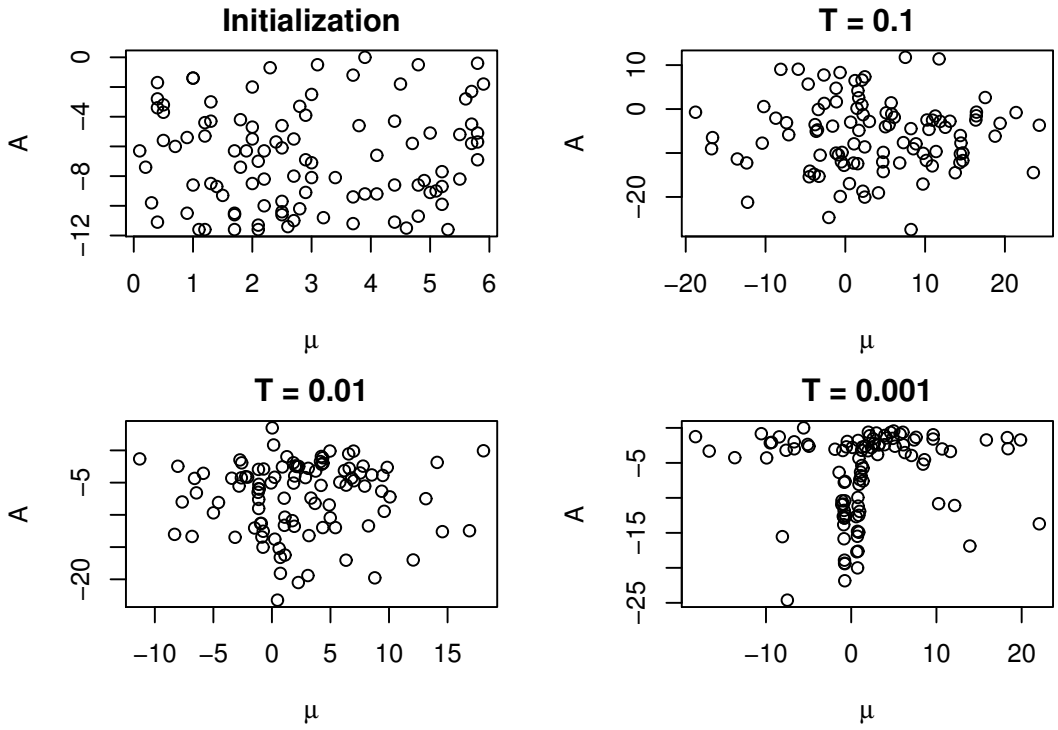

Figure 4.10: Initial position of ants in the Simulated annealing calibration method and steady-state configurations for temperatures $T \in \{0.1, 0.01, 0.001\}$. Note the symmetry in the μ-direction for low temperatures.

The performance of the two methods is rather similar and it should be pointed out that the density is estimated on the final positions of the ants. The k in the kNN method has been set by tuning to $k = 5$ for market 1 and market 10, where the differences in the results for different k's are negligible, and to $k = 10$ for market 10 where the differences are greater. In turn, for the kernel method, the choice of the bandwidth reflects highly the value of the estimated probability density. A too small bandwidth will make the estimates too large and vice versa. Thus, for the nominal value of the probability density, the unbiased estimate from the kNN method should be more reliable.

However, the nominal value of the density is not important for the posterior maximum likelihood estimation per se, but is relevant for the question regarding the size of the bandwidth itself. In this case, all the points where the distribution is estimated contain a perfect observation (since the ants' positions are used). Thus, all the points have a contribution of equal size, and comparing the most likely ant position with the most unlikely one will provide an insight on the coupling of the position of the ants with the probability density. The estimation yields that the smallest probability is about 10^{-13}, i.e., the difference with the most likely position is distinct and the estimation can be

considered reliable.

Table 4.5: Bayesian parameter estimation results obtained by maximizing the posterior probability distribution $\hat{p}(\theta|x)$ given by kernel density estimation of ants for market 1, market 10 and market 15. The values in the parenthesis are the true parameter values.

	y_0	α	A	β	B	μ	λ	$\hat{p}(\theta\|x)$
Market 1	1.0 (0.2)	-3.6 (-1)	-1.0 (-10)	-0.6 (-1)	3.0 (1)	1.7 (1)	-0.4 (1)	$2.6 \cdot 10^{-9}$
Market 10	3.5 (0.2)	-4.4 (5)	1.9 (-1)	-1.7 (-1)	3.1 (1)	1.0 (5)	2.5 (3)	$6.4 \cdot 10^{-10}$
Market 15	-1.1 (0.2)	-2.5 (-1)	4.0 (1)	-7.3 (-1)	2.2 (1)	1.9 (2)	0.7 (1)	$1.2 \cdot 10^{-9}$

Table 4.6: Bayesian parameter estimation results obtained by maximizing the posterior probability distribution $\hat{p}(\theta|x)$ given by the kNN method of ants for market 1, market 10 and market 15. The values in the parenthesis are the true parameter values.

	y_0	α	A	β	B	μ	λ
Market 1	1.7 (0.2)	-4.3 (-1)	-0.4 (-10)	-0.7 (-1)	1.7 (1)	1.0 (1)	-0.4 (1)
Market 10	-0.7 (0.2)	-4.9 (5)	7.4 (-1)	-0.9 (-1)	1.2 (1)	1.9 (5)	3.0 (3)
Market 15	-2.0 (0.2)	-0.5 (-1)	3.9 (1)	-3.7 (-1)	5.4 (1)	1.6 (2)	-0.4 (1)

	$\hat{p}(\theta\|x)$	k	$\widehat{\mathrm{Var}}(\hat{p}(\theta\|x))$
cont. Market 1	$5.1 \cdot 10^{-5}$	5	$8.6 \cdot 10^{-10}$
cont. Market 10	$1.4 \cdot 10^{-4}$	5	$6.2 \cdot 10^{-9}$
cont. Market 10	$2.7 \cdot 10^{-6}$	10	$8.9 \cdot 10^{-13}$

Turning to the performance of the two methods on market 1, the stiff direction (μ) is well estimated, and the 10 most probable ants are roughly in the same region. As seen before, y_0 is poorly estimated, and the same applies to the sloppy A-direction. It should again be noted that the sign of λ is irrelevant because of the symmetric (quadratic) dependence in the function f^{NL}. The symmetry is visible for instance in Figure 4.10. For the signs of all parameters, the kernel method performed perfectly while the kNN method succeeded in 70 % of the cases. In summary, the method performs well on market 1.

For market 10, again the issue of attraction of the mispricing time-series x to the limiting cycle occurs. Once again, this results in opposite estimation of the α and A parameters. However, the calibration performance is seemingly good for this "new market".

The calibration of market 15 is more troublesome. For the kNN method, the result is highly dependent on the number of neighbours, and for small ks are the estimated values not in the vicinity of the true parameters. The probability density is maximized close

to another point in parameter space. However, increasing k transforms the optimum towards a more realistic point in the parameter space. This indicates that the probability density has different local maxima, which makes the parameter estimation more difficult. Thus, when applying the method on real data, a spectrum of different ks should be used. For the kernel estimation this problem does not occur, i.e., the bandwidth seems to be well-suited for the calibration. Nevertheless, the estimates of the parameters for both methods on market 15 are fairly good, even though the uncertainty for the kNN method is rather high (high standard deviation of estimates). The stiff μ-direction is almost perfectly estimated, while both the A and the β parameters are not as good. However, these direction are not stiff, and hopefully the predictions could still be reliable.

4.4.2 Analysis of Prediction Power

The performance of the predictions upon using the estimates reported in Tables 4.5 and 4.6 is exhibited in the same fashion as in Section 4.3.3, which contains details regarding the construction of the prediction distributions. As can be seen in Tables 4.7 and 4.8, the prediction on the 1 week horizon is almost spot-on and the standard deviations are small. For 1 month, the results are also very good and exhibit small standard deviations.

A remark for the very promising results is the assumption about knowledge of y_0. A sensitivity analysis has been conducted, and its results are summarized in Table B.5. Nevertheless, for the shorter time-period, the prediction performance is generally still great, but the 1 month prediction performance is slightly worse. However, for a longer time-frame it should be possible to estimate y more precisely, and thus the input parameter should be more reliable. An additional remark concerns the usability of the predictions. These tests show that the calibration and out-of-sample predictions work well when applied to synthetically generated data. To test the performance on real asset prices, a real financial time-series is needed and this is studied in Chapter 5.

In summary, both the Evolutionary algorithm and Simulated annealing seem to perform rather well in forecasting the synthetic time-series on a weekly time-horizon. The two Simulated annealing approaches perform somewhat better on a longer time-scale and might hence be preferable, even though the differences are not very large. Regarding parameter estimation, the Simulated annealing approach also performed slightly better, e.g., in terms of determining signs of parameters (y_0 exempted).

Table 4.7: Descriptive statistics of estimated mispricing distribution $\hat{x}$ (mean and standard deviation) and the true outcome x_{obs} after 1 week and 1 month, respectively, using the adapted kernel estimation. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	1 week				1 month			
	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$
Market 1	0.55	0.55	0.01	0.71	0.55	0.53	0.02	0.19
Market 10	-4.09	-4.09	0.01	0.76	-4.30	-4.32	0.02	0.25
Market 15	3.25	3.25	0.01	0.72	3.29	3.26	0.02	0.15

Table 4.8: Descriptive statistics of estimated mispricing distribution $\hat{x}$ (mean and standard deviation) and the true outcome x_{obs} after 1 week and 1 month, respectively, using the kNN method. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	1 week				1 month			
	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$
Market 1	0.55	0.55	0.01	0.70	0.55	0.53	0.02	0.16
Market 10	-4.09	-4.09	0.01	0.71	-4.30	-4.33	0.02	0.12
Market 15	3.25	3.25	0.01	0.70	3.29	3.27	0.02	0.23

Chapter 5

A Study of a Financial Time-Series

In this chapter, a real-world financial time-series is studied. Motivation of the chosen time-span and index is presented in Section 5.1. The two parameter estimation methods studied in this thesis are applied to the time-series in Sections 5.2 and 5.3, respectively. The analyses use an estimation window of 8 years and study a forecast period of up to 1 year, to measure the performance of the asset pricing model in a real world environment.

5.1 Motivation of Choice of Studied Time-Series

To be able to apply the Evolutionary algorithm and the Simulated annealing approach, the same assumptions as stated in Chapter 2 and summarized in Section 2.2.4 are needed. In particular, this specifies the fundamental price p_f and the hyperparameters μ_f, σ_x and σ_y as defined in Section 2.2.3. Moreover, since the market price is assumed to follow a type of Ornstein - Uhlenbeck process, the deterministic fundamental price needs to be extracted from the market price time-series. This can be done in various ways, but with ex-post knowledge about the financial climate in certain time periods this is easily done by a simple trend-fitting of the time-series of, what could be considered as, market prices in equilibrium.

In this thesis, the stock market index Standard & Poor's 500 (S&P 500) is studied. The index consists of the 500 largest companies in terms of market capitalization listed on NYSE or NASDAQ. Thus, the asset pricing model is applied to a collection of stocks, which implies that the model parameters to be estimated should be interpreted as the joint market behaviour. As previously seen, the calibration is highly dependent on the out-of-equilibrium properties of the mispricing. Therefore, the time period from 2000 to 2008 is chosen, since it begins with the crash of the .com-bubble (climax in March 10, 2000) and ends with the start of the Great recession, e.g., the filing of bankruptcy of Lehman Brothers on September 15, 2008. In contrast to earlier applications of the calibration methods, an estimation window of 8 years instead of 7 years is used. This uses more information from the time-series and should improve calibration. However, the most important reason is that this will end the estimation window at the last trading

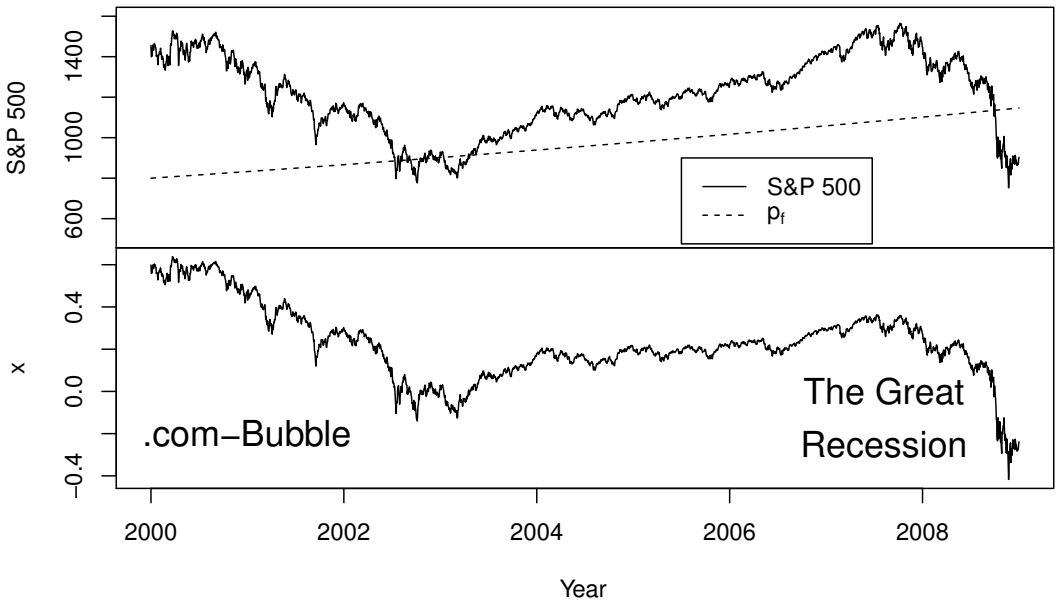

Figure 5.1: Time-series of the S&P 500-index from 2000 to 2008 and the extracted mispricing x. The graph emphases the two most important financial events during this time period.

day of 2007, i.e., before the Great recession. Thus, it will be possible to examine how well an event like the Great recession is projected by the asset pricing model.

In Figure 5.1, the time-series of the S&P 500-index is illustrated for the chosen time period. It includes the fundamental price p_f fitted to the S&P 500 time-series with $\mu_f = 0.04$ as well as the extracted mispricing component x. The fit highlights that the mispricing varies from a maximal mispricing of about 0.6 to a minimal mispricing of -0.4.

5.2 Evolutionary Algorithm Calibration

In this section, the application of the Evolutionary algorithm on the S&P 500 time-series is presented and analysed.

5.2.1 Parameter Estimation

Calibrating the model on S&P 500 data using the Evolutionary algorithm and $\mu_f = 0.06$ (cf. Section 5.2.2) gives the results shown in Figure 5.2 with descriptive statistics in Table 5.1. For the real data calibration, the uncertainty bounds are seemingly large,

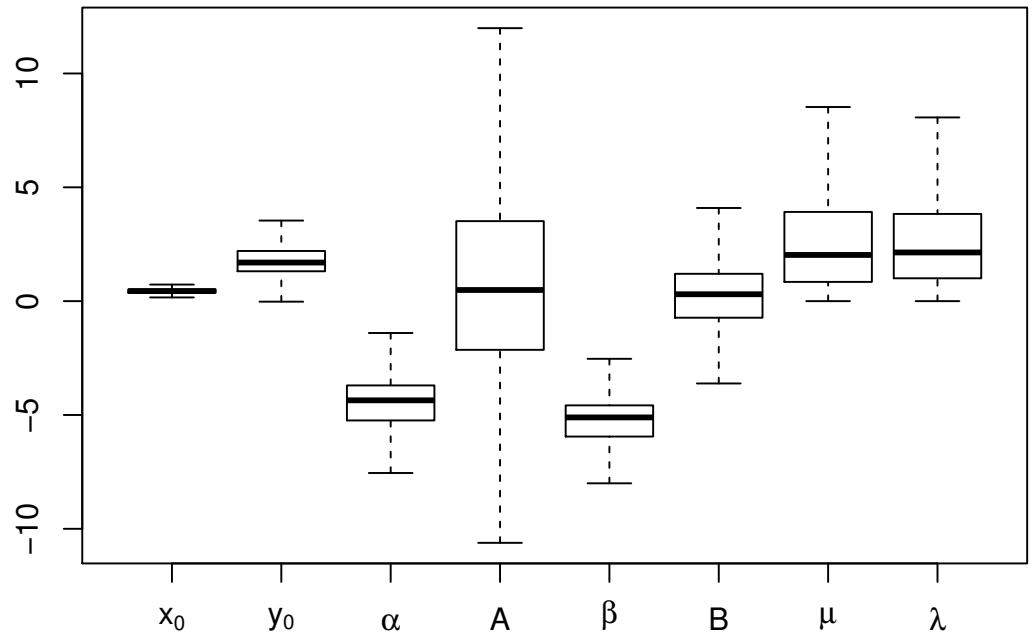

Figure 5.2: Box plot of the parameter estimation results for the S&P 500 time-series including initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations for 25 parallel estimations.

especially for the A parameter. However, as seen for the calibration of market 15 in Section 4.3.2, the parameter estimation could still be rather satisfactory in forecasting, but a higher number of chromosomes would be necessary to decrease the uncertainty. The interesting α parameter is estimated to be negative, i.e., that individuals follow a mean-reversal strategy. This seems reasonable in the aftermath of the .com-bubble. In addition, the median value of A is positive, indicating speculative joint behaviour of the market participants. However, the uncertainty in this estimate is very high.

Figure B.4 provides the phase plot of the corresponding deterministic system given by the median values of the most fit chromosomes (the values in Table 5.1). This deterministic system has only one stable attractor at the equilibrium point $(x, y) = (0, 0)$. Thus, this calibration seems to yield a model that describes well how the market retreats from the bubble-phase in beginning of the sample data, but that probably cannot capture the market development after January 1, 2008 when the market changes to a negative regime. This is investigated further in the next section.

Table 5.1: Descriptive statistics (median and standard deviation in parenthesis) of the 10 % fittest chromosomes of the extended Evolutionary algorithm calibration on the S&P 500 data with 25 parallel runs of the algorithm and 20000 chromosomes in each run.

x_0	y_0	α	A
0.44 (0.12)	1.70 (1.09)	-4.36 (1.69)	0.49 (5.60)
β	B	μ	λ
-5.11 (1.69)	0.30 (2.25)	2.03 (2.84)	2.14 (2.55)

5.2.2 Sensitivity Analysis of Fundamental Price

As a consequence of a deterministic fundamental price, the choice of the market growth-rate μ_f directly affects the behaviour and amplitude of the mispricing x. A smaller growth-rate yields a larger contribution of the mispricing component in market prices and vice versa. Therefore, the calibration results for the cases $\mu_f = 0.02$, $\mu_f = 0.04$, $\mu_f = 0.06$ and $\mu_f = 0.13$ are investigated in this section. Table 5.2 reports the median values and standard deviations of the parameter estimates for the different values of μ_f.

As can be seen in this table, the calibration results depend heavily on the choice of μ_f. For $\mu_f = 0.02$ and $\mu_f = 0.04$, the estimates do not seem to converge. In particular, the uncertainties in the estimates are huge and the estimates are unreasonable regarding both signs and sizes. Meanwhile, the results for $\mu_f = 0.06$ and $\mu_f = 0.13$ are similar. From a practical point of view, the explanation is a consequence of the realization of the mispricing x. Comparing Figure 5.1 ($\mu_f = 0.04$) with Figure B.6 ($\mu_f = 0.06$), the states of the mispricing x with $\mu_f = 0.06$ are more distinguishable. This means that the mispricing changes from a positive regime, to equilibrium and further on towards a negative mispricing. In the case $\mu_f = 0.04$, the mispricing does not stay at or close to equilibrium. Instead it grows again after the .com-Bubble, before it drops as a consequence of the Great recession.

With this in mind, the calibration results suggest that the asset pricing model is useful for determining market characteristics only when it is possible to clearly distinguish mispricing regimes. This goes in hand with the long-term objective of the model to be useful for calculations of regime switching probabilities, but demands an increased user awareness and pin-points the drawbacks of a deterministic fundamental price. This means that small fluctuation in market prices should be encountered by the fundamental price, not the mispricing, for calibration to be meaningful, which is not possible in the deterministic case. Furthermore, this does not mean that in general $\mu_f = 0.06$ is a better choice than $\mu_f = 0.04$, since the increases in the mispricing after the .com-Bubble could be a consequence of stochasticity of the fundamental price, cf., a trajectory of a random walk. Nevertheless, in this particular application of the Evolutionary algorithm, the calibration

Table 5.2: Descriptive statistics (median and standard deviation in parenthesis) for different values of μ_f of the 10 % fittest chromosomes of the extended Evolutionary algorithm calibration on the S&P 500 data with 25 parallel runs of the algorithm and 20000 chromosomes in each run.

	x_0	y_0	α	A
$\mu_f = 0.02$	0.59 (0.28)	2.46 (6.43)	5.47 (9.61)	-48.12 (96.70)
$\mu_f = 0.04$	0.51 (0.22)	2.08 (3.20)	-0.03 (1.65)	-17.56 (20.99)
$\mu_f = 0.06$	0.44 (0.12)	1.70 (1.09)	-4.36 (1.69)	0.49 (5.60)
$\mu_f = 0.13$	0.67 (0.11)	1.55 (1.04)	-4.73 (1.44)	3.63 (4.61)
	β	B	μ	λ
cont. $\mu_f = 0.02$	-10.03 (15.46)	-6.18 (11.58)	9.80 (17.07)	6.37 (13.77)
cont. $\mu_f = 0.04$	-7.09 (6.17)	-1.97 (6.56)	5.39 (7.31)	3.17 (7.36)
cont. $\mu_f = 0.06$	-5.11 (1.69)	0.30 (2.25)	2.03 (2.84)	2.14 (2.55)
cont. $\mu_f = 0.13$	-4.72 (1.05)	0.17 (1.13)	1.32 (1.94)	2.14 (2.55)

results are much more reliable for $\mu_f = 0.06$ and is thus used for the prediction study in the following section. In fact, it is not even a too unrealistic estimation of the long-term growth rate of the S&P 500 index. Using $\mu_f = 0.04$ and changing the parameters with one standard deviation leads to no consistency in the prediction distributions.

In conclusion, this sensitivity analysis suggests that choosing a deterministic fundamental price might be a too big simplification of the real-world dynamics when applying the Evolutionary algorithm for parameter estimation. For the calibration to be applicable, it is necessary to be able to distinguish in which regime the mispricing is.

5.2.3 Prediction Power Analysis

To be able to analyse the prediction power of the calibrated model, the same procedure as in Chapter 4 is used. However, for the real financial time-series it is not possible to use an exact value for y at the start of the prediction period, and the estimation method presented in Section 4.2.2 is applied. Performing this smoothing, the y value at January 1, 2008 seems to be in the interval $[-0.5, 0.5]$. Therefore, both the cases -0.5 and 0.5 are analysed, and the results are shown in Table 5.3.

As can be seen in the Table 5.3, the estimated mispricing distributions are narrow (the standard deviation is small). However, the model is unable to predict the true values, i.e., the x_{obs} are located in the tails of the prediction distributions. This is non-satisfactory, since it suggests poor performance when using the model to forecast asset prices. In addition, the 1 year result indicates that the Great recession was a one-in-ten million year event. As of now, with the understanding of what happened in 2008 and the risk

Table 5.3: Descriptive statistics of the estimated mispricing distribution $\hat{x}$ (mean and standard deviation) and the true outcome x_{obs} after 1 week, 2 weeks, 1 month and 1 year, respectively, when using the parameter estimates found in Table 5.1. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	$y = -0.5$				$y = 0.5$			
	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$
1 week	0.15	0.10	0.01	$3.0 \cdot 10^{-6}$	0.15	0.12	0.01	0.01
2 weeks	0.11	0.10	0.01	0.12	0.11	0.13	0.01	0.91
1 month	0.06	0.08	0.02	0.86	0.06	0.15	0.02	$1 - 2.5 \cdot 10^{-7}$
1 year	-0.44	-0.00	0.06	$1 - 1.3 \cdot 10^{-7}$	-0.44	0.11	0.06	$1 - 3.2 \cdot 10^{-7}$

management methods used, the likelihood of this event was probably to be assigned a higher probability. To get a deeper understanding of the prediction performance, these results should be compared with the Simulated annealing results.

5.3 Simulated Annealing Parameter Estimation

In this section, the corresponding analysis to the one in Section 5.2 is performed. Furthermore, a comparison between the prediction performance of the two different parameter estimation approaches is conducted.

5.3.1 Calibration of Parameters

In comparison with the results from the Evolutionary algorithm, the Simulated annealing method is not as sensitive to changes in μ_f, especially not the adapted kernel estimator. The results are summarized in Tables 5.4 and 5.5, and the ants seem to reach a steady-state for all the different parametrizations of μ_f. However, the uncertainty is high for the kNN method with $\mu_f = 0.02$. The estimated sign of the A parameter differs between the two calibration methods, but the uncertainty for the Evolutionary algorithm is high and might therefore be an explanation. The corresponding phase plots of the deterministic systems are, however, similar. There is only one attractor at equilibrium $(x, y) = (0, 0)$ and the out-of-equilibrium mispricing converges towards this value. Nevertheless, a comparison of Figures B.4 and B.5 yields that the convergence follows different paths, implying distinguished properties of the mispricing behaviour. With the results for the time-series in mind, the calibrated models describe the behaviour during the in-sample data well, and an extensive discussion of the performance in the post estimation window is presented in the following section.

The deviations of the calibrated parameter values between the two different methods for probability density estimation are small. However, the uncertainty is generally higher

for the kNN method. For instance the A parameter (the parameter estimate that differs most from the Evolutionary algorithm results) has very large uncertainty bounds. This indicates that this parameter is particularly difficult to estimate for the considered time-series, and a qualitative interpretation should be made with caution. Furthermore, the results from the Simulated annealing method ought to be more reliable than the Evolutionary algorithm results. The positions of the ants seem to have converged well into the steady-state distribution, while this is clearly not the case for the chromosomes in the Evolutionary algorithm for small values on μ_f.

Table 5.4: Position of the most fit ant and the standard deviation in parenthesis of the 10 fittest ants for different μ_f using the adapted kernel density for 12 000 ants.

	y_0	α	A	β
$\mu_f = 0.02$	0.06 (0.53)	-1.43 (0.65)	-4.11 (1.80)	-7.90 (0.78)
$\mu_f = 0.04$	1.07 (0.53)	-3.20 (0.80)	-3.63 (1.78)	-7.38 (0.77)
$\mu_f = 0.06$	1.18 (0.59)	-4.83 (0.55)	-2.43 (1.47)	-7.58 (0.62)
$\mu_f = 0.13$	0.78 (0.41)	-5.35 (0.75)	-0.55 (1.38)	-7.84 (0.58)

	B	μ	λ
cont. $\mu_f = 0.02$	0.19 (1.65)	4.01 (1.33)	2.53 (0.42)
cont. $\mu_f = 0.04$	1.38 (1.00)	4.02 (0.59)	3.32 (1.09)
cont. $\mu_f = 0.06$	1.19 (1.51)	4.57 (0.91)	2.83 (0.93)
cont. $\mu_f = 0.13$	-0.37 (1.59)	1.65 (1.73)	2.15 (0.80)

Table 5.5: Position of the most fit ant and the standard deviation in parenthesis of the 10 fittest ants for different μ_f using the kNN method on 12 000 ants.

	y_0	α	A	β
$\mu_f = 0.02$	-0.43 (0.67)	-1.66 (1.00)	5.25 (8.02)	-8.89 (2.37)
$\mu_f = 0.04$	0.13 (1.01)	-1.70 (0.81)	-7.37 (4.63)	-7.32 (2.37)
$\mu_f = 0.06$	0.91 (0.36)	-3.96 (1.48)	-2.55 (3.89)	-5.72 (1.67)
$\mu_f = 0.13$	1.20 (0.56)	-5.66 (0.93)	-0.48 (4.95)	-7.79 (1.62)

	B	μ	λ	k
cont. $\mu_f = 0.02$	3.55 (1.47)	1.31 (3.13)	1.93 (0.52)	5
cont. $\mu_f = 0.04$	0.21 (3.02)	3.93 (2.19)	1.97 (1.94)	5
cont. $\mu_f = 0.06$	3.16 (3.01)	2.67 (3.25)	0.83 (1.13)	5
cont. $\mu_f = 0.13$	-1.71 (3.19)	0.81 (2.48)	1.24 (1.79)	5

5.3.2 Investigation of Prediction Power

In this section, a corresponding analysis of Section 5.2.3 is conducted for the parameter estimates of the Simulated annealing method. Once again, two different starting points for y are chosen (-0.5 and 0.5) and the results are presented in Tables 5.6 and 5.7, respectively. The performance of the predictions are not satisfactory, since the true outcomes are located in the tails of the narrow prediction distributions. This could imply that the model might not be able to explain the American stock market well, or that the chosen calibration window is unrepresentative. Note that the values of x_{obs} differ compared to the Evolutionary algorithm predictions, since different values of μ_f are used. Nevertheless, the inability to predict the actual realizations of the mispricing of the S&P 500 index are alike for all three methods of generating the prediction distribution. This emphasises that the calibrated model has difficulties in describing the true market performance. In particular, the results are present for different values of μ_f, i.e., different interpretations of the mispricing x, and for different ys, which suggests that the poor predictions are a consequence of the model's properties.

As previously stated, the phase plots of the deterministic systems precisely describe the convergence of out-of-equilibrium prices to the mispricing $(x, y) = (0, 0)$ and reflect the time-period January, 2000 to December 2007 well. Thus, when trying to describe actual market prices with the model of Yukalov et al. with constant parameters, other market phases should presumably also be used to calibrate the model. This means that the parameter estimates at hand only describe the convergence of the mispricing from a positive attractor, i.e., the burst of a bubble. One suggestion could be to perform similar calibrations on the building-up phase of bubbles and also for the beginnings and recoveries of recessions. From this, it might be possible to construct dynamic parameters depending on the present mispricing x to be able to predict the actual market prices more precisely. More effort in this direction could improve the knowledge about these issues. The usage of a relatively short time-period goes in hand with the earlier studies on synthetic data, and the main constraint for longer time-series is the computational power needed for repeated simulation. Hence, using longer estimation windows could be profitable, since the data would include different regime switches and the calibrated model could capture these transitions.

In conclusion, these introductory results on performance of calibration and prediction of the asset pricing model of Yukalov et al. in a real-world context are not as satisfactory as desired, neither for the Evolutionary algorithm nor the Simulated annealing approach. It is even hard to motivate if one of the approaches are preferable, but the Simulated annealing approach with an adapted kernel estimator is least sensitive to changes in μ_f and should hence be more reliable. One idea that might improve the performance is to study enlarged time-series to capture other regime switches than the bursting of the .com-Bubble.

Table 5.6: Descriptive statistics (mean and standard deviation) of the estimated mispricing distribution $\hat{x}$ from the Simulated annealing method with kernel density approximation and the true outcome x_{obs} after 1 week, 2 weeks, 1 month and 1 year, respectively, using the estimates in Table 5.4 for $\mu_f = 0.04$. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	$y = -0.5$				$y = 0.5$			
	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$
1 week	0.31	0.26	0.01	$7.2 \cdot 10^{-7}$	0.31	0.28	0.01	0.01
2 weeks	0.27	0.26	0.01	0.10	0.27	0.29	0.01	0.89
1 month	0.22	0.24	0.02	0.86	0.22	0.30	0.02	$1 - 6.5 \cdot 10^{-7}$
1 year	-0.26	0.13	0.06	$1 - 1.4 \cdot 10^{-7}$	-0.26	0.22	0.06	$1 - 1.5 \cdot 10^{-7}$

Table 5.7: Descriptive statistics (mean and standard deviation) of the estimated mispricing distribution $\hat{x}$ from the Simulated annealing method with the kNN method and the true outcome x_{obs} after 1 week, 2 weeks, 1 month and 1 year, respectively, using the estimates in Table 5.5 for $\mu_f = 0.04$. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events.

	$y = -0.5$				$y = 0.5$			
	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$	x_{obs}	$\mathrm{m}(\hat{x})$	$\mathrm{sd}(\hat{x})$	$P(\hat{x} > x_{obs})$
1 week	0.31	0.26	0.01	$1.5 \cdot 10^{-7}$	0.31	0.28	0.01	0.01
2 weeks	0.27	0.26	0.01	0.11	0.27	0.29	0.01	0.88
1 month	0.22	0.24	0.02	0.86	0.22	0.30	0.02	$1 - 1.4 \cdot 10^{-7}$
1 year	-0.26	0.16	0.07	$1 - 1.6 \cdot 10^{-7}$	-0.26	0.25	0.06	$1 - 1.7 \cdot 10^{-7}$

Chapter 6

Conclusion

In terms of the objective of this Master thesis, to get a deeper understanding of why the calibration of the asset pricing model by Yukalov et al. is difficult, both the Sloppy model analysis and the performance of the two different parameter estimation approaches are highly informative. Firstly, the calibration of the model is highly dependent on the access to information about out-of-equilibrium dynamics. In the limit of deterministic prices in equilibrium, there is no information at all that is useful for calibration. Secondly, the Sloppy model analysis shows that the deterministic systems corresponding to different market types have sloppy directions. This means that in the sloppy directions, the information needed to calibrate the model simply would not be enough.

Hence, the calibration methods studied so far are probably adequately involved to match the dimension of applicability given by the parameter estimation problem in general. The methods perform similarly and it is not clear whether one of the approaches are preferable, but the Simulated annealing method with an adapted kernel density approximation yields lowest uncertainty bounds. If new method are to be studied, one idea is to develop a type of maximum likelihood estimation based on numerical Fokker-Planck solutions of the mispricing probability distribution function conditioned on the model parameters. Nevertheless, the insights from Sloppy model-theory suggests that the usage of a Sloppy model could still be rich, e.g., for forecasting purposes.

Thus, the prediction power of the model is studied. For the calibrated model in terms of the synthetic data, the forecasting performs seemingly well. The prediction distributions are thin and the actual outcomes are often located well inside the distributions. However, for an application of the parameter estimations methods on a real time-series taken from the American stock index S&P 500, the performance of the prediction is poor. The prediction distributions are still thin, but the true outcomes are located in the tails. The explanation could be that the sample time-series is taken from a period of a burst of a bubble into equilibrium pricing, and that the calibrated model only describes this behaviour (as can be seen in the corresponding deterministic phase plots).

Therefore, to enhance the performance of the prediction power, dynamic parameter estimates depending on the state of the mispricing should be considered, and the time-series for calibration should be chosen accordingly. This has not been studied in this thesis, and could be the starting point of further studies. In particular, the study of the prediction performance and the real world applicability of the model is highly relevant for understanding if the model by Yukalov et al. has explanatory insights on asset pricing, or if it is a complex way to describe historical data.

In contrast to earlier studies of the asset pricing model, the fundamental price has been modelled as deterministic. This is a less sophisticated choice, but tries to increase the transparency of the fundamental properties of the model. To get an even deeper understanding of the sloppiness structure of the model, case-specific stochastic Sloppy model-tools might be possible to find to allow for comparison with the results from the deterministic system. There is no clear way how to do this as of now, but there are other more straight forward analyses that could be conducted. For instance, the model manifold and the intrinsic sloppiness of the model should be studied for improved knowledge about what parts of the sloppiness that are model specific and what parts that are dependent on the parametrization.

Furthermore, the study of a real-world financial time-series suggests that the assumption on a deterministic fundamental price presumably is too simplistic. For the model to be useful, small changes in market prices should not be explained by movements in the mispricing. In addition, the underlying idea of the mispricing approach is to be able to distinguish between different regimes, not small fluctuations in the market price. Luckily, the Simulated annealing method is rather robust with respect to this assumption. Thus, it might not be of highest priority to analyse stochastic fundamental prices in a first further study.

In conclusion, this Master thesis mainly improves on earlier works in terms of insights about why the calibration of the asset pricing model is hard, cf., the model's sloppiness structure. Furthermore, it indicates that the usage of the model for forecasting is not completely straight forward, since dynamic approaches might be needed. The long-term objective of usage of the model for forecasting transition probabilities between different mispricing regimes is distant, but still seems tractable with base in an enhanced prediction power analysis of more sophisticated parameter dynamics.

References

[1] M. Arulampalam, S. Maskell, N. Gordon and T. Clapp, 2002, *A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking*. IEEE Transactions on Signal Processing, 50(2), 174-188.

[2] L. Bachelier, 1900, *Théorie de la Spéculation*. Annales Scientifiques de l'École Normale Supérieure, 3(3), 21-86.

[3] A. Bertolace, 2009, *Study of a Nonlinear Model of the Price of an Asset: Kalman Filter Calibration to data*. `http://www.er.ethz.ch/publications/MAS_Thesis_Bertolace.pdf`

[4] C. Berzuini and W. Gilks, 2001, *Following a Moving Target - Monte Carlo Inference for Dynamic Bayesian Models*. Journal of the Royal Statistical Society, 63(1), 127-146.

[5] F. Black and M. Scholes, 1973, *The Pricing of Options and Corporate Liabilities*. The Journal of Political Economy, 81, 637–654.

[6] K. Burke and G. Kendall, 2005, *Search Methodologies - Introductory Tutorials in Optimization and Decision Support Techniques*. Springer Science and Business Media, New York.

[7] K. Brown, C. Hill, G. Calero, C. Myers, K. Lee, J. Sethna and R. Cerione, 2004, *The Statistical Mechanism of Complex Signaling Networks: Nerve Growth Factor Signaling*. Physical Biology, 1, 184-195.

[8] K. Brown and J. Sethna, 2003, *Statistical Mechanical Approaches to Models with Many Poorly Known Parameters*. Physical Review, 68.

[9] S. Cecchetti, P. Lam and N. Mark, 1990, *Mean Reversion in Equilibrium Asset Prices*. American Economic Review, 80, 398-418.

[10] P. Christensen and K Larsen, 2012, *Incomplete Continuous-time Securities Markets with Stochastic Income Volatility*. `http://arxiv.org/abs/1009.3479v2`.

[11] K. Cranmer, 2001, *Kernel Estimation in High-Energy Physics*. Computer Physics Communications, 136, 198-207.

[12] C. Darwin, 1859, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle of Life.* Journal of Researches During H.M.S. Beagle's Voyage round the world.

[13] A. Fraser, 1957, *Simulation of Genetic Systems by Automatic Digital Computers. II: Effects of Linkage on Rates under Selection.* Australian Journal of Biological Science, 10, 492–499.

[14] D. Goldberg, 2002, *Design of Innovation: Lessons From and For Competent Genetic Algorithms.* Kluwer, Boston, Massachusetts.

[15] J. Hamilton, 1989, *A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle.* Econometrica, 57, 357-384.

[16] H. He, and H. Leland, 1993, *On Equilibrium Asset Price Processes.* Review of Financial Studies, 6, 593-617.

[17] O. Hellwich, 1998, *Model Parameter Estimation using Simulated Annealing.* International Archives of Photogrammetry and Remote Sensing, 32, 233–238.

[18] F. Hoffmeister and T. Bäck, 1991, *Genetic Algorithms and Evolution Strategies: Similarities and Differences.* Springer, Berlin, Heidelberg.

[19] J. Holland, 1975, *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, Michigan.

[20] S. Kirkpatrick, C. Gelatt and M. Vecchi, 1983, *Optimization by Simulated Annealing.* Science, 220(4598), 671-680.

[21] A. Koenig, 2002, *A Study of Mutation Methods for Evolutionary Algorithms.* Advanced Topics in Artificial Intelligence, CS 447.

[22] D. Koller and N. Friedman, 2009, *Probabilistic Graphical Models.* MIT Press, Cambridge, Massachusetts.

[23] A. Kong, J. Liu, and W. Wong, 1994, *Sequential Imputations and Bayesian Missing Data Problems.* Journal of the American Statistical Association, 89, 278–288.

[24] W. Lin, W. Yuan and T. Hong, 2003, *Adapting Crossover and Mutation Rates in Genetic Algorithms.* Journal of Information Science and Engineering, 19, 889-903.

[25] D. Luenberger and P. Woehrmann, 2007, *On kNN Density Estimation.* National Centre of Competence in Research - Financial Valuation and Risk Management, Working Paper no. 417.

[26] Benjamin Machta, Ricky Chachra, Mark Transtrum and James Sethna, 2013, *Parameter Space Compression Underlies Emergent Theories and Predictive Models.* Science, 342, 604-607.

[27] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, 1953, *Equation of State Calculations by Fast Computing Machines*. Journal of Chemical Physics, 21(6), 1087-1092.

[28] S. Robert, 2012, *Sequential Monte Carlo Methods for a Dynamical Model of Stock Prices*. `http://www.er.ethz.ch/publications/MAS_Thesis_Sylvain_Robert_Aug2012.pdf`

[29] A. Savitzky and M. Golay, 1964, *Smoothing and Differentiation of Data Simplified Least Squares Procedures*. Analytical Chemistry, 36(8), 1627-1639.

[30] D. Sornette, 2004, *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, Princeton.

[31] D. Sornette and P. Cauwels, 2014, *1980-2008: The Illusion of the Perpetual Money Machine and What it Bodes for the Future*. Risks, 2, 103-131.

[32] M. Transtrum, B.Machta and J. Sethna, 2012, *Why are Nonlinear Fits to Data so Challenging?* Physical Review Letters, 104(6), 060201.

[33] M. Wand and M. Jones, 1995, *Kernel Smoothing*. Monographs on Statistics and Applied Probability, 60.

[34] V. Yukalov, 1991, *Method of Self-Similar Approximations*. Journal of Mathematical Physics, 32(5), 1235–1239.

[35] V. Yukalov, D. Sornette and E. Yukalova, 2009, *Nonlinear Dynamical Model of Regime Switching Between Conventions and Business Cycles*. Journal of Economic Behaviour & Organization, 70(1-2), 206-230.

Appendices

Appendix A

Mathematical Derivations and Expressions

In this chapter, some mathematical results for the asset pricing model by Yukalov et al. are derived and presented.

A.1 Hessian Matrix of the Cost Function at Minimum

This section treats the mathematical derivation of the expression for the Hessian matrix of the deterministic system given by Equation (4.6).

Lemma A.1.1 *For the cost function*

$$\chi^2(\theta) = \frac{1}{2} \sum_{t\in\{0,1,\ldots,T-1\}} (f^{NL}(X_t, Y_t) - \bar{y}_t^{obs})^2, \tag{A.1}$$

the Hessian matrix at the minimizing point in parameter space is given by

$$H_{i,j}^{\chi^2} = \sum_{t\in\{0,1,\ldots,T-1\}} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_i} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_j}. \tag{A.2}$$

Proof $H_{i,j}^{\chi^2}$ is given by

$$\begin{aligned} H_{i,j}^{\chi^2} &= \frac{\partial}{\partial \theta_j} \frac{\partial}{\partial \theta_i} \chi^2 \\ &= \frac{\partial}{\partial \theta_j} \sum_{t\in\{0,1,\ldots,T-1\}} (f^{NL}(X_t, Y_t) - \bar{y}_t^{obs}) \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_i} \\ &= \sum_{t\in\{0,1,\ldots,T-1\}} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_j} \frac{\partial f^{NL}(X_t, Y_t)}{\partial \theta_i} + \sum_{t\in\{0,1,\ldots,T-1\}} (f^{NL}(X_t, Y_t) - \bar{y}_t^{obs}) \frac{\partial^2 f^{NL}(X_t, Y_t)}{\partial \theta_j \partial \theta_i}. \end{aligned}$$

At the minimizing point, the difference $f^{NL}(X_t, Y_t) - \bar{y}_t^{obs}$ equals zero, and the claim follows.

□

A.2 Calculation of the Elements of the Hessian Matrix

The elements of the Hessian matrix given by Equation (4.6) is easily calculated from the following derivatives:

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial \alpha} = X_t, \tag{A.3}$$

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial A} = X_t^3 \exp(-X_t^2/\mu^2), \tag{A.4}$$

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial \beta} = Y_t, \tag{A.5}$$

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial B} = Y_t^3 \exp(-Y_t^2/\lambda^2), \tag{A.6}$$

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial \mu} = \frac{AX_t^5 \exp(-X_t^2/\mu^2)}{\mu^3}, \tag{A.7}$$

$$\frac{\partial f^{NL}(X_t, Y_t)}{\partial \lambda} = \frac{BY_t^5 \exp(-Y_t^2/\lambda^2)}{\lambda^3}. \tag{A.8}$$

Appendix B

Results

In this chapter, supplemental results not directly stated in the report from the calibration of synthetic data as well as parameter estimation of the S&P 500 time-series are presented.

B.1 Calibration of Synthetic Data

Supplemental results from calibration of the time-series of synthetic data both for the Evolutionary algorithm and the Simulated annealing method are presented in this section. In addition, the sensitivity of y in predictions are investigated.

B.1.1 Evolutionary Algorithm Supplemental Results

For the Evolutionary algorithm, the supplemental results are presented in Figures B.1 - B.3 and in Tables B.1 and B.2 on the following pages.

B.1.2 Simulated Annealing Supplemental Results

The supplemental Simulated annealing results are summarized in Tables B.3 and B.4 on the following pages.

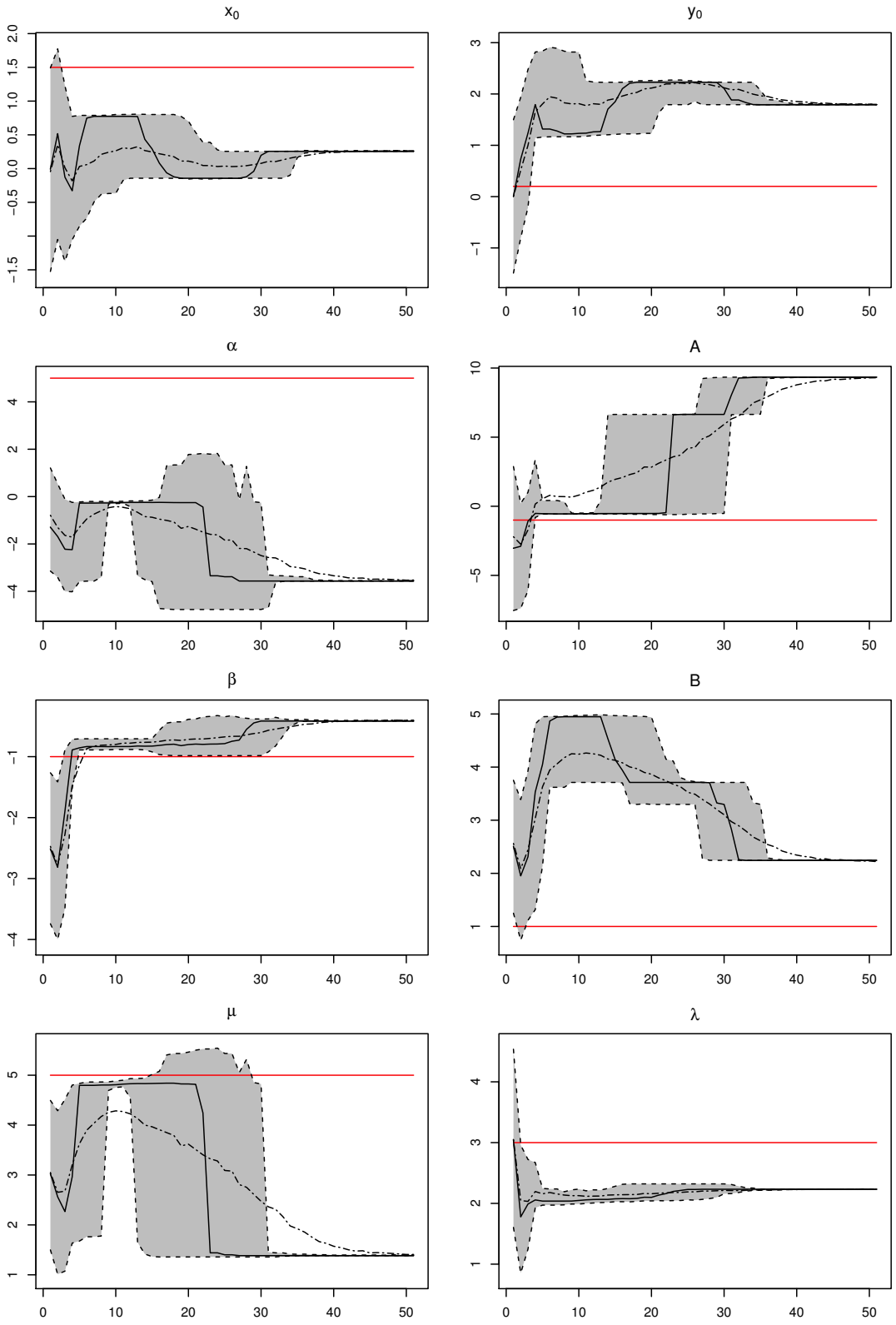

Figure B.1: Market 10. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

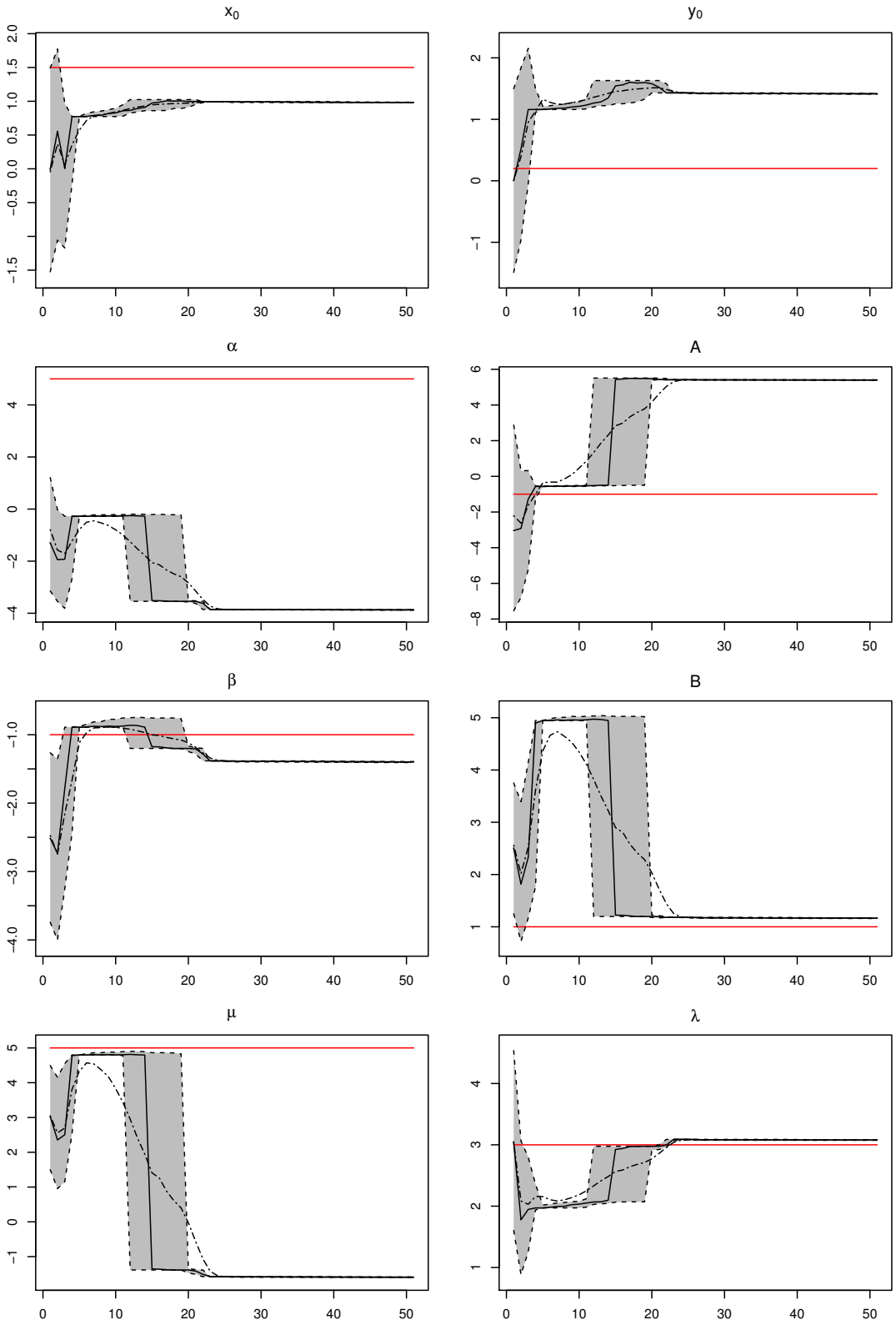

Figure B.2: Market 10, deterministic. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

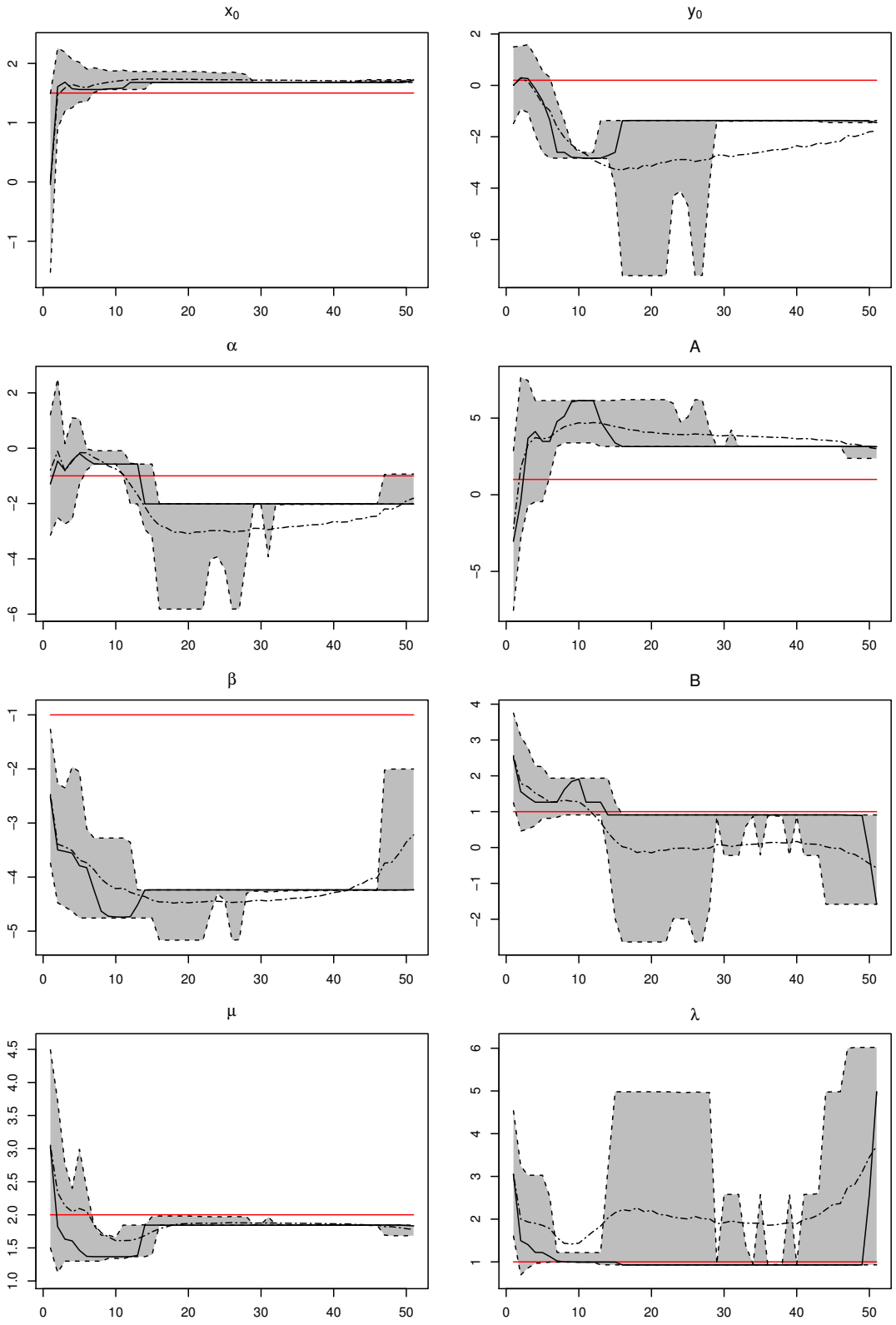

Figure B.3: Market 15, deterministic. Calibration results of initial conditions (x_0, y_0) and parameters α, A, β, B, μ and λ using the 10% fittest chromosomes out of 20000 and 50 generations (x-axis). The grey areas are the inter-quartile ranges, the solid lines the medians and the dashed lines the means. The red lines correspond to the true values.

Table B.1: Descriptive statistics (median and standard deviation) of the 10 % fittest chromosomes of the extended Evolutionary algorithm calibration with 25 parallel runs of the algorithm with 20000 chromosomes in each run. The sign(parameter) is the percentage of chromosomes estimated with the correct sign for the corresponding parameter, and sign(θ) is the portion of chromosomes with all genes having the correct sign. Note that the signs of μ and λ are irrelevant.

	m(x_0)	sd(x_0)	sign(x_0)	m(y_0)	sd(y_0)	sign(y_0)	m(α)	sd(α)	sign(α)	m(A)	sd(A)	sign(A)
Market 1	1.29	0.22	1.00	2.33	0.86	0.99	0.12	0.83	0.44	-5.89	2.23	1.00
Market 1 0	0.68	0.78	0.87	1.34	1.61	0.71	-4.24	2.37	0.08	5.57	4.00	0.13
Market 15	1.85	0.41	1.00	-1.81	1.12	0.04	-0.45	1.86	0.61	1.56	1.59	0.61

	m(β)	sd(β)	sign(β)	m(B)	sd(B)	sign(B)	m(μ)	sd(μ)	sign(μ)	m(λ)	sd(λ)	sign(λ)
cont. Market 1	-1.27	0.30	1.00	1.25	1.81	0.80	2.94	1.49	-	0.50	0.52	-
cont. Market 10	-1.04	1.08	1.00	1.57	1.02	1.00	1.69	1.12	-	2.81	0.44	-
cont. Market 15	-2.25	1.40	0.99	1.05	1.80	0.81	1.86	1.55	-	1.17	1.42	-

	sign(θ)
cont. Market 1	0.36
cont. Market 10	0.04
cont. Market 15	0.02

Table B.2: Descriptive statistics (median and standard deviation) of the 10 % fittest chromosomes of the extended Evolutionary algorithm calibration of the re-parametrized model with 25 parallel runs of the algorithm with 20000 chromosomes in each run. The sign(parameter) is the percentage of chromosomes estimated with the correct sign for the corresponding parameter, and sign(θ) is the portion of chromosomes with all genes having the correct sign. Note that the signs of μ and λ are irrelevant.

	m(x_0)	sd(x_0)	sign(x_0)	m(y_0)	sd(y_0)	sign(y_0)	m(α)	sd(α)	sign(α)	m(A)	sd(A)	sign(A)
Market 1	1.47	0.16	1.00	1.17	0.76	0.93	-2.09	0.96	1.00	-3.99	4.69	0.92
Market 1 0	0.58	0.68	0.97	1.79	1.44	0.76	-3.95	1.67	0.04	5.78	4.83	0.20
Market 15	1.88	0.38	1.00	-1.73	1.00	0.03	-0.40	1.15	0.88	2.31	1.25	0.88

	m(β)	sd(β)	sign(β)	m(B)	sd(B)	sign(B)	m(μ)	sd(μ)	sign(μ)	m(λ)	sd(λ)	sign(λ)
cont. Market 1	-0.70	0.22	1.00	2.14	1.10	0.99	0.55	5.70	-	1.58	4.43	-
cont. Market 10	-1.33	1.04	1.00	1.66	1.10	1.00	1.52	0.87	-	2.92	0.55	-
cont. Market 15	-2.04	1.12	0.98	2.39	1.66	0.93	1.59	2.15	-	0.87	1.04	-

	sign(θ)
cont. Market 1	0.85
cont. Market 10	0.04
cont. Market 15	0.02

Table B.3: Descriptive statistics (mean and standard deviation) of the 10 fittest ants of the Simulated annealing method approximating the probability distribution $\hat{p}(\theta|x)$ by the adaptive kernel estimation. The sign(parameter) is the percentage of the ants estimated with the correct sign for the corresponding parameter, and sign(θ) is the portion of ants with all parameters having the correct sign. Note that the signs of μ and λ are irrelevant.

	m(y_0)	sd(y_0)	sign(y_0)	m(α)	sd(α)	sign(α)	m(A)	sd(A)	sign(A)	m(β)	sd(β)	sign(β)
Market 1	1.23	0.49	1.00	-3.64	0.38	1.00	-1.03	0.58	1.00	-0.83	0.38	1.00
Market 10	2.55	0.66	1.00	-4.05	0.76	0.00	2.78	1.09	0.00	-1.33	0.72	1.00
Market 15	-0.48	1.02	0.40	-1.05	0.68	1.00	3.29	0.53	1.00	-6.78	0.67	1.00

	m(B)	sd(B)	sign(B)	m(μ)	sd(μ)	sign(μ)	m(λ)	sd(λ)	sign(λ)	sign(θ)
cont. Market 1	2.41	0.63	1.00	1.84	0.68	-	-0.29	0.68	-	1.00
cont. Market 10	2.91	0.84	1.00	0.63	0.62	-	2.50	0.30	-	0.00
cont. Market 15	2.09	0.93	1.00	1.69	0.12	-	0.64	0.59	-	0.40

Table B.4: Descriptive statistics (mean and standard deviation) of the 10 fittest ants of the Simulated annealing method approximating the probability distribution $\hat{p}(\theta|x)$ using kNN estimation. The sign(parameter) is the percentage of the ants estimated with the correct sign for the corresponding parameter, and sign(θ) is the portion of ants with all parameters having the correct sign. Note that the signs of μ and λ are irrelevant.

	m(y_0)	sd(y_0)	sign(y_0)	m(α)	sd(α)	sign(α)	m(A)	sd(A)	sign(A)	m(β)	sd(β)	sign(β)
Market 1	0.90	0.49	1.00	-4.23	0.39	1.00	0.46	2.92	0.70	-0.65	0.22	1.00
Market 10	-0.06	0.66	0.50	-4.65	0.31	0.00	6.19	1.39	0.00	-1.18	0.34	1.00
Market 15	-1.57	0.47	0.00	2.47	3.04	0.30	0.64	2.08	0.30	-5.18	1.77	1.00

	m(B)	sd(B)	sign(B)	m(μ)	sd(μ)	sign(μ)	m(λ)	sd(λ)	sign(λ)	sign(θ)	k
cont. Market 1	2.39	0.78	1.00	1.36	1.65	-	0.14	0.38	-	0.70	5
cont. Market 10	1.22	0.43	1.00	1.90	0.09	-	3.05	0.30	-	0.00	5
cont. Market 15	3.77	1.31	1.00	4.24	1.78	-	0.08	0.41	-	0.00	10

B.1.3 Sensitivity of y in Predictions

In this section, a stress test of the prediction performance on synthetic data when the assumption of perfect observation of y is alleviated is presented in Table B.5. Similar results hold for the Evolutionary algorithm, Simulated annealing with a kernel density as well as for Simulated annealing with a kNN density for market type 10. Only the sensitivity of the kNN method on market 1 and market 15 are shown for clarity.

Table B.5: Descriptive statistics of the estimated mispricing distribution $\hat{x}$ (mean and standard deviation) and the true outcome x after 1 week and 1 month, respectively, for different values of y and using the kNN method, cf., Table 4.5. $P(\hat{x} > x_{obs})$ gives a measure of the probability of more extreme events. The second column shows how many multiples of the true y that is used for the predictions.

		1 week				1 month			
		x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$	x_{obs}	m($\hat{x}$)	sd($\hat{x}$)	$P(\hat{x} > x_{obs})$
Market 1	y	0.55	0.55	0.01	0.70	0.55	0.53	0.02	0.16
	$0.5y$	0.55	0.55	0.01	0.78	0.55	0.54	0.02	0.28
	$2y$	0.55	0.55	0.01	0.55	0.55	0.52	0.02	0.04
	$-y$	0.55	0.56	0.01	0.92	0.55	0.57	0.02	0.76
	$-0.5y$	0.55	0.56	0.01	0.88	0.55	0.55	0.02	0.61
	$-2y$	0.55	0.57	0.01	0.97	0.55	0.58	0.02	0.94
Market 15	y	3.25	3.25	0.01	0.70	3.29	3.27	0.02	0.23
	$0.5y$	3.25	3.25	0.01	0.61	3.29	3.26	0.02	0.11
	$2y$	3.25	3.25	0.01	0.87	3.29	3.29	0.02	0.55
	$-y$	3.25	3.24	0.01	0.35	3.29	3.23	0.02	$3.9 \cdot 10^{-3}$
	$-0.5y$	3.25	3.25	0.01	0.43	3.29	3.24	0.02	0.02
	$-2y$	3.25	3.24	0.01	0.17	3.29	3.21	0.02	$02.7 \cdot 10^{-4}$

B.2 Calibration of a S&P 500 Time-Series

In this section, the supplemental results from the calibration studies on the S&P 500 time-series are presented.

B.2.1 Phase Plots of Deterministic Systems

The phase plot given by the deterministic counterpart of the calibrated system is presented in this section.

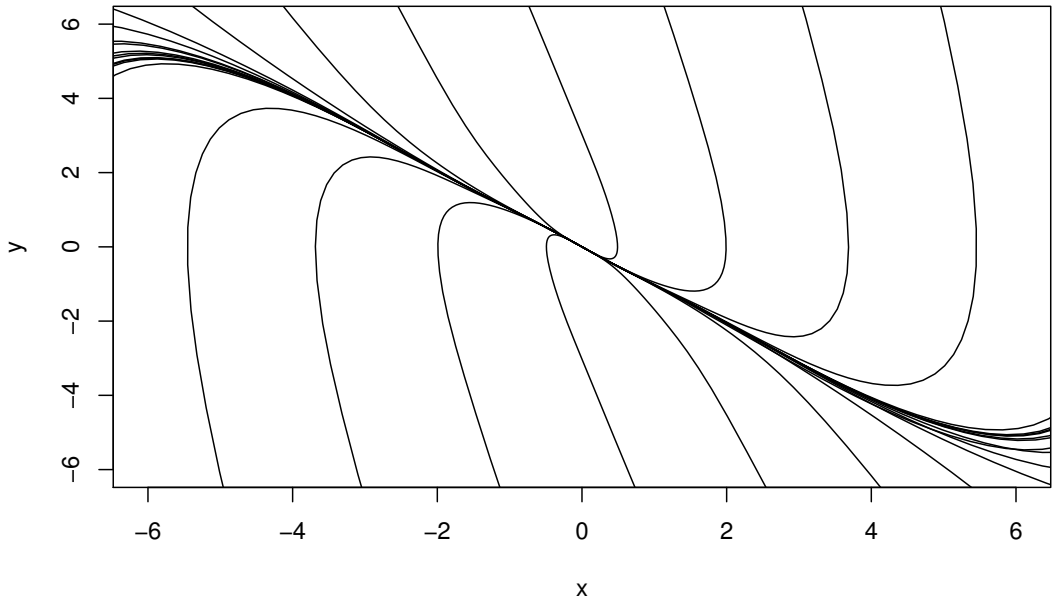

Figure B.4: Phase plot of the, by the Evolutionary algorithm, calibrated deterministic system for the S&P 500 time-series. The estimates used for the parameters are the median values found in Table 5.1.

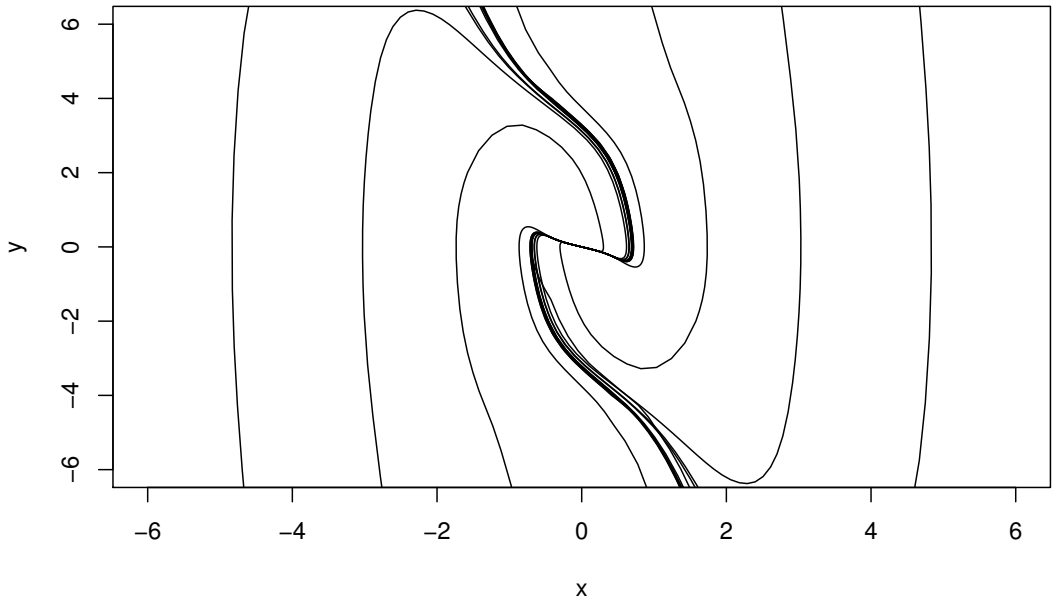

Figure B.5: Phase plot of the calibrated deterministic system for the S&P 500 time-series when using the Simulated annealing method with the adapted kernel density estimator. The estimates used of the parameters are the median values found in Table 5.4 for $\mu_f = 0.04$.

B.2.2 Mispricing Time-Series for $\mu_f = 0.06$

As a comparison with Figure 5.1 with $\mu_f = 0.04$, the corresponding graph for $\mu_f = 0.06$ is presented in this section.

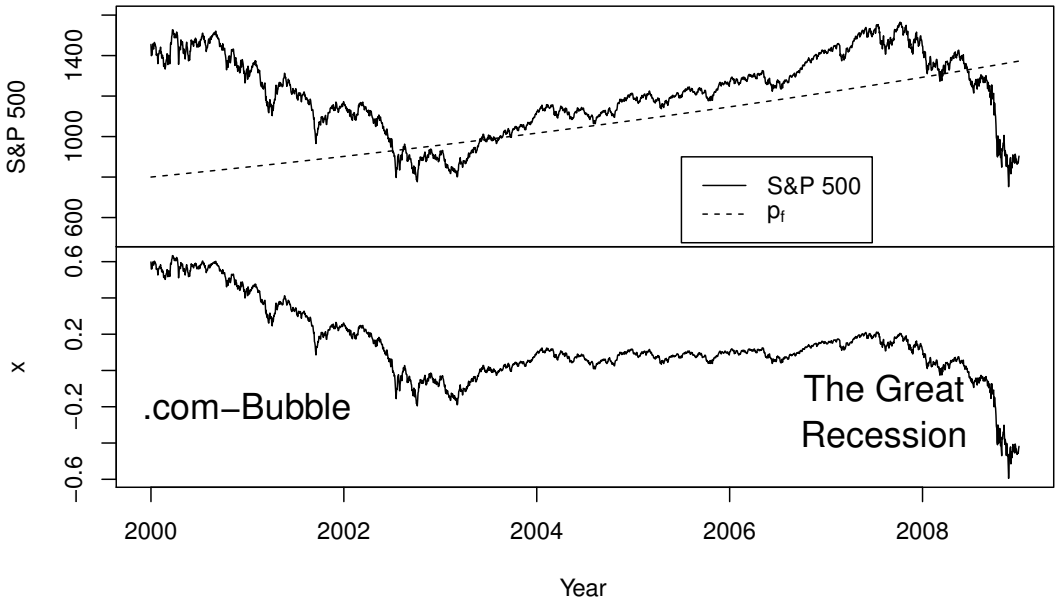

Figure B.6: Time-series of the S&P 500-index from 2000 to 2008 and the extracted mispricing x using $\mu_f = 0.06$, emphasising the two most important financial events during this time period.