THESIS / THÈSE

MASTER IN BUSINESS ENGINEERING PROFESSIONAL FOCUS IN DATA SCIENCE

Simulated Bayesian estimation of a financial agent-based model

Werland, Arthur

Award date:
2020

Awarding institution:
University of Namur

Simulated Bayesian estimation of a financial agent-based model

Arthur WERLAND

Directeur: Prof. S. BEREAU

Mémoire présenté
en vue de l'obtention du titre de
Master 120 en ingénieur de gestion, à finalité spécialisée
en data science

ANNEE ACADEMIQUE 2019-2020

Université de Namur, ASBL
Faculté des Sciences économiques, sociales et de gestion – Département des Sciences de gestion

Rempart de la Vierge 8, B-5000 Namur, Belgique, Tel. +32 [0]81 72 48 41/49 58, Fax +32 [0]81 72 48 40

Abstract

This master thesis attempts to introduce a Bayesian estimation which used the Non-Parametric Simulated Maximum Likelihood Estimator (NPSMLE) in order to compute an approximation of the agent-based model likelihood. The main achievement of this master thesis is the creation of an adaptive Gibbs sampler which takes into account the shape of the likelihood function in order to explore the parameter space in a proper way.

To test the accuracy of our Gibbs sampler, laboratory experimentation have been conducted in order to assess the extent to which our simulated Bayesian method is better than the NPSMLE. Finally, our Bayesian method has been tested on the S&P500 index in normal and crisis economic conditions. We used the most famous Heterogeneous Agent Models, the Adaptive belief system (Brock and Hommes, 1998). This new Bayesian method is more accurate than its frequentist equivalent in laboratory conditions because it has a smaller variance and bias. The sample size of our bias-variance analysis is small, therefore it constitutes a strong limitation to our work. Cloud computing has to be used in order to increase this sample and have a better estimation of the true bias and variance of our estimator. To conclude, it is clear that a paradigm shift is necessary in order to improve calibration accuracy in agent-based models.

Acknowledgments

I would like to express my grateful to my supervisor Sophie Béreau for her number of advice and her knowledge sharing during the creation of my master thesis. I also would like to thank all the teachers and assistants at the University of Namur for their learning through all my studies.

I would like to express my gratitude to all my family and friends who supported me during all my studies.

I must express my very profound gratitude to my mother who helped during the writing of my master thesis and my girlfriend for supporting and encouraging me every day throughout this year.

Thank you.

Contents

List of abbreviations

- ABM: Agent-Based Model
- HAM: Heterogeneous Agent Models
- ABS: Adaptive Belief System
- ACF: Agent-Based Computational Finance
- ACFM: Agent-based Computational Finance Model
- EMH: Efficient Market Hypothesis
- DGP: Data Generating Process
- rwDGP: real-world DGP
- mDGP: model-DGP
- GA: Genetic Algorithm
- MCMC: Markov Chain Monte Carlo
- NPSMLE: NonParametric Simulated Maximum Likelihood Estimator

Introduction

For a long time, the efficient market hypothesis (Fama, 1970) was the dominant hypothesis in finance but also in the economy in general. But the rational framework suffers from too restrictive assumptions about agents' behaviour, for instance a rational behaviour implies that agents are able to know all traders' strategies in the market in order to price the security rationally. It is pretty obvious that kind of assumption will not stands for the market without biases. However, this extremely simplified version of the reality is not totally wrong because this theory can be seen as a theoretical benchmark to confront emergent theories.

Behavioural finance has become appealing because this paradigm is able to address issues that the traditional finance was not enabled to face during many years. One of the biggest issues with the traditional paradigm is that stylized facts found in financial stock can not be explained in a rational framework. In fact, through the last three decades, behavioural finance has received a lot of attention from academic literature because this theory is much more realistic than the former with less restrictive assumptions about traders and the market as a whole. Moreover, behavioural finance is able to resolve some puzzles which are unanswered for a long time[1].

Behavioural models have received a lot of attention in order to construct new models based on agents who have more realistic behaviours as cognitive biases, limited cognition, learning or interaction. Those models are called Heterogeneous Agent Models

[1]For more argument on behavioural finance, see the excellent survey of Barberis and Thaler (2003).

(HAMs) if they are very simple, they only stand for a simplified version of the reality. These models are composed with boundedly rational agents using simple rules to predict the future stock price, these agents are heterogeneous and they are able to switch from a strategy to another based on profitability heuristics. Simplest HAMs are analytically tractable (at least numerically), but if we move to models which are closer to the reality, computational approaches are needed. These more complex models are generally called agent-based computational economic models. More precisely, in this master thesis, we will use this kind of model in the sub-field of finance, therefore we called this type of model, Agent-based Computational Finance Models (ACFMs). These models are constructed along the exponential increased of new technologies and optimization methods which are largely due to the increasing popularity of the machine learning. In fact, these models are called *computational* because they are strongly linked to the utilization of numerical methods and algorithms which are able to solve these models.

In this master thesis, three main assumptions have been made in order to use HAMs and ACFMs in financial markets. First of all, agents are boundedly rational. As argued by Simon (1955), the empirical refutation of the EMH does not mean that agents are irrational but more, they are limited in their degree of rationality. In other words, agents make choices which are satisfactory, not especially optimal. Secondly, agents have heterogeneous expectation and use simple forecasting rules to make predictions about the future market prices. Finally, the performance of their strategies is evaluated according to profitability measures, hence they are able to learn from historical performance of strategies in order to adapt their forecasting rule properly. Moreover, agents are supposed to be able to interact with each other. The evolution of the market fraction[2] and the interactions between agents both influence endogenously the market prices of stock, which means that the price is not solely influenced by exogenous news shocks.

[2]The market fraction is the fraction of the different types of trading strategies that exist in financial markets.

Interaction and learning are both encapsulated into the concept of complex adaptive system in ACFMs. A complex system is, as quoted by Tesfatsion (2006, p.7) :

> "The system exhibits emergent properties, that is, properties arising from the interactions of the units that are not properties of the individual units themselves. "

And then a complex adaptive system is, as quoted by Tesfatsion (2006, p.7) :

> "A complex adaptive system is a complex system that includes goal-directed units, i.e., units that are reactive and that direct at least some of their reactions towards the achievement of built-in (or evolved) goals."

During the last three decades, new ACFMs have emerged in the academic literature. The complexity in term of learning, interaction and heterogeneity of these models has increased more and more until these models have become unable to be estimated with direct methods. In other words, these new complex models are usually not tractable (analytically or numerically), therefore researchers are not able to derive a closed-form solution of their likelihood function and their moments conditions. Direct estimation as Maximum likelihood or OLS can not be used in order to calibrate the parameters of these new complex models. Hence, new methods have to be created in order to estimate these models, simulation-based econometric methods[3] are able to overcome the fact that these models are not tractable (analytically or numerically). Indirect frequentist approach has been mainly discussed last two decades in the empirical literature, but the Bayesian paradigm is left unexplored to indirect estimation. The only attempt to use Bayesian inference in ACFMs is made by Grazzini et al. (2017). They proposed a simplified version of a Bayesian inference with the most simple Markov chain Monte Carlo, the Metropolis-Hastings algorithm. However, their first attempt to estimate

[3]In this master thesis, these methods are called indirect estimation methods

ACFMs with Bayesian inferences works well, their method produces accurate estimation of most of the parameters, but their method is limited in order to explore the parameter space accurately when the dimensionality increases.

Bayesian estimation seems to be largely ignored in the empirical literature (Platt, 2020). But as empirically demonstrated by Platt (2020), in a wide range of cases, Bayesian inference outperforms other up to date frequentist methods. Hence, he suggests that a paradigm shift is required, with improvement of existing Bayesian estimation techniques. This master thesis proposes an innovative method to use Bayesian inference with ACFMs[4] in order to address the lack of methodology in sampling methods. The main achievement of this master thesis is the creation of an adaptive Gibbs sampler which takes into account the shape of the likelihood function. We use the method introduced by Kristensen and Shin (2012) to compute an approximation of the likelihood function, this method is called the NonParametric Simulated Maximum Likelihood Estimator (NPSMLE). Our simulated Bayesian method can be seen as an extension of Kukacka and Barunik (2017) because they used the NPSMLE to estimate an HAM, the adaptive belief system created by Brock and Hommes (1998). Our empirical estimation is first made in laboratory conditions where the true value of parameters are known. This section will estimate the parameters of a two-type adaptive belief system created by Brock and Hommes (1998) in order to assess the extent to which this Bayesian method is able to retrieve the true value of the parameters. The main purpose of this section is to compare the accuracy in term of bias and variance of the classic NPSMLE and our Bayesian method. Finally, this method will be used to estimate a two-type adaptive belief system on the S&P500 index in normal and crisis economic conditions.

In order to collect all required information in this master thesis, an academic review has been made first. This step is essential to gather all information about HAMs and

[4]This method can also be used with HAMs.

ACFMs created in the academic literature and how these models are estimated and validated in practice. The second step of this master thesis is to gather all the theoretical background necessary in order to construct our new Bayesian estimation method. The final step of this thesis is to test this new method in laboratory conditions[5] and on real financial data. The papers which have inspired the general structure of this thesis are ; Béreau (2014), Chen et al. (2012), Kukacka and Barunik (2017) and Fagiolo et al. (2019). Béreau (2014), Chen et al. (2012) and Fagiolo et al. (2019) have inspired the general structure of the literature review. The structure of Kukacka and Barunik (2017) have been naturally followed by the empirical estimation chapter because this master thesis can be seen as an extension of this paper.

The present master thesis is organized as follows: the first chapter is dedicated to review the literature review. This chapter in subdivided in two sections : the first one is devoted to a literature review on agent-based model and the second discusses empirical aspect of agent-based model estimation. The second chapter presents all the theoretical background and methodology necessary to construct our new simulated Bayesian estimator. The third chapter addresses the results of empirical estimation and their interpretation. Finally, the last part attempts to draw overall conclusions of our work and limitation which suggest further researches.

[5]Laboratory conditions means that the true value of parameters are known.

Chapter 1

Literature review

1 Agent-based models

> "Agent-based modeling is a computational method that enables a researcher to create, analyze, and experiment with models composed of agents who interact within an environment." (Gilbert, 2008, p.1)

Agent-based model (ABM) is a widely used model around many disciplines, as in ecology, economics, geography, supply chain, stock market and many others. This type of model is useful when the modeller has to deal with complex adaptive system[1] composed of heterogeneous agents who interact with each other. The behaviours of agents are described by simple rule and are influenced by the behaviours of other agents. The outcomes of those models show some regularities, structure and behaviour which are not explicitly programmed, but arise from the behaviours and interaction of agents (Macal et al., 2010).

The popularity of ABMs is nowadays strongly related to the fast computing power

[1] A system is complex if it exhibits emergent properties, that is, properties arising from the interactions of agents. A complex adaptive system is a complex system that includes goal-directed units, i.e., units that are reactive and that direct at least some of their reactions towards the achievement of built-in (or evolved) goals. Both definitions came from Tesfatsion (2006, p.7)

of our contemporaneous computers. In fact, in most cases ABMs do not have aggregation equation which can be analytically derived, so classic econometric methods cannot be applied. Fortunately, Simulation-based Econometric Methods (SEM) have been created to deal with this type of models which cannot be analytically derived. In practice, Monte Carlo simulation can be applied to derive aggregate function and econometric methods are applied directly on simulated aggregations (Chen et al., 2012), the section 2.1 will describe this step in details. However, this kind of econometric methods needs a huge number of simulations to deal with agent-based models and therefore is strongly dependant on the computing power of the modeller's computers. Moreover, parallelization of the computation enables to reach a much larger number of simulations in a much shorter time. The significant improvement in computers' technologies enables ABMs to be used by a lot of researchers in many disciplines.

Behavioural finance argues that agents are subject to cognitive biases when they form their beliefs and on people's preference (Tversky and Kahneman, 1974), these biases have two major consequences in the modelization of the economy. First of all, agents are not fully rational that means that agents do not act in a way to maximise their expected outcomes. A number of experimental works have shown that people systematically violate the expected utility framework (EU) in practice (Barberis and Thaler 2003). Secondly, it is impossible that all agents are subject to cognitive bias in the same way and intensity. Hence, it is unlikely that agents are the same in the economy, the high trading volume strengthens the idea of heterogeneity in financial market[2]. Thus, financial markets are composed of different types of agents and they are non fully rational. But as recalled by Barberis and Thaler (2003), experimental evidence on cognitive bias should not be taken for granted because people are able to learn from experience or with more powerful incentive, therefore psychological bias can decrease in intensity or completely disappear. So, agents are able to learn from their previous mistakes (bias) but they immediately violate it in a specific application.

[2]The section 1.2.1 will discuss extensively on the heterogeneity in financial markets.

Behavioural finance suggests that agents are not fully rational, heterogeneous and are able to learn from their experiences. ABMs seem to be perfect to mimic the financial market. Moreover, as suggested by Tesfatsion (2003, p.1):

> "Decentralized market economies are complex adaptive systems, consisting of large numbers of adaptive agents involved in parallel local interactions. These local interactions give rise to macroeconomic regularities"

In finance, regularities can be seen as the stylized facts which emerge from bounded rational agents who are able to learn from their previous experiences and interact with other agents in local clusters. ABMs seem to be the perfect way to represent financial markets in a simplified version of the reality (Gilbert, 2008), with only a few types of agents who interact locally, follow simple rules and are able to learn. The following section will provide all hypothesis that this master thesis makes in order to use ABMs in financial markets.

1.1 Hypotheses

We have to formulate three main hypotheses to use ABMs in financial markets. First of all, agents are considered as bounded rational, secondly they are able to learn from past strategies performance and finally they are heterogeneous.

1.1.1 Bounded rationality

As already introduced above, agents are subject to cognitive bias which can have a strong influence on their behaviour when they have to take a decision (Barberis et al. 2003). Moreover, this influence is strengthened when agents are facing risks, uncertainty, incomplete information or when the complexity of the task is very high (Tseng, 2006). All those factors lead to question the efficient market hypothesis (EMH) and more precisely the idea that agents behave rationally. As a reminder, EMH is composed of three assumptions (Shleifer, 2000). First of all, market participants are assumed to be rational and therefore are able to price securities rationally. Secondly, if a share of investors is not rational, their trades are random therefore on average their trades will be cancelled with each other without affecting prices. Thirdly, to the extent that a share of investors is irrational in similar ways, these trades are cancelled in the market by rational arbitrageurs who eliminate their influences on prices. Let's focused on the first assumption of the EMH, according to Barberis and Thaler (2003, p.2) :

> "Rationality means two things. First of all, when traders receive new information, agents update their beliefs correctly, in a manner described by Bayes' law. Secondly, given their beliefs, agents make choices that are normatively acceptable, in the sense that they are consistent with Savage's notion of Subjective Expectation utility (SEU)."

Let's focus on the first part of this quote. As shown by Kahneman and Tversky (1974) people apply the Bayes law incorrectly. When people try to evaluate the probability that a data set was generated by a model, they use the representative heuristics. This means that they are trying to evaluate this probability by evaluating the probability that the model has generated the data set. But the representative heuristics generate some biases. One of the biases is the base rate neglect. In fact, when agents estimate the probability that their data set was generated by a model, they put too much weight on $P(data|model)$[3] which capture the representativeness but too little weight on the prior. In the Bayes' rule $P(model)$ represent the prior.

$$P(model|data) = \frac{P(data|model)P(model)}{P(data)} \tag{1}$$

In practice, agents are not able to apply the Bayes' rule correctly which is not consistent with the first assumption of the EMH. Moreover, a number of experimental works has shown that people systematically violate the expected utility framework (EU) in practice (Barberis and Thaler, 2003). EMH does not seem to be the perfect

[3]This term is captured by the likelihood of the data given a model.

framework to modelize the economy because its assumptions are too strict to be fulfilled in practice. Herbert A. Simon has introduced the concept of bounded rationality in 1957. According to Simon (1997, vol 3, p. 291):

> "The term 'bounded rationality' is used to designate a rational choice that takes into account the cognitive limitations of the decision-maker, limitations of both knowledge and computational capacity. Bounded rationality is a central theme in the behavioral approach to economics, which is deeply concerned with the ways in which the actual decision-making process influences the decisions that are reached."

Considering market participants as bounded rational seems a much better assumption than modelizing agents as fully rational in financial markets. More precisely, bounded rational theory as described by Simon, relaxes some purely theoretical assumptions supported by the EMH. Simon (1955)[4] suggests that it is almost impossible that agents are able to perform this kind of optimization required in the rational framework. He argued that because optimization costs too much and agents have limited cognition, a more natural way to consider agents is to see them as "satisfying". In other words, agents make choices which are satisfactory, not especially optimal. The term bounded rational does not mean that they are irrational but more they are limited in their degree of rationality (Lo, 2004).

1.1.2 Complex adaptive system

Another important feature to use ABMs in financial market is to assume that agents are able to learn from their past strategies to adapt their belief on the future stocks price. A boundedly rational agents forms their expectations about the future based upon observable quantities and are able to adapt their forecasting rule as soon as additional information become available (Hommes, 2006). Moreover, we have to make the hypothesis that agents are able to interact with each others. For instance, Föllmer

[4] This information come from Lo (2004)

(1974) shows that short range interaction among agents may propagate through the economy and lead to macro-level uncertainty causing a breakdown of price equilibria (Hommes, 2006).

Hence, stock market is composed of interacting and adapting agents which lead to the definition of complex adaptive system given by Tesfatsion (2002, p.1) :

> "Decentralized market economies are complex adaptive systems, consisting of large numbers of adaptive agents involved in parallel local interactions. These local interactions give rise to macroeconomic regularities such as shared market protocols and behavioral norms which in turn feed back into the determination of local interactions. The result is a complicated dynamic system of recurrent causal chains connecting individual behaviors, interaction networks, and social welfare outcomes."

In finance, the stock market's regularities can be seen as stylized fact which emerges from bounded rational agents who are able to learn from their experience and interact with other agents in local clusters.

Let's define more formally what a complex system is and then what a complex adaptive system is.

A system is considered as complex if it shows the following two properties (Tesfatsion 2006, p.7)[5]:

- The system is composed of interacting agents.
- The system exhibits emergent properties, that is, properties arising from the interactions of agents.

Unfortunately, there is not consensus on the definition of a complex adaptive system. But three nested definitions are considered (Tesfatsion, 2006, p.7):

[5] See Flake (1998) for more information on the properties of complex system.

- **Definition 1**: A complex adaptive system is a complex system that includes reactive units, i.e., units capable of exhibiting systematically different attributes in reaction to changed environmental conditions.

- **Definition 2**: A complex adaptive system is a complex system that includes goal-directed units, i.e., units that are reactive and that direct at least some of their reactions towards the achievement of built-in (or evolved) goals.

- **Definition 3**: A complex adaptive system is a complex system that includes planner units, i.e., units that are goal-directed and that attempt to exert some degrees of control over their environment to facilitate achievement of these goals.

In summary, complex and adaptive in a complex adaptive system stand for, respectively in ABMs the interaction and the learning of agents.

1.1.3 Heterogeneous expectations

> "One of the things that microeconomics teach you is that individuals are not alike. There is heterogeneity , and probably the most important heterogeneity here is heterogeneity of expectations. If we did not have heterogeneity, there would be no trade. But developing an analytic model with heterogeneous agents is difficult." (Colander et al., 2004, p. 301)[6]

As argued by Hommes (2006), in a market where all agents are rational, there will be no trade at all. For instance, if a trader has information that gives him incentive to sell a stock and other traders in the market do not have this information, the informed traders cannot benefit from his information. In fact, if other trades are rational, they will anticipate that he has private information and therefore they do not buy the stock to him. This fact is strongly in contrast with the high daily trading volume observed

[6]This citation comes from Hommes (2006, p.1)

in every market all around the world (Hommes, 2006). The trading volume observed in practice, reinforces the idea that markets are composed of agents who have radically different expectations. We suppose that agents have heterogeneous expectation about the future of the stock price through this thesis.

1.2 Density of ABMs

This section aims to do a literature review on agent-based models in finance and especially on articles which focus on stock markets. The literature on ABM was very scarce at the end of the nineties, but the literature has grown very fast up to now. Nowadays, ABM is an intensive research area in most of disciplines where these models are useful.

The review will be articulated in three subsections which are the heterogeneity, learning process and interaction as proposed by Chen et al. (2012). These three subsections will describe the three building blocks required in order to construct an agent-based model. It is important to categorize ABMs in term of their density[7] (from simple to complex) in order to compare different ABMs properly.

1.2.1 Heterogeneity

Heterogeneity is an essential component of ABMs because it describes the composition of the markets and its degree of diversification in strategies (beliefs). In ABMs, the heterogeneity has many dimensions, for instance, expectation about the future price, risk aversion, strategy, wealth and many others (Chen et al. 2012). But, in practice it is impossible to deal with multidimensional heterogeneity, but fortunately, models have to be only simplified representation of the reality (Gilbert, 2008). So, most of models deal only with few dimensions of the heterogeneity. Different expectations about the future price is the most common heterogeneity embedded in ABMs (see Brock and Hommes,

[7] The density of an ABM is defined as its degree of complexity in term of heterogeneity, learning process and interactions through its environment. The term "density" is proposed by Chen et al. (2012). It should be noticed that the term density in ABM is not related to the density in statistic.

1998). Different degrees of risk aversion have been taken into account in ABMs, for instance, Chiarella & He (2002), have studied the model of Brock and Hommes (1998) but relax the assumption of homogeneous risk aversion. The different degrees of risk aversion among traders is one of the main reasons why the trading volume is high in stocks market.

We can differentiate simple heterogeneity to complex heterogeneity (LeBaron, 2000). Simple heterogeneity is models with a few-type which means there is only a small number of types of traders modelized in the artificial market. Typically, models with two or three types of agents are considered as few-type models[8]. On the other hand, complex heterogeneity is a model with many types of different traders in the artificial market.

Before starting to describe different types of design, let's focus on the motivation which has inspired designer of ABMs in term of heterogeneity. Experiments have been conducted during the late 1980s and early 1990s and have accumulated a number of empirical evidences about how financial agents forecast in practice. Empirical evidences have been discovered by several ways, as questionnaires, surveys, financial specialists, dealers and so on (Frankel and Froot ,1990; Allen and Taylor ,1990). As mentioned by Chen et al. (2012), there are two interesting results which are generally extracted from empirical data. First of all, data indicated that there are two kinds of expectations which co-exist in markets. The first type of agents is the fundamentalist, they are characterized to be the stabilizing force of the market. At the opposite, there is the chartist (also called trend follower or technical analyst) which is the destabilized force of the markets. Fundamentalist and chartist will be described more accurately afterwards through examples of ABMs modelization. Secondly, the proportion of these two types of agents (market fraction) is not constant over time. This result suggests that agents are able to adapt through time. In other words, agents are able to learn from their past

[8] It is an arbitrary number and it can change across researchers.

performance to improve their general performance. This finding is a strong supportive of the complex adaptive systems that we suppose through this master thesis.

Heterogeneous agent models

The simplest Heterogeneous Agent Model (HAM) is a model with only two types of agents, we can also call it few-type design. The classic HAM is the fundamentalist-chartist model. Fundamentalist forms their expectation about the future stock price according to market fundamental such as dividend, growth, unemployment, etc. (Hommes, 2006). This type of agents can be seen as a rational trader[9] who believes that the price of the stock is determined according to the EMH (Brock and Hommes, 1998). At the opposite, there is the chartist who based their expectation about future stock price and their trading strategies according to historical price. Generally, they try to extrapolate price trend with historical price movement[10].

One of the very first ABM used to modelizing the stock market or exchange rates has been created by Zeeman (1974). The main hypothesis behind this model is that the market is composed of two types of investors : fundamentalist and chartist. This model aims to offer a qualitative description of stylized facts observed in bull and bear markets[11].

Brock and Hommes (1998) have created a very influential ABM where the classic opposition of fundamentalist and chartist are modelized as an adaptive belief system. The main idea is that agents are able to switch from the fundamentalist (chartist) beliefs to the chartist (fundamentalist) beliefs based on the past realized profit of both strategies. The switching mechanism is embedded in the ABM with the logit regres-

[9] Remark : fundamentalists are not rational traders in a sense that they are not able to know all the strategies of all other traders in the market, but rather traders which are able to stabilize the market price around its fundamental value.

[10] They use models such ARMA, ARIMA or linear forecasting rule (Brock and Hommes, 1998).

[11] Bull and bear markets is a metaphorical representation which describe how stocks markets are doing in general (Investopedia). Link : https://www.investopedia.com/insights/digging-deeper-bull-and-bear-markets/

sion as described later. This model will be described more accurately in the section 1 of the theory and methodology chapter. Another two-type design which is closely related to the idea of fundamentalist and chartist is the model created by Barberis, Shleifer and Vishny (1998). This model assumes that investors are affected by two psychological biases when they form their expectations about future cash flows. The two biases are the conservatism (people underweight new information relative to their prior) and representativeness (people think that only a small sample reflects the properties of the entire population, this bias is often known as the "law of small number" (Rabin, 2002)). Investors suppose that there are two distinct regimes which generate earning : a "mean-reverting" regime, where earnings are more mean-reverting than in reality, and a "trending" regime where earning trend more than in reality. The "trending" regime captures the effect of representativeness and the "mean-reverting" regime capture the conservatism. Investors believe that the regime which generates earning changes exogenously over time and their task is to choose the right regime at the right time (Barberis and Thaler, 2003). These models succeed in incorporate cognitive bias into investor belief. The "mean-reverting" and "trending" regime could be seen as respectively fundamentalist and chartist. Many other ABMs use only two types of traders (fundamentalist and chartist), see Hommes, 2006 and Lux 2009 for review.

Another two-type design which was popular at the end of the 90s is the rational vs noise traders. This model is similar from the fundamentalist vs chartist discuss above because neither of the two types of traders is fully rational. In fact, to be fully rational, a trader has to take into account the presence of other traders and knows their strategies (Hommes, 2006). But chartist can be considered as a king of noisy traders. One of the purpose of these models is to address to the Friedman hypothesis : stabilizing investors ("smart"/rational traders) perform better than destabilizing investors (noise trader) which suggests the following question : are noise traders able to survive in the markets ?

DeLong et al. (1990a) has proposed a model with two types of traders, sophisticated traders and noise traders. The market is composed with two assets, a safe asset (paid a fixed dividend each period) and a risky asset which pays an uncertain dividend. Noise traders base their beliefs upon incorrect information from several sources of information, resulting that their beliefs are incorrect. In contrast, sophisticated traders take advantage of the incorrect belief of noise traders. In other words, they make arbitrage. A surprising conclusion of DeLong et al. (1990a) is that in a finite horizon, a constant fraction of noise traders is able to earn higher expected returns than rational agents. In other words, noise traders are able to survive in the market.

Extension of the two-type design to three or four types of agents is easy to set up, we have just to add another type of agents in the system. But there are still little doubts about the fact that strategies of financial agents can be more complex than a simple two-type design (Chen et, 2012). A simple example of three-type design is to consider an extension of the two-type design adaptive belief system (ABS) introduced by Brock and Hommes (1998). They proposed a three belief types where there is fundamentalist vs opposite biases traders or fundamentalist vs trend vs biases traders.

There are ABMs with more than three or four types of agents. Unfortunately the number of possible combinations of agents' type increases exponentially with the number of agents' type in the system. It is one of the reasons which explains that ABM is so complicated in practice. And it is another argument to keep the number of agents' type in the system relatively low.

Many-type design

Generally, many-type design are models with four or more types of traders[12]. There are no consensus on how to build a many-type design, but the easiest approach is to add new types of traders to two or three-type design. N-type design can be viewed

[12] This is subjective and this number can change across researchers.

as a generalization of the few-type design composed with two to four or more types of traders.

Brock, Hommes and Wagener (2005), henceforth BHW05, have developed a theoretical framework to study the evolution of the markets in presence of many types of traders. They have introduced the Large Type Limit (LTL), this notion enables to make a low dimensional approximation of the evolution of a market forms with many types of traders (Hommes, 2006). Diks and van der Weide (2005) have introduced the notion of Continuous Belief Systems (CBS) which is a generalization of the notion of large type limit. The idea behind CBS is to form traders' belief according to a continuous density function. The distribution of belief is a continuous distribution from which the observed beliefs are sampled. According to the definition of a continuous distribution, the number of different beliefs is infinite, so the CBS is an infinite-type design by construction. The infinite N-type design works directly on the distribution while the finite N-type design works on a sample of beliefs. In other words, the finite N-type design can be seen as a sample of size N which is drawn from the continuous density function (Chen et al., 2012). An advantage of working with distribution is that most of distribution can be described with few parameters[13] to represent an "infinite"[14] degree of heterogeneity in markets composed with an infinite number of different types of agents. For more information on notions of Large type Limit and Continuous Belief Systems, Hommes (2006) gives a nice summary of those notions.

HAM vs many-type design

The paradigm behind the HAM is to make clusters in the population and traders who have similar strategies or beliefs fall into the same cluster[15]. But this notion of

[13]For instance, the Normal distribution can be described with only two parameters, the mean and the variance.

[14]A continuous density function has an infinite number of different values, but in practice there is not difference between two types of agents who have almost the same parameters which govern their beliefs. We consider them as the same, but with a residual heterogeneity.

[15]If we have all information about all traders in a market, the belief of each trader type in a few-type design can be obtained by performing a K-means with K equal to the number of types of beliefs we

clusters does not mean that traders who fall into a same class are identical but they are quite similar. The literature on chartist and fundamentalist falls into this idea of cluster. For instance, if two traders are fundamentalists, but they do not have the same belief, it does not mean they are different because if we compare a fundamentalist and a chartist, their belief will be fundamentally different on average. In other words, the difference between the two fundamentalists will be negligible on average.

The advantage of HAMs is those models are usually tractable (numerically at least) as mentioned by Coqueret (2017). Hence, These tractable models are quite easy to solve. Unfortunately, they are always too simple in term of density to stands for complex markets, they underfit the real complexity of markets.

The paradigm of the many-type design is to be closer to the reality in term of heterogeneity. In fact, in reality all traders are different, therefore the number of significant different traders is much more than 2 or 3. In a many-type design, when two traders have tiny difference in their beliefs, they fall into different classes of traders (Chen et al., 2012).

The disadvantage and advantage of many-type design are the opposite of HAMs. The advantage is that this kind of model is much more closer to the reality in term of heterogeneity and the disadvantage is that there are rarely analytical or numerical tractable. Both paradigms will be extensively compared in the sections 2.1.2 and 2.2.2.1 in term of econometric methods used to estimate their parameters and their ability to reproduce stylized facts respectively.

want to modelize. The beliefs of a type of agents will be the average (median) beliefs inside this given cluster.

1.2.2 Learning process and interaction

The distinction between the learning and the interaction of agents in ABMs is not every time clear, it is the reason why both will be described at the same time in this section. In this section, the most popular ABMs in finance will be described in term of their learning rate and interaction. The described ABMs are not exhaustive, but this is a good introduction to understand how to take into account the interaction and the adaptation of agents in a complex adaptive financial system.

Adaptive belief system

One of the most famous ABM is the Adaptive Belief System (ABS) created by Brock and Hommes (1997, 1998). This model will be described in details in the section 1 of the theory and methodology chapter, so the explanation here will be quite brief. The learning and interaction of agents in the ABS are embedded in the switching mechanism. Agents' belief is not fixed which means that they can switch from one type of agents (belief) to another at each iteration according to the result of past realized performances of each strategy. In the simplest case where there are only two types of agents, this choice is binary between the two types of beliefs. Formally, a logit model can be used to estimate the probability that one type of strategy will be better that the other. So, the probability that an agent chooses a strategy is equivalent to the probabilities that this strategy will be more profitable than the other. In other words, the logistic regression takes into account the learning and the interaction of agents.

The idea that agents are able to switch from one type of belief to another are based on papers which have shown that heterogeneity in strategies (beliefs) may lead to complicate dynamics or instability in the price. These complicated dynamics are, for instance, chaotic fluctuation, cycle in financial markets (Day and Huang, 1990; Chiarella, 1992; De Grauwe et al., 1993; Lux, 1995; Sethi, 1996). In these nonlinear models, fluctuation of asset price is caused by an endogenous mechanism relating the market factions of fundamentalist and chartist to the difference between the current and the fundamental price. Fluctuation of price is caused by the fraction of chartist and fundamentalist, which are respectively the destabilized and stabilized forces in the stock market (Brock and Hommes, 1998). Moreover, as recalled by Chen et al. (2012) variation in the market fraction have been considered to be a cause of a large set of stylized facts (Kirman, 1993; Hommes, 2002).

Kirman's ANT Model

The Kirman's ant model (Kirman, 1993) is inspired by experimentation conduct by the two following papers : Pasteels et al. (1987) and Deneubourg et al. (1987). They found that when ants are facing two sources of food, the distribution of ants in both sources are not equal, on average the proportion of remaining food in both sources follows a ratio around 80:20. This behaviour can be explained only if interaction between individuals is taken into account. Moreover, this herding behaviour involved in ants society are also observed in human behaviour and so in financial markets (Wang et al., 2018). The idea of Kirman is to take into account herding mechanism which takes place in financial market when agents form their expectation. Formally, the ANT model is a two-type design model with the classic chartist and fundamentalist. They are able to switch from one type of agents to another by two mechanisms (parameters). The first mechanism is "self-conversion" which enables agents to switch with a given probability without being influence by the system. The second is the herding mechanism which changes the opinion of an agent with a given probability with respect to the belief of another random agent. Alfarano et al., (2008) have studied a continuous-time version of the 'ant process'. They have embedded the herding mechanism into a simple equilibrium asset pricing model and they showed that is possible to derive a closed-form solution.

Lux's Model (Interacting agent hypothesis)

Lux (1995 and 1998) has created a three-type design where there is the classic fundamentalist and chartist but the latter is subdivided into optimist and pessimist.

The idea of Lux follows the Kirman's ant model by adding an herding mechanism into the process of switching. Lux has succeeded to encapsulate the learning and the interaction into a switching mechanism with a transition rate function. The transition rate function has two components : the relative profit and the herding mechanism. Agents switch from one strategy to another based on the relative profit of their strategy to the other. When a chartist's strategy has earned very high returns subsequently, it is very likely that fundamentalist will switch to become chartist. This is similar to the idea of the adaptive belief system developed by Brock and Hommes (1998). The second component is the herding mechanism, non informed traders (optimist and pessimist chartist) are also influenced in their switching by the number of traders in each strategy. If almost all non informed traders are in an optimistic mood, the probability that pessimist agents will switch to an optimistic mood is very high. So, Lux makes the assumption that market sentiment has an influence on the dynamic of financial price.

Many other types of models appear in the literature, we can mention Game Theory (GT), Minority Games (MG), Microscopic Simulation (MS), Prospect-Theory-based model (PT), Threshold Model (TM).

1.2.3 Autonomous-agent designs

In N-type design, agents have to choose between N strategies based on their past performance. This design has several restrictions in a context where strategies have to evolve to survive in competitive systems as financial stock markets. In N-type design, agents have to choose only between all proposed strategies, moreover, these strategies are fixed at the beginning and no new rules are added and unsuccessful rules are not driven out.

A natural way to improve the complexity in term of learning for N-type design is to give more flexibility to agents. Here, flexibility means that agents are able to discover by their own new rules or strategies. It is a much more realistic view of how financial agents behave in real markets.

Genetic algorithms and the SFI-ASM model

A way to improve the degree of autonomy of agents is to let them learn by themselves. Genetic algorithm[16] (GA) offers a proper perspective to modelize system composed of many agents which are able to learn or adapt progressively. GA enables agents to learn from their past experience (strategies) but additionally they could be creative by developing new strategies as well (Chen et al., 2012).

In 1994, five researchers have introduced the concept of GA in a model where adaptive agents can buy and sell stock on a financial market in the Santa Fe Institute (Palmer et al.,1994; Arthur et al., 1996). This model is called the Santa Fe Institute Artificial Stock Market (SFI-ASM). The main difference between the N-type design and the autonomous-agent designs is that agents are not grouped into a fixed number of clusters or represented by a continuous density function, but instead there are all different and their behaviours are customized with a GA. This type of learning gives much more autonomous to agents' behaviour than N-type design. GA enables to increase the complexity in term of heterogeneity and learning in order to achieve the real complexity in financial market[17]. The SFI-ASM model, sees each artificial agent as a machine learning trader who uses regression trees to forecast the stock price in the next period. For more information about the SFI-ASM, let's see Chen et al. (2012) and original papers (Palmer et al.,1994; Arthur et al., 1996).

The main difference between SFI model and N-type design is that the SFI model enables agents to be heterogeneous ex ante while agents in N-type design are heterogeneous only ex post. In the SFI model, agents learn from their own experiences and not

[16]Genetic algorithm has been created by John Holland.

[17]Note : ABMs should not be as complex as real world financial market because ABMs aim only to be a simplified and realistic view of the reality.

from a shared experience as the one used in the ABS which uses a logistic regression (Chen et al., 2012).

There are many other types of ABMs which have been created in the literature, for the interested reader, many surveys are already found (Hommes, 2006; LeBaron, 2006; Lux, 2009; Chen et al., 2012; L Wang et al., 2018)

2 The empirics of agent-based model

This section will discuss about the empiric of agent-based models and its contact with econometric and statistical methods. To introduce this section, let's define what an agent-based computational economics is (ACE). ACE can be defined as the computational study of economic systems viewed as a complex adaptive system (Tesfatsion, 2003). Therefore, agent-based computational finance (ACF) is the application of the ACE in the sub-field of finance. Hence, agent-based computational finance model (ACFM) is a type of model which is useful to understand how agents behave in a controlled environment and how macro-regularities are formed.

This section is organized as follows : firstly a review on how to estimate those models and what to estimate will be described. Then, we will discuss on recent papers which suggest some empirical validations of ACFMs.

2.1 Estimation of ACFMs with econometrics

Calibration of ACFMs parameters can be difficult in most cases. In fact, complexity in agents interactions and the presence of nonlinearities (even if the density of the ABM is very low) make these models complex to estimate. These two properties of ACFMs result in an impossibility of deriving a closed-form solution of the likelihood function and of the moments conditions in of the most cases (Fagiolo et al., 2019).

Hence, other econometric methods have been created to deal with ABMs in general. This section aims to review the econometric literature on how to estimate HAMs and ACFMs with real financial data. First of all, we will discuss what to estimate in ACFMs and secondly, how to estimate these parameters in practice with econometric methods. Finally, a review on estimated ACFMs on financial data will be described[18].

2.1.1 What to estimate in ACFMs ?

So far the main concern of ABMs is first to encapsulate cognitive biases into agents' behaviour, secondly to modelize heterogeneous belief about the future price in order to reproduce price dynamics and finally understanding how stylized facts appear in financial stock market. But, how cognitive biases and heterogeneous belief are taken into account in ABMs ?

For instance, in an adaptive belief system, chartist can be interpreted as a trend follower, they have a cognitive bias which is the representativeness. This bias occurred when people think that only a small sample reflects the properties of the entire population. This bias is sometimes known as the "law of small number" (Rabin, 2002). In ABS, there are predefined behavioral rules for each type of agents. Chartist is characterised by a trend-continuing belief and this type of belief is represented by an extrapolating coefficient α_c in the equation 2. The value of this extrapolating coefficient expresses the degree by which chartist extrapolates past variation in price to future variation (Chen, 2012). The chartist behavioural rule can be written as follows:

$$E_{c,t}[P_{t+1}] = P_t + \alpha_c(P_t - P_{t-1}),\ with\ \alpha_c \geq 0 \tag{2}$$

The equation 2 can be interpreted as follows : an agent who has a chartist belief expects that the future price P_{t+1} will depend linearly on the previous change in price $(P_t - P_{t-1})$ ponderated by an extrapolating coefficient α_c[19]. In other words, behavioural

[18] This section will be focused on adaptive belief system (Brock and Hommes, 1998).
[19] See Chen et al. (2012) for further explanation on ABS.

rules deal with the heterogeneity of behaviour (belief) and taken into account cognitive biases.

Moreover, financial market is assumed to be a complex adaptive system in this master thesis, so learning and interaction have to be tracked by parameters to be taken into account.

For instance, the learning is taken into account in ABS and giving traders the ability to change their belief type from chartist (fundamentalist) to fundamentalist (chartist). In this setting, agents are supposed to be adaptive, because they are able to choose the most profitable strategies to maximise their profit. The switching mechanism is represented by a logit model which gives the probability that an agent chooses the strategy X (f for fundamentalist and c for chartist) given the temporal realized profit $V_{x,t-1}$ of each strategy. In other words, if the fundamentalist strategies have performed well in the past, the likelihood that an agent will choose the fundamentalist strategy is high. The logit model can be written as follows[20]:

$$P(X = f, t) = \frac{exp(\beta V_{f,t-1})}{exp(\beta V_{f,t-1}) + exp(\beta V_{c,t-1})} \tag{3}$$

In this case, the parameter which influences the learning rate of agents is characterised by the intensity of choice (β) in the equation 3. This parameter is called intensity of choice because it measures the sensitivity of agents to the difference between realized profit in both strategies[21]. For instance, if this parameter is close to 0, but the difference between both strategies, in term of past realized profit is quite high, the incentive to switch from the worst to the best profitable strategy is low.

[20] This mechanism was first proposed by Vigfusson (1997).

[21] The logit model can be extended to design with more than two types of traders.

An example of interaction was given by the kirman's ant model already explained in the section 1.2.2. The ant model is governed by two parameters, the self-conversion rate and the herding mechanism (conviction rate). Basically, the model is composed of N agents and each of them must choose a belief about the risky asset price for the next period. They can choose between a chartist and a fundamentalist belief. The self-conversion rate and the conviction rate can affect the belief of agents at each period of time. Each agent is randomly matched with another agent, the first agent is converted to the belief of the second with a probability $(1-\delta)$, this is the conviction rate. Moreover, there is also a probability ε that the agent changes by himself, his belief, this is the self-conversion rate (Hommes, 2006).

It has been found that when the equation 4, holds, the market fraction (fraction of the fundamentalist) does not stay around 0.5 but rather around extreme value, near 0 or 1 (Chen et al., 2012).

$$\varepsilon = \frac{(1-\delta)}{N-1} \tag{4}$$

Behavioral rules, learning and interaction of agents are all represented in ABMs by parameters which capture their intensity. Different values of those parameters will obviously lead to different price dynamics[22], this is the reason why the selection of these values is a crucial point in ABMs to reproduce stylized facts and proper price dynamics.

The next section will review econometric methods which aim to estimate parameters embedded in ABMs to approximate the real-world data generating process (rwDGP) as accurately as possible.

[22]See chaos theory for more information

2.1.2 Methods used to estimate ACFMs

Estimation procedure is methods which are able to tune model parameters in order to find parameters which provide relevant dynamic of the financial time series.

As already said, in many cases, it is impossible to derive a closed-form solution of the likelihood function and of the moments conditions of ACFMs. Direct estimation is therefore impossible in most of the ACFMs because it is difficult if it is not impossible to derive the aggregation equation with analytical method. Hence, other econometric methods have to be created to deal with these types of model.

2.1.2.1 Direct estimation

The content of this section comes from Kukacka and Barunik (2017) and Hurlin (2007).

Direct estimation is only usable for tractable models where the likelihood function and moment condition may be derived. The advantage of direct estimations is they are able to calibrate accurately the parameters of interest in a short time period. Unfortunately, a tiny share of ACFMs can use direct estimation in order to calibrate their parameter vector. In fact, only very simple HAMs are able to derive a closed-form solution of their likelihood function and their moments conditions because those models are usually tractable (numerically at least) (Coqueret, 2017). But the disadvantage of simple HAMs (small density) is there are very simplified models and they are not able to capture all complexity which erupt from real world financial markets.

Although, simple HAMs are largely estimated in the empirical literature because they are able to generate a relatively large number of stylized facts observed in financial markets and are "easy" to estimate with direct estimation. For these two reasons, HAMs are still popular in the empirical literature of ABM.

The most popular method of direct estimation is the Generalized Method of Moments (GMM) developed by Hansen (1982), this method is the general case of many other well-known methods as Ordinary Least Squares (OLS), Maximum Likelihood (ML) or Nonlinear Least Squares (NLS). Many empirical estimations have used GMM to estimate whether an adaptive belief system (Brock and Hommes, 1998), Lux's Model (Lux 1995,1998) or the Kirman's ANT Model (Kirman, 1993).

Quasi Maximum Likelihood (QML) estimator is also a popular method in ACFMs[23].

Direct estimation can be only applied on simple models, hence it is not possible to estimate more complex model (complex density) with those methods. Therefore, a new type of method has to be used in order to estimate these more complex models.

2.1.2.2 Indirect estimation

A solution to get around the issue of infeasible derivation of the aggregation equation in an analytical way (or numerically) is to use Monte Carlo simulations. Econometric methods are then applied on aggregated Monte Carlo simulations. The aim of these methods is they can estimate any type of ACFMs (from simple to complex density). Unfortunately, these methods are computationally expensive when the model complexity and the number of parameters increase.

Methods which calibrate parameters without a closed-form solution are widely shared by other econometric and economic models. Therefore, they have received much interest in the last three decades which have led to the development of new procedures based on simulations (Chen et al., 2012). This procedure is called simulation-based econometric methods[24].

[23] We can mention three other methods which are less popular : Empirical Martingale Simulation (EMS), Interactive Evolutionary Computation (IEC) and Vector AutoRegression (VAR).

[24] Gouriéroux and Monfort (1997) reviews simulation-based econometric methods

Frequentist approach

The majority of calibration exercises in the empirical literature falls into the frequentist paradigm. Certainly the most popular simulation-based method is the Method of Simulated Moments (MSM). The idea behind MSM is to calibrate the parameter vector of an ACFMs in order to the properties of the simulated time series looks like to the one observed. Following Chen et al. (2012), the MSM can be summarized as follows :

The parameter vector is set to given values and the ACFM is run with this given parameter vector in order to generate a simulated time series. Then, we can extract specific moments from the generated time series (simulated moments $\boldsymbol{Y}$) and from the real time series (sample moments $\boldsymbol{X}$). Both moments vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are used to form a distance function (objective function). Hence, the goal of this method is to minimize the distance between the two vectors. The objective function can be written as follows :

$$\hat{\boldsymbol{\theta}}_{MSM} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, \mathcal{L}(\boldsymbol{X}, \boldsymbol{Y}; \boldsymbol{\theta}) \tag{5}$$

The shortcoming of this method is the choice of moments are generally arbitrary, different choice of moments may lead to different parameter vector (Fagiolo et al., 2019). Moreover, a subset of specific moments will only represent a subset of aggregate properties of the real time series, therefore this method risk of not capturing important dynamics (Platt, 2020). Despite of this shortcoming, MSM is very popular in the recent empirical literature[25]. In fact, multiple MSM estimation on each of the three most popular HAMs/ACFMs (ABS[26], Lux's Model[27], the Kirman's ANT model[28]) are attempted in the recent literature.

[25] This popularity is related to its well-understood statistical properties.

[26] Examples of ABS estimated by the MSM : Franke (2009) and Franke and Westerhoff (2012).

[27] Examples of Lux's Model estimated by the MSM : Franke and Westerhoff (2012) Chen and Lux (2016).

[28] Examples of Kirman's ANT model estimated by the MSM : Gilli and Winker (2003) and Winker et al. (2007) .

Indirect Inference (II) introduced by Gourieroux et al. (1993) proposed to use an auxiliary model instead of moments condition. This model is a simple model that is tractable, therefore a direct estimation can be applied on (as maximum likelihood). The objective function is constructed by estimating the auxiliary model on both real and simulated data and comparing optimal parameters obtained. The goal is to minimize the objective function according to the parameter vector of the simulated data. Unfortunately, II suffers from the arbitrary choice of the auxiliary model (Platt, 2020). Only few papers have attempted to estimate a model with II, Bianchi et al. (2007) have, for instance, attempted to use the II in an ABMs context.

MSM and II are both particular cases of the Simulated Minimum Distance (SMD) method. This method involves the construction of an objective function that measures the distance between the real and the simulated time series with some set of functions applied on both time series. The aim of these functions is to summarize time series with only few values. When these functions are moments, we get the MSM. And when these functions are the parameters of the auxiliary model, we get the II (Grazzini and Richiardi, 2015).

Both MSM and II suffer from the arbitrary selection of the moments or the auxiliary model. Some recent works have been conducted in order to solve this issue, rather than focusing on aggregated properties, they have tried to construct metrics that take into account structure and pattern of a given time series (Platt, 2020). Most significant metrics are the Generalized Subtracted L-divergence (GSL-div) and the Markov Information Criterion (MIC) created by Lamperti (2018) and Barde (2017) respectively.

The Simulated Maximum Likelihood (SML) is a simulation-based econometric method used to overcome the fact that ABMs do not have a closed-form solution to their likelihood function and therefore cannot derive the classic MLE. The SML can be considered as an extension of the Methods of Simulated Moments (MSM) because the SML method does not have to choose arbitrary moment (Fagiolo et al., 2019). The SML adapts the classic MLE with an estimation of the density with a non-parametric KDE and Monte Carlo simulation in order to compute the likelihood as suggested by Kristensen and Shin (2012), this method is called NonParametric Simulated Maximum Likelihood Estimator (NPSMLE)[29]. The Simulated Maximum Likelihood will be explained extensively in the section 2 of the theory and methodology chapter because this method will be used in the chapter 4 : empirical estimation of a two-type adaptive belief system.

Kukacka and Barunik (2017) have estimated an ABS with the NPSMLE. They have succeeded to empirically demonstrate that in laboratory conditions, this method is able to calibrate accurately the intensity of choice and the behavioural parameters of a two-type design ABS[30].

Bayesian approach

Most of the simulation-based econometric methods used to estimate ACFMs (or ABMs in general), follow a frequentist approach. The first attempt to estimate ACFMs with Bayesian inference has been introduced by Grazzini, Richiardi and Tsionas (2017). They have attempted to estimate an ABM with one parameter with two versions of the Bayesian inference. Both versions are respectively the NPSMLE and the MSM which are embedded into a Bayesian framework. They have also tried to estimate an ABM with nine parameters with the NPSMLE embedded into a Bayesian framework.

One of the main advantages of the Bayesian approach relatively to the frequentist approach is that it exploits information from the whole distribution and can extract specified moments from the posterior distribution in order to make prediction on the parameters of interest. The second advantage of this method is that a prior distribution (beliefs of the specialist) can be added to the inference. Bayesian inference will be

[29] Note that it is also possible to assume a parametric density. The classic assumption is, for instance the Normal distribution (Grazzini et al., 2017).

[30] The two-type design is composed of the classic fundamentalist and chartist.

explained accurately in the appendix B because this paradigm has a key role in this master thesis. Unfortunately, Bayesian method has a high computational cost when it is applied to ACFMs in order to estimate the likelihood function and the posterior distribution.

Platt (2020) proposed an empirical exercise which compares accuracy of up-to-date indirect estimation (frequentist and Bayesian) on simple models as AR(1), ARMA(2,2) and ABS model. The main conclusion of his empirical exercise is that Bayesian inference seems to be ignored in ABM estimation, but paradoxically it works better than other up-to-date indirect methods in a wide range of cases. He suggests to shift the dominant paradigm from frequentist to Bayesian approach.

2.1.3 Review of adaptive beliefs system

The purpose of this section is to review the empirical academic literature on adaptive belief system created by Brock and Hommes (1998).

The section 1 reviews ABMs which have been developed in the special purpose of "explaining" the economy and especially financial markets. The section 2.1.2 reviews how to estimate ABMs in practice. This section, will review papers which have attempted to estimate ABS.

Kukacka and Barunik (2017) have summarized 47 papers which have attempted to estimate ABMs with financial data. 30 models out of the 47 are based on the ABS. A large majority of the 30 papers has focused their estimation on behavioural rules associated with the different types of traders and on the intensity of choice. These parameters are the most relevant for economic interpretation and for price dynamics. One half of the paper used daily financial data. The other part of studies used low-frequencies data (week to yearly), the main advantage of using low-frequencies data is that fundamental value of the stock can be computed using dividend or other related relevant features. The preferred methods to estimated ABS is the direct estimation with 26 out of the 30 papers and the indirect estimation have only three attempts.

Kukacka and Barunik (2017) have used the Simulated Maximum Likelihood (SML) to estimate an ABS. They focused their estimation on the intensity of choice and behavioural parameters on simulated data. They have succeeded with a good accuracy to retrieve the intensity of choice and behavioural parameters of a two-type ABS. As mentioned by Kukacka and Barunik (2017, p.23) :

> "Nonetheless, the intensity of choice is a crucial and very robust driver of the data-generating process behind switching FABMs and to a large extent determines the behaviour of the system in a very consistent manner"

Hence, it is important to ensure that the method used is able to retrieve a good approximation of the intensity of choice. Kukacka and Barunik (2017) have empirically proved that the NonParametric Simulated Maximum Likelihood Estimator (NPSMLE) is able to retrieve accurately the true value of the intensity of choice and behavioural rules in laboratory conditions.

2.2 Empirical validation of ACFMs

Empirical validation aims to evaluate how inputs or outputs of Monte Carlo simulation looks like, to some well defined real world statistical properties, like behavioural rules or stylized facts (Fagiolo et al., 2019). The section 2.2.1 will discuss about input validation and the section 2.2.2 of output validation.

2.2.1 Input validation

Input validation aims to validate, behavioral rules assumptions embedded in the ACFM, selecting the initial conditions and exploring the possibly high parameter space. The following explanation of the three input validations is provided by Fagiolo et al., (2019).

Selection of behavioral rules

One of the first input validation of an ACFM is to perform laboratory experimentation, this step is useful to the modeler to verify how different types of agents behave in an environment where everything is controlled. For instance, experiments have been conducted to test specific assumptions about the behaviour of agents in ABM (see Hommes, 2011). For instance, "normative understanding" following Tesfatsion (2006) is the laboratory process where researchers try to discover relevant economic designs. As quoted by Tesfatsion (2006, p.9) :

> "ACE researchers pursuing this objective are interested in evaluating whether designs proposed for economic policies, institutions, and processes will result in socially desirable system performance over time."

Selection of initial conditions

ACFM is very sensitive to initial conditions because even the simplest and deterministic ACFM (a deterministic ACFM with a very low density for instance) can display chaotic price dynamic and very tiny variation in the initial condition between two configurations of ACFM can generate extremely different time series. See Brock and Hommes (1997 and 1998) for further explanation.

Exploration of the parameter space

The empirical input validation of the parameters is different from the section 2.1 in a sense that validation does not care about how to estimate ABMs but rather how to explore the parameter space in order to evaluate the impact of very close settings of parameters on the dynamic of the ABM and to perform policy analysis exercise.

2.2.2 Outputs validation

This section is largely inspired by Fagiolo et al. (2019) and Chen et al. (2012). In practice, the true data generating process (DGP) which generates financial time series is unknown, this DGP refer to the real-world DGP (rwDGP)[31]. In most cases, this rwDGP is a very complicated stochastic process[32] which governs the generation of a single instance of a financial time series and its associated set of stylized facts. ACFMs aim to provide an approximation of the rwDGP. The model has to be a simplified version of the rwDGP, which is called the model-DGP (mDGP). Ultimately, main concerns of ACFMs are to provide relevant explanations of the mechanism which generates stylized facts observed in practice and to generate a correct representation of the dynamic of the true time series.

The purpose of empirical output validation is to assess how far the mDGP is a good approximation of the rwDGP. In other words, output validation can be seen as the process which assesses the extent which the output of the simulation is a good representation of the real world observation.

2.2.2.1 What validate ?

As already mentioned, stylized facts are main concern in ACFMs, therefore it is important to validate accurately if ACFMs are able to replicate stylized facts or not. This section will first explain why stylized facts are so important in ACFMs. Afterwards, we will discuss ACFMs which are able to explain some stylized facts. Finally, we will try to establish the relationship between density and the number of stylized facts reproduce by ACFMs. In other words, does the density of an ACFM have a direct influence on its performance ?

Is ACFMs able to explain stylized facts in financial market ?

One of the main motivations to estimated ACFMs in practice is to provide an explanation to stylized facts which can not be explained with traditional finance. In fact, empirical observation, in U.S stock market in the eighties claims that stocks prices

[31] The DGP that we observe in the real world.

[32] The process is complicated in a sense that possibly a lot of parameters affect the rwDGP.

exhibit excess volatility, according to movement in underlying economic fundamentals. Hence, the excess volatility seems difficult to be explained by a rational model. Other empirical observations suggest that the rational agent model is not the good paradigm to explain stylized fact in stock markets[33]. These empirical findings have supported bounded rational and heterogeneous agents model to explain price movement in financial stock markets (Hommes, 2006).

The literature about excess volatility and overreaction of stock price to news suggests that psychological biases and boundedly rational behavior could provide an explanation for stylized facts (Lux, 2009). So, behavioural model for financial market have increased in popularities at the end of the eighties and early nineties. For a review on stylized facts and their relation to ABMs see e.g. (Hommes, 2006; Lux, 2009)[34].

A list of the most important stylized facts is presented below[35] :

- **Martingale property of financial prices**
- **Excess volatility**
- **Volatility clustering**
- **Heavy tails**
- **Bubble et crash**
- **Long memory of return**
- **Absence of autocorrelation in returns**
- **Conditional Heavy Tails**
- **Volatility Volume Correlations**
- **High trading volume**

[33]See Lux (2009) and Hommes (2006) for extensive explanations on the link between stylized facts and rationality.

[34]For a review which are more focus on stylized fact see Cont (2001 and 2007)

[35]Here, important means that it is the most studied stylised facts in the literature.

General performance

Chen et al. (2012) have summarized the result of 50 ACFMs which are able to replicate stylized facts in the literature. In their study, there are 18 two-type designs (36%), 9 three-type designs (18%), 11 N-type designs (22%) and 12 Autonomous-agent designs (24%).

They have noticed that only 12 different stylized facts are replicated among all 50 estimated ACFMs. Moreover, the stylized facts of high-frequency financial time series are left unexplained in all 50 papers. This can be explained because high-frequency data set is rarely used to estimate ACFMs. Another explanation, is that most of ACFMs uses the Walrasian tatonnement or the market-maker scheme to determine the price of the asset.

In the 12 stylized facts of low-frequency financial time series replicated at least one time, there are 4 which have received much attention. They are Fat Tails (FT), Volatility Clustering (VC), Absence of Autocorrelation (AA) and the Long Memory of return (LM) which have been replicated respectively 41, 37, 27 and 20 out of the 50 papers. Chen et al. (2012) explained this intensive attention to FT and VC can be explained to the dominance of GARCH-type models in financial econometrics. LM is closely linked to VC because LM describes the long term dependencies between the returns of a time series, whereas VC is the tendency of large (small) variation in stock price to be likely followed by large (small) variation.

What is the role of the density on performance ?

The density of an ACFM is its degree of complexity in terms of heterogeneity, learning process and interaction. A natural question is : does the density of ACFMs have an impact on the stylized facts replicated ? Answering this question is not devoid of biases because density is composed of three elements (heterogeneity, learning process and interaction) which are interlinked. For instance, complex (simple) heterogeneity often comes with complex (simple) learning process and interaction.

Among the 12 stylized facts replicated in the 50 ACFMs, 8 are replicated by two-type designs, 8 by three-type designs, 11 by N-type designs and 9 by autonomous-agent designs.

First of all, heterogeneity does not seem to have an influence on the number of stylized facts replicated by ACFMs. In fact N-type designs are able to reproduce 11 stylized facts, whereas two and three-type designs are able to explain 9 together. Chen et al. (2012) argue that the two additional stylized facts explained by the N-type designs are outliers because these two results come from the same paper which focused on a specific aspect of the financial agents. Hence, performance of few-type design and many-type design are not significantly different.

Secondly, learning does not have an influence on the performance of ACFMs. In fact, the only additional stylized fact added by autonomous-agent designs is the Long Memory of Volume (VLM) and there is no reason why N-type models can not explain this stylized fact.

To conclude, complexity in terms of heterogeneity and learning in ACFMs seems not to add explaining power to stylized facts. In conclusion few-type design seems enough to explain stylized facts in financial stock market, this is a good news because they are much easier to estimate than many-type design and autonomous-agent designs. Much attention should be brought to the link between the density and the explaining power of ACFMs.

2.2.2.2 How validate ?

More sophisticated statistical techniques have recently been developed for a better discrimination between different ACFMs which reproduces same stylized facts. Mark

(2013) has examined three measures of the similarity between two time series, which are able to discriminate what the "best" models is[36]. These three are Kullback-Leibler information-theoretic, State Similarity Measure and Generalized Hartley Metric. Lamperti (2018) has developed a new information theoretic criterion that measures similarity in terms of dynamic between two time series, called the Generalized Subtracted L-divergence (GSL-div). See Fagiolo et al., (2019) for a review of new validation approach.

[36]These measures are not perfect, hence they are not able to choose the best model, but at least a good model.

Chapter 2

Theory and methodology

This chapter aims to describe all the theory and methodology necessary in order to construct a new version of a Bayesian estimator in a context where the likelihood function has to be estimated with an indirect estimator. Moreover, this master thesis focuses on the comparison between a frequentist and a Bayesian approach in an indirect estimation framework. Hence, in this chapter, all the relevant theories about the frequentist approach used in this thesis will be described.

1 Agent-based computational finance

This section aims to describe the model which will be estimated in the chapter 3 of this thesis. In this section, we will describe the general set up of the Adaptive Belief System (ABS) model created by Brock and Hommes (1998).

1.1 Adaptive belief system

As quoted by Hommes (2006, p.47) :

> "An ABS is in fact a standard discounted value asset pricing model derived from mean-variance maximization, extend to the case of heterogeneous beliefs"

This model assumes that agents are boundedly rational (as described in the section 1.1.1) and select their investment strategies based on historical performance of different strategies. An interesting property on this model is the fact that it can be formulated in term of deviation from a fundamental price. Hence, ABS can be used in term of deviation from the rational expectation benchmark. The deviation can be computed as follows : $y_t = p_t - p_t^*$. Where p_t and p_t^* denote respectively the observed and the fundamental price of the stocks. The fundamental price can be defined with the following formula :

$$p_t^* = \sum_{l=1}^{\infty} \frac{E_t[d_{t+l}]}{(1+r)^l} = \sum_{l=1}^{\infty} \frac{\bar{d}}{(1+r)^l} = \frac{\bar{d}}{r} \tag{6}$$

In a word, where all agents have rational expected, the stock price is solely determined by economic fundamentals therefore by the discounted sum of expected future dividends as described by the equation 6. Hence, the fundamental price depends on the stochastic process of the dividend d_t. An IID stochastic process of the dividend with a constant mean $E_t[d_t] = \bar{d}$ is assumed for simplicity. Hence, the fundamental price follows a stationary process. Where r stands for a constant risk-free interest rate and R is fixed gross rate, R=1+r.

The ABS is widely approached in the literature, therefore only main equations will be described here. Formally, the ABS can be summarized in three nested equations as follows :

$$Ry_t = \sum_{h=1}^{H} n_{h,t-1} f_{h,t} + \varepsilon_t = \sum_{h=1}^{H} n_{h,t-1}(g_h y_{t-1} + b_h) + \varepsilon_t \tag{7}$$

$$n_{h,t-1} = \frac{exp(\beta U_{h,t-1})}{\sum_{h=1}^{H}(exp(\beta U_{h,t-1}))} \tag{8}$$

$$\begin{aligned} U_{h,t-1} &= (y_{t-1} - Ry_{t-2})\frac{f_{h,t-2} - Ry_{t-2}}{\alpha\sigma^2} \\ &= (y_{t-1} - Ry_{t-2})\frac{g_h y_{t-3} + b_h - Ry_{t-2}}{\alpha\sigma^2} \end{aligned} \tag{9}$$

The equation 7 represents how the deviation y_t is computed. Where the forecasting rule f_h stands for the expectation of the agent type $h \in \{1, ..., H\}$ about the deviation of the observed price from its fundamental value. The forecasting rule is supposed to be linear and with only one lag in its simplest form[1], it can be written as follows :

$$f_{h,t} = E_{h,t-1}[y_t] = g_h y_{t-1} + b_h \tag{10}$$

Where g_h and b_h are called the trend and bias of the traders type h.

- If $b_h = 0$:
  1. The traders is called pure trend chaser if $g_h > 0$ (strong pure trend chaser if $g_h > R$)
  2. The traders is called contrarian if $g_h < 0$ (strong contrarian if $g_h < -R$)
- If $g_h = 0$:
  1. The traders is purely upward biased if $b_h > 0$
  2. The traders is purely downward biased if $b_h < 0$
- If $b_h = g_h = 0$, the trader is called fundamentalist

ε_t in the equation 7 is an error term which represents market unpredictable events or uncertainty about economic fundamentals. Generally, ε_t follows a Normal distribution as $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

The equation 8 defines how to compute the market fraction n_h of trader type $h \in \{1, ..., H\}$. n_h is computed with a multinomial logit regression, an interesting property of this method is that the fraction of all types of traders is summed up to 1, $\sum_{h=1}^{H} n_h = 1$. Where $U_{h,t-1}$ and β denoted respectively the profitability measure for the trading strategy $h \in \{1, ..., H\}$ at time t-1 and the intensity of choice that captures how fast traders are willing to switch between strategies based on historic profitability. Formally, a logit model can be used to estimate the probability that a type of strategy will be better (or worst) that the other. Hence, the probability that an agent chooses a strategy

[1] The forecasting rule can have more than one lag. For instance, the forecasting rule can take into account L lagged deviation, $f_{h,t} = f_h(y_{t-1}, ..., y_{t-L})$.

is equivalent to the probabilities that this strategy will be more profitable than the other. In other words, the logistic regression takes into account the learning of agents.

Finally, equation 9 derives a profitability measure $U_{h,t-1}$ based on past strategies performance. Where α and σ^2 denote respectively the risk aversion coefficient and the beliefs about the conditional variance of excess returns. α and σ^2 are supposed constant over time and type of traders.

The ABS have three main parameters which are the intensity of choice β, the trend $\boldsymbol{g} = (g_1, ..., g_H)$ and the bias $\boldsymbol{b} = (b_1, ..., b_H)$ of the traders' type $h \in \{1, ..., H\}$. These three parameters will be called behavioural parameters thought this master thesis. The last parameter of interest is the variance of the error term σ_ε^2 in the equation 7.

2 Simulated Maximum Likelihood

This section aims to introduce the Simulated Maximum Likelihood (SML) method that we will use in the empirical estimation exercise in the chapter 3 of this thesis. The SML is a simulation-based econometric method used to overcome the fact that most ACFMs do not have a closed-form solution to the likelihood function and therefore cannot derive the classic MLE. The SML can be considered as an extension of the Methods of Simulated Moments (MSM) because the SML method does not have to choose arbitrary moment (Fagiolo et al., 2019). The SML adapts the classic MLE with an estimation of the density with a non-parametric KDE and Monte Carlo simulation in order to compute the likelihood as suggested by Kristensen and Shin (2012), this method is called NonParametric Simulated Maximum Likelihood Estimator (NPSMLE). Kukacka and Barunick (2017) proposed a general explanation of the NPSMLE for estimated ABMs. Here, we describe this method for the ABS model described in the section 1.1. These explanations are largely inspired from Kristensen and Shin (2012).

In the ABS model as described in the section 1.1, y_t denoted the deviation of the observed price from its fundamental value. The length of the time series is equal to T, $\{(y_t, x_t)\}_{t=1}^{T}$. Where x_t includes lagged values of the deviation y_t. Moreover, the observed time series $\boldsymbol{y} = \{y_1, ..., y_T\}$ is assumed to be generated by a fully parametric model :

$$y_t = g_t(x_t, \varepsilon_t, \boldsymbol{\theta}),\ t = 1, ..., T \tag{11}$$

Where ε_t dented an IID sequence of the error term with a known distribution F_ε. The vector of unknown parameters $\boldsymbol{\theta}$ includes behavioural parameters β, $\boldsymbol{g} = (g_1, ..., g_H)$ and $\boldsymbol{b} = (b_1, ..., b_H)$ as described in the section 1. Moreover, $\boldsymbol{\theta}$ includes the variance of distribution F_ε of the error term in the equation 7. Hence, the parameter vector can be written as, $\boldsymbol{\theta} = (\beta, \boldsymbol{g}, \boldsymbol{b}, \sigma_\varepsilon)^T$. Finally, x_t contain the last three lagged deviations, $(y_{t-1}, y_{t-2}, y_{t-3})$.

An estimator of the vector of unknown parameters $\boldsymbol{\theta}$ is the maximum of the following conditional log-likelihood :

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \ell(\boldsymbol{\theta}) \tag{12}$$

Where

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T} log\ p_t(y_t | x_t, \boldsymbol{\theta}) \tag{13}$$

Unfortunately, $p_t(y_{t|x_t,\boldsymbol{\theta}})$ does not have a closed-form solution therefore the classic MLE is not feasible (analytically and numerically). In order to solve this issue, a sample of simulated observations can be generated from the parametric model as described in the equation 11. For each period t, N IID error terms $\{\varepsilon_i\}_{i=1}^{N}$ are drawn from their distribution F_ε and N observed deviation $y_{t,i}^{\boldsymbol{\theta}}$ can be generated from the followed parametric model :

$$y_{t,i}^{\boldsymbol{\theta}} = g_t(x_t, \varepsilon_i, \boldsymbol{\theta}), i = 1, ..., N \tag{14}$$

Where the N simulated deviation $\{y_{t,i}^{\boldsymbol{\theta}}\}_{i=1}^{N}$ are generate with the same set of parameters $\boldsymbol{\theta}$.

By construction, the N simulated IID random variable $\{y_{t,i}^{\boldsymbol{\theta}}\}_{i=1}^{N}$, follows the target distribution : $y_{t,i}^{\boldsymbol{\theta}} \sim p(.|x_t, \boldsymbol{\theta})$, i=1,...,N. Hence, they can be used in order to estimate the conditional density $p_t(y_{t|x_t,\boldsymbol{\theta}})$ with a kernel methods as described in the appendix B. The non-parametric density estimator of $p_t(y_{t|x_t,\boldsymbol{\theta}})$ can be written as follows :

$$\hat{p}_t(y_{t|x_t,\boldsymbol{\theta}}) = \frac{1}{N}\sum_{i=1}^{N} K_h(y_t - y_{t,i}^{\boldsymbol{\theta}}) \tag{15}$$

Under regularity constrain on p_t and the kernel K, we obtain:

$$\hat{p}_t(y_{t|x_t,\boldsymbol{\theta}}) = p_t(y_{t|x_t,\boldsymbol{\theta}}) + \mathcal{O}_P(\frac{1}{\sqrt[2]{Nh}}) + \mathcal{O}_P(h^2) \tag{16}$$

When $N \to \infty$, $h \to 0$ because $h \propto N^{-1/5}$. $\sqrt[2]{Nh} \propto \sqrt[2]{NN^{-1/5}} = N^{2/5}$, so $\sqrt[2]{Nh} \to \infty$. Hence, $\frac{1}{\sqrt[2]{Nh}} \to 0$ and $h^2 \to 0$.

Now, we have an estimation of the conditional density $\hat{p}_t(y_{t|x_t,\boldsymbol{\theta}})$, we can construct the simulated MLE as follows :

$$\hat{\boldsymbol{\theta}}_{NPSMLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \hat{\ell}(\boldsymbol{\theta}) \tag{17}$$

Where :

$$\hat{\ell}(\boldsymbol{\theta}) = \sum_{t=1}^{T} log\ \hat{p}_t(y_t|x_t, \boldsymbol{\theta}) \tag{18}$$

Since $\hat{p}_t(y_{t|x_t,\boldsymbol{\theta}}) \to p_t(y_{t|x_t,\boldsymbol{\theta}})$, $\hat{\ell}(\boldsymbol{\theta}) \to \ell(\boldsymbol{\theta})$ when $N \to \infty$ for $T \geq 1$ under regularity conditions. Kristensen and Shin (2012) have shown that the approximation of $\hat{\boldsymbol{\theta}}$ by the NPSMLE inherits the properties of the classic MLE when $T, N \to \infty$ under suitable conditions. It should be noted that the estimator works on non-stationary time series because the density estimator is not affected by the dependence structure of the series. See Kristensen and Shin (2012) to have more information about the properties of the NPSMLE.

Now, optimisation methods can be used in order to find the maxima of the log-likelihood function according to the parameter vector $\boldsymbol{\theta}$. Obviously, the optimal value of the log-likelihood function will be an approximation of the true parameter vector $\boldsymbol{\theta}$. Note that, the same draws of error terms $\{\varepsilon_i\}_{i=1}^{N}$ can be used across different parameter vector $\boldsymbol{\theta}$ and t.

3 Optimization methods

The aim of this section is to describe optimization methods which will be used in the chapter 3 of this thesis in order to approximate the true parameter vector $\boldsymbol{\theta}$. The chapter 3 will use Bayesian and frequentist estimation of the true parameter vector $\boldsymbol{\theta}$, but all optimization methods describe in this section are not suitable for both approaches. The section 3.1, 3.2 and 3.3 will describe respectively how to use a grid search, a genetic algorithm and a Markov Chain Monte Carlo with the NPSMLE.

3.1 Grid search

The idea of the Grid Search (GS) is very simple, it consists of testing many different parameter vectors $\boldsymbol{\theta}$ in order to find an approximation of the true parameter vector $\hat{\boldsymbol{\theta}}$. The GS is a useful method when the parameter space is in low dimension. In fact, the figure 1 depicts a grid search in low dimension (with only three parameters), so the number of parameters set to test stays low in order to find a good approximation of the true parameter vector $\hat{\boldsymbol{\theta}}$. Unfortunately, the number of parameters set to test grows quickly with the dimensionality of the parameter space. To conclude, this method is very suitable when the dimensionality of the parameter space is low, but reaches bad performance when the dimensionality increases.

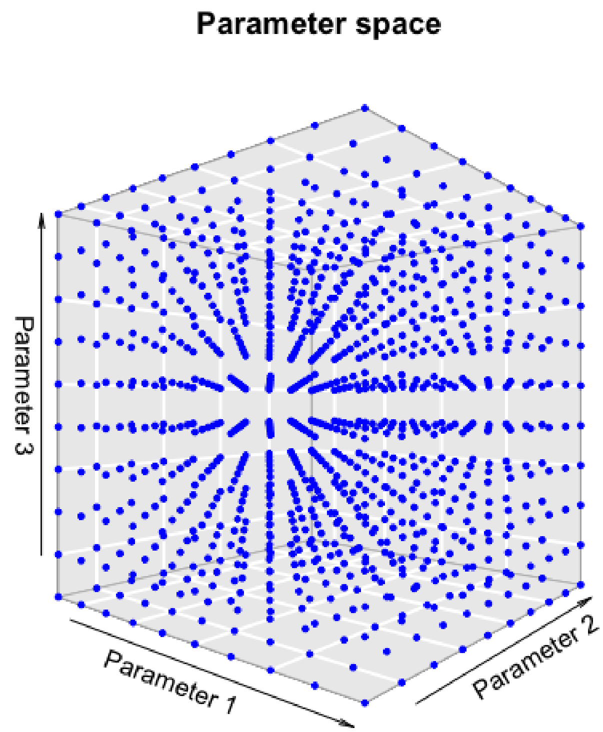

Figure 1: Representation of a grid search with three parameters, for each parameter, 10 equidistant values are tested.

3.2 Genetic algorithm

This part is inspired from the paper of Calvez and Hutzler (2006) and from the blog post on towards data science[2].

Genetic algorithm is an optimization method inspired from the evolution and the natural selection. The figure 2 shows a general schema of genetic algorithms.

[2]link of the blog post on towards data science : https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3

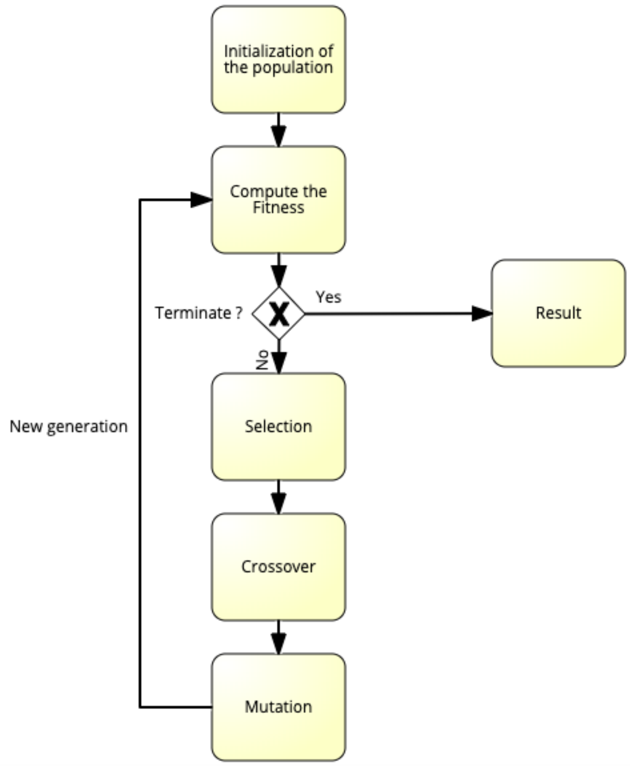

Figure 2: Diagram of a typical genetic algorithm

First of all, the process begins by creating a set of random individuals which forms a population. An individual can be viewed by analogy to the natural selection as a chromosome composed of genes which stands for the parameter vector $\boldsymbol{\theta} = (\theta_1, ...\theta_k)$ of the model. The second step of the algorithms is to compute the fitness score associate with each individual using a fitness function. The fitness score measures how well an individual is good with respect to the optimization problem. Obviously, the fitness score in the case of the NPSMLE is the log-likelihood. Afterwards, a selection of the best individual is made, based on their fitness score. Hence, only these individuals are allowed to pass their genes to the next generation. The crossover phase matches randomly two parents from the population selected to form an offspring based on a mixture of the parents' genes. For instance, each child's gene is picked up randomly from one of both parents. Once the crossover is done, a mutation can occur on the child's chromosome. This phase mutates each gene with some low probabilities. This step adds randomness to the process in order to avoid local convergence. Once, all the new individuals (children) are formed, the next generation is ready to start the process again. Note that the population is set to a fixed size and the number of generations is fixed in advance.

3.3 Markov Chain Monte Carlo

This section is largely based on the two following sources ; Andrieu et al. (2003) and Hartig et al. (2011).

Markov Chain Monte Carlo (MCMC) is a very useful algorithm to solve integration and optimization problem when the parameter space is in high dimension. Both problems are recurrent in many fields like machine learning, statistic, econometric and many others.

This algorithm can by used to solve the integration problem in the Bayes' rule. In fact, the intractable integration problem is central in Bayesian statistics. To obtain the posterior distribution $p(\boldsymbol{\theta}|D)$ given the likelihood $\mathcal{L}(\boldsymbol{\theta})$ and the prior $p(\boldsymbol{\theta})$, the normalization factor of the Bayes' rule needs to be computed in order to obtain a good approximation of the posterior distribution. Where $\boldsymbol{\theta} = (\theta_1, ...\theta_k)$ is a vector of k parameters.

As a reminder the Bayes' rule can be written as follows:

$$p(\boldsymbol{\theta}|D) = \frac{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_A p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}, where A \subseteq \mathbb{R}^k \tag{19}$$

The normalization factor is : $p(D) = \int_A p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$

In the case of the NPSMLE, it is unfeasible to test all set of parameters $\boldsymbol{\theta}$ in order to compute $p(D)$, hence, more sophisticated techniques have to be used to estimate the normalization factor.

But in practice, only the numerator matter and $p(D|\theta)$ is replaced by the likelihood. Hence, we can write the Bayes' rule as follows:

$$p(\boldsymbol{\theta}|D) \propto \mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{20}$$

This integration problem is "solved" by the MCMC algorithms. MCMC is an algorithm which generates sample and explores the parameter space (state space χ) using a Markov chain mechanism[3] (Andrieu et al., 2003). This mechanism tries to concentrate effort of sampling in area associated with high likelihood $\mathcal{L}(\boldsymbol{\theta})$. In other words, MCMC sampling tries to estimate the normalization factor of the Bayes' rule by testing parameter vector $\boldsymbol{\theta}_i = (\theta_1, ...\theta_k)$ as much as possible in zone of high likelihood. A parameter vector which has an associated low value of likelihood, does not have a significant impact on the normalization factor instead of a parameter vector which has a high likelihood.

Let's define some notions and terminology that we will use in the following section. The target distribution is the posterior distribution $p(\boldsymbol{\theta}|D)$ in a Bayesian framework. The proposal distribution is the Probability Density Function (pdf) used to explore the parameter space, throughout this master thesis the proposal distribution will be noted $K(\theta)$ and the target distribution will be noted $p(\boldsymbol{\theta}|D)$ in the Bayesian framework.

Formally, the MCMC algorithm constructs a Markov chain of parameter vector $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$, where each iteration is constructed by proposing a random move conditional to the previous parameters values $\boldsymbol{\theta}_i = (\theta_1, ...\theta_k)$ which lead to a new parameter vector $\boldsymbol{\theta}_{i+1} = (\theta_1, ...\theta_k)$, this move is accepted or rejected conditionally on the ratio of likelihood $\frac{\mathcal{L}(\boldsymbol{\theta}_{i+1})}{\mathcal{L}(\boldsymbol{\theta}_{i+1}}$. As cited by Hartig et al. (2011, p.822) :

[3]The chain is homogeneous, which means the evaluation of the chain only depends on the current state of the chain.

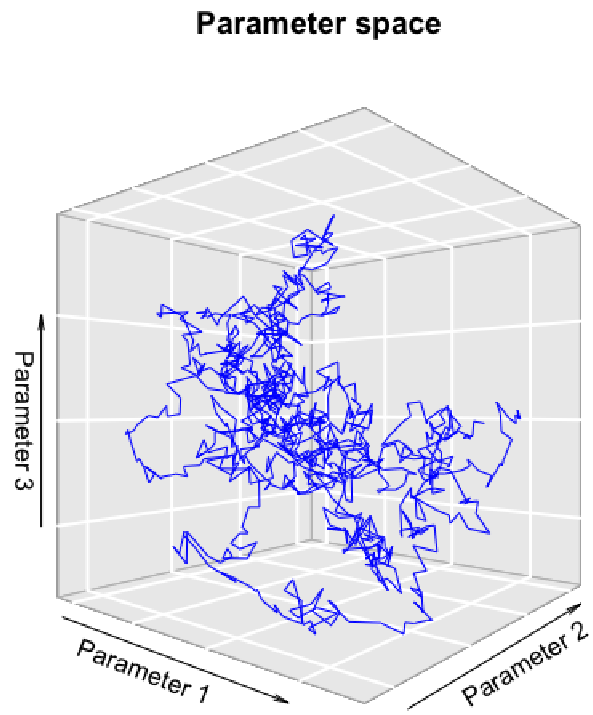

Figure 3: Representation a Markov chain in a parameter space with three dimensions where all random moves for each parameter is drawn from a Normal distribution and all moves are accepted.

> "Given that certain conditions are met (see, e.g Andrieu et al., 2003), the Markov chain of parameters value will eventually converge to the target distribution $\mathcal{L}(\boldsymbol{\theta})$."

For instance, the figure 3 depicts a Markov chain in a parameter space with three dimensions where all moves for each parameter is drawn from a Normal distribution and all moves are accepted.

The advantage of the MCMC is that the time needed to reach an acceptable convergence is shorter than other method because sampling effort is concentrated in zone of high likelihood (Hartig et al., 2011).

3.3.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithms is the most popular MCMC algorithm (Hastings, 1970; Metropolis et al., 1953). A simple example of the MH algorithms is the Metropolis algorithm, this sampling method assumes a symmetric random walk proposal $K(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = K(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{i+1})$. The new candidate $\boldsymbol{\theta}_{i+1}$ is accepted with the following probability :

$$A(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = min\{1, \frac{p(\boldsymbol{\theta}_{i+1}|D)}{p(\boldsymbol{\theta}_i|D)}\} \tag{21}$$

The ratio $\frac{p(\boldsymbol{\theta}_{i+1}|D)}{p(\boldsymbol{\theta}_i|D)}$ is not directly used in the equation 21. In practice the following ratio is used :

$$A(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = min\{1, \frac{\mathcal{L}(\boldsymbol{\theta}_{i+1})p(\boldsymbol{\theta}_{i+1})}{\mathcal{L}(\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}\} \tag{22}$$

Where $p(\boldsymbol{\theta}_i|D) \propto \mathcal{L}(\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)$ and $p(\boldsymbol{\theta}_{i+1}|D) \propto \mathcal{L}(\boldsymbol{\theta}_{i+1})p(\boldsymbol{\theta}_{i+1})$

Interesting properties of the MH algorithms are that the normalization factor of the Bayes' rule is not required and the target distribution needs to be known up to a constant of proportionality. Unfortunately, the choice of the covariance matrix of the proposal distribution is difficult in practice. As empirically demonstrated by Andrieu et al. (2003), the success or the failure of the MH algorithm often depends on the choice of the proposal distribution and its associate covariance matrix. For more information on MCMC, see Andrieu et al. (2003).

3.3.2 The Gibbs sampler

A solution to improve the HM algorithm in term of the choice of the proposal distribution and its associate covariance matrix is to use the Gibbs sampler. The key idea of this sampling algorithm is that, rather than proposing a random move for all parameters at once, a random move is made for each parameter separately, where each move is conditionally to all other dimensions of the parameter space (Resnik et al., 2010). The proposal distribution is therefore the conditional distribution as follows :

$$p(\theta_j|\boldsymbol{\theta}_{-j}^{(t+1),(t)}) = \frac{p(\boldsymbol{\theta}^{(t+1),(t)})}{p(\boldsymbol{\theta}_{-j}^{(t+1),(t)})} \tag{23}$$

Where $\boldsymbol{\theta}^{(t+1),(t)} = (\theta_1^{(t+1)}, ..., \theta_{j-1}^{(t+1)}, \theta_j^{(t+1)}, \theta_{j+1}^{(t)}, ..., \theta_k^{(t)})$
and $\boldsymbol{\theta}_{-j}^{(t+1),(t)} = (\theta_1^{(t+1)}, ..., \theta_{j-1}^{(t+1)}, \theta_{j+1}^{(t)}, ..., \theta_k^{(t)})$

It should be noticed that notation has changed, parameters are indexed by j and the iteration is represented by t. The idea of Gibbs sampling is to sample new value of the parameters θ_j according to the conditional proposal distribution given all other parameters $\boldsymbol{\theta}_{-j}$. Be aware that new value of parameters are used as soon as they are obtained. Unfortunately, in practice the target distributions (posterior distributions) are not known families, hence it is impossible to use this sampling algorithm in this form. As proposed by Grazzini et al., (2017), new candidate can be sampled by a stochastic process as follows :

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \eta_j \varepsilon_j \tag{24}$$

Where ε_j is the random noise which follows a given distribution and η_j is a parameter governing the acceptance rate[4].

It should be noticed that the rest of this section is developed for the special purposed of this master thesis.

In our case, the following stochastic process will be used :

$$\theta_j^{(t+1)} = \theta_j^{(t)} + \varepsilon_j \tag{25}$$

[4]This parameter is adjusted to accept approximately 25% of the candidate tested.

The proposal distribution is the pdf which randomly generate ε_j. Thought this thesis, we assume that the family of this proposal distribution follows a Normal distribution, as $\varepsilon_j \sim \mathcal{N}(\mu,\, \sigma^2)$[5].

In this setting, the new candidate is accepted with probability :

$$A(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) = min\{1, \frac{p(\theta_j^{(t+1)}|D, \boldsymbol{\theta}_{-j}^{(t+1),(t)})}{p(\theta_j^{(t)}|D, \boldsymbol{\theta}_{-j}^{(t+1),(t)})}\} \tag{26}$$

Where :

$$p(\theta_j^{(t+1)}|D, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \propto \mathcal{L}(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \prod_{i=1}^{j} p(\theta_i^{(t+1)}) \prod_{i=j+1}^{k} p(\theta_i^{(t)}) \tag{27}$$

$$p(\theta_j^{(t)}|D, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \propto \mathcal{L}(\theta_j^{(t)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \prod_{i=1}^{j-1} p(\theta_i^{(t+1)}) \prod_{i=j}^{k} p(\theta_i^{(t)}) \tag{28}$$

Parameters are supposed independent. Hence, we can rewrite the acceptance rate as follows :

$$A(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) = min\{1, \frac{\mathcal{L}(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \prod_{i=1}^{j} p(\theta_i^{(t+1)}) \prod_{i=j+1}^{k} p(\theta_i^{(t)})}{\mathcal{L}(\theta_j^{(t)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) \prod_{i=1}^{j-1} p(\theta_i^{(t+1)}) \prod_{i=j}^{k} p(\theta_i^{(t)})}\} \tag{29}$$

The acceptance rate can be simplified as follows :

$$A(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) = min\{1, \frac{\mathcal{L}(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})p(\theta_j^{(t+1)})}{\mathcal{L}(\theta_j^{(t)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})p(\theta_j^{(t)})}\} \tag{30}$$

We can rewrite the acceptance rate in term of log-likelihood as follows :

$$A(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) = min\{1, \frac{exp(\alpha\ell(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})/n)p(\theta_j^{(t+1)})}{exp(\alpha\ell(\theta_j^{(t)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})/n)p(\theta_j^{(t)})}\} \tag{31}$$

[5]The mean μ of the proposal distribution is equal to 0 in practice. There is no reason to bias the random walk.

Simplification :

$$A(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)}) = min\{1, exp(\frac{\alpha}{n}(L_{candidate} - L_{current}))\frac{p(\theta_j^{(t+1)})}{p(\theta_j^{(t)})}\} \quad (32)$$

Where $L_{candidate} = \ell(\theta_j^{(t+1)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})$ and $L_{current} = \ell(\theta_j^{(t)}, \boldsymbol{\theta}_{-j}^{(t+1),(t)})$

Where n is the length of the time series and alpha α is the parameters which controls the acceptance rate, if α is high, the acceptance rate is low. The figure 4 shows the link between the value of the acceptance rate and alpha α. The figure 5 shows the link between $(L_{candidate} - L_{current})$ and the acceptance rate for different values of $\frac{\alpha}{n}$. We can see that when $\frac{\alpha}{n}$ decreases, the exponential curve is more and more flat. It is important to not fix the ratio $\frac{\alpha}{n}$ to 1 because when the length of the time series is high (more than 100 for instance), the difference between the log-likelihood of the candidate and the current parameter vector becomes high too with respect to the exponential curve. The consequence of this difference is that when the candidate has a lower log-likelihood than the current parameter vector, the acceptance rate will be very low on average (close to 0). Hence, decreasing the ratio $\frac{\alpha}{n}$ enables to accept more candidates with lower log-likelihood.

The value of α can be updated as follows :

$$\alpha^{new} = \alpha^{current} - \gamma_\alpha(\bar{A} - A) \quad (33)$$

where γ_α is the learning rate of α, A is the acceptance rate and $\bar{A}$ is the "optimal" acceptance rate.

The main issue with the Gibbs sampler, where the likelihood function is computed with Monte Carlo simulations and the posterior distribution is not known family, is that it is impossible to know optimal values of standard deviations of proposal distribution associated with each parameter in order to accept around 25% of the candidates tested

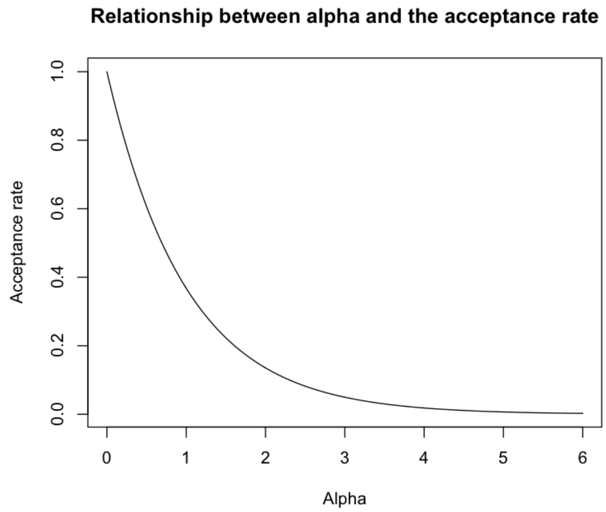

Figure 4: Plot of the link between the acceptance rate and alpha when the difference in the log-likelihood is equal to -1 ($L_{candidate} - L_{current}$).

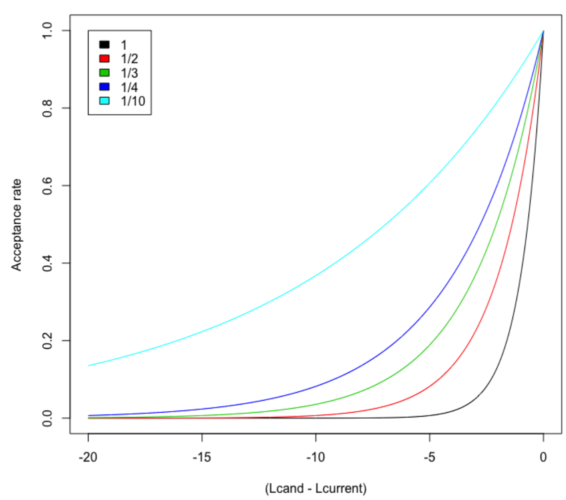

Figure 5: Link between ($L_{candidate} - L_{current}$) and the acceptance rate for different values of $\frac{\alpha}{n}$.

(Grazzini et al., 2017). Hence, the idea is to update standard deviation of the proposal distribution associated with each parameter. The update is characterized with a learning rate $\gamma_{\sigma_{\theta_j}}$, this term updated at each iteration of the Gibbs sampler the value of the standard deviation of the proposal distribution associated with the parameter θ_j. The update formula is as follows :

$$\sigma_{\theta_j}^{new} = \sigma_{\theta_j}^{current} - \gamma_{\sigma_{\theta_j}}(\bar{A} - A) \tag{34}$$

Where A and $\bar{A}$ are respectively the acceptance rate and the "optimal" acceptance rate. In other words, the standard deviations of the proposal distributions of the parameter vector $\boldsymbol{\theta}$ are updated at each iteration of the Gibbs sampler to converge to the acceptance rate $\bar{A}$[6]. Therefore, this algorithm among to the class of adaptive Gibbs samplers.

In the Gibbs sampler, each parameter θ_j is drawn sequentially and the conditional posterior probability $p(\theta_j^{(t+1)}|D, \boldsymbol{\theta}_{-j}^{(t+1),(t)})$ associated with each new candidate is computed given the other parameters $\boldsymbol{\theta}_{-j}^{(t+1),(t)}$. Unfortunately, the time needed in order to converge thought the target distribution is time consuming if we test sequentially all parameters. To solve this issue, multiple moves according to different parameter dimensions can be tested together as a new candidate rather than test each movement separately. All parameter dimensions tested together is called a block. The size of the blocks is crucial to establish because there is a trade-off. If each parameter is tested separately, the time needed to explore the parameter space is very long, especially when the number of dimension is high. This issue is worsened if the correlation between parameters is high. At the opposite, if we test all the parameters at the same time, the probability of accepting the candidate tends to be low (Andrieu et al., 2003). The acceptance rate has to be rewritten as follows if we test several parameters at the same time :

$$A(\boldsymbol{\theta}_S^{(t+1)}, \boldsymbol{\theta}_{-S}^{(t+1),(t)}) = min\{1, \frac{exp(\alpha\ell(\boldsymbol{\theta}_S^{(t+1)}, \boldsymbol{\theta}_{-S}^{(t+1),(t)})/n)\prod_{\boldsymbol{S}_{b=1}}^{B} p(\theta_{\boldsymbol{S}_{b=1}}^{(t+1)})}{exp(\alpha\ell(\boldsymbol{\theta}_S^{(t)}, \boldsymbol{\theta}_{-S}^{(t+1),(t)})/n)\prod_{\boldsymbol{S}_{b=1}}^{B} p(\theta_{\boldsymbol{S}_{b=1}}^{(t)})}\} \tag{35}$$

Where $\boldsymbol{\theta}_S^{(t+1)}$ and $\boldsymbol{\theta}_{-S}^{(t+1),(t)}$ stands for the parameters into and without the block respectively. B is the size of the block and n the length of the time series.

[6]For more information about adaptive MCMC see Andrieu and Thoms (2008).

Our pseudo adaptive Gibbs sampler algorithm can be written as follows :

Algorithm 1: Pseudo Gibbs sampler algorithm

Result: A Markov chain of length T

Initialization of the Markov chain :

- For each parameter of the parameter vectors, sample one value from its associated prior distribution, as $\theta_j^1 \sim p(\theta_j)$. The parameter vector $\boldsymbol{\theta}^1$ denotes the first state of the Markov chain.
- Compute the log-likelihood associated with $\boldsymbol{\theta}^1$ with the NPSMLE method.

t = 2;

while $t < T$ *(T is the target length of the Markov chain)* **do**

- Select randomly B parameters. The indexes of the parameters selected are stocked in the vector $\boldsymbol{S}$.
- For each parameter selected, sample one value from its associated proposal distribution.

for i *in* $\boldsymbol{S}$ **do**

- $\varepsilon_i \sim K(\theta_i)$
- $\theta_i^t = \theta_i^{t-1} + \varepsilon_i$

- $\boldsymbol{\theta}^1$ is updated with the new values of the B selected parameters (θ_i^t) to form the candidate parameter vector $\boldsymbol{\theta}_{candidate}^t$.
- The log-likelihood associated with the candidate parameter vector $\boldsymbol{\theta}_{candidate}^t$ is computed with the NPSMLE method.
- The acceptance rate (equation 35) is computed.
- $r \sim runif(0,1)$
- Update the value of α in the equation 35 as follows : $\alpha^{new} = \alpha^{current} - \gamma_\alpha(\bar{A} - A)$, where γ_α is the learning rate of α and $\bar{A}$ is the "optimal" acceptance rate.
- Update values of the standard deviations of proposal distribution associated with the B selected parameters. The update formula is as follows : $\sigma_{\theta_j}^{new} = \sigma_{\theta_j}^{current} - \gamma_{\sigma_{\theta_j}}(\bar{A} - A)$, where $\gamma_{\sigma_{\theta_j}}$ is the learning rate of the SD associated with the proposal distribution of the parameter θ_j

if $A(\boldsymbol{\theta}_S^{(t+1)}, \boldsymbol{\theta}_{-S}^{(t+1),(t)}) > r$ **then**

- $\boldsymbol{\theta}^t = \boldsymbol{\theta}_{candidate}^t$
- t = t+1

if $t < T$ **then**
  Continue
else
  End

else
  Continue

4 Simulated Bayesian estimator

Our Simulated Bayesian estimator can be summarized in three computational steps as follows :

1. Rather than initializing the Markov chain randomly and throwing out the beginning of the Markov chain that is considered as burn-in, we find a parameter vector $\boldsymbol{\theta}$ which is located in a region of high likelihood (log-likelihood)[7]. Hence, the first step of the estimation is to find the starting state of the Markov chain associated with the Gibbs sampler algorithm[8]. This parameter vector $\boldsymbol{\theta}$ is found with a genetic algorithm as described in the section 3.2. Hence, the starting state of the Markov chain is the optimal value of of the parameter vector as found by the GA, $\hat{\boldsymbol{\theta}}_{GA}$.

2. The second step of the Bayesian estimation is to find the value of alpha which enables the Gibbs sampler to accept around 25% of the candidates given fixed values of standard deviations associated with proposal distributions of the parameter vector[9]. This step takes into account the shape of the likelihood function in area of high likelihood.

   The Gibbs sampler algorithms as described in the section 3.3.2 is used in order to find the "optimal" value of alpha which enables the Gibbs sampler to accept around 25% of the candidates given fixed values of standard deviations associated with proposal distributions of the parameter vector.

   The mean of all updated alpha[10] through the construction of the Markov chain is considered as the "optimal" alpha in order to accept around 25% of the candidates.

[7] For more information about the burn-in period, see the following blog post : https://www.johndcook.com/blog/2016/01/25/mcmc-burn-in/

[8] Both methods have the same goal, finding a region of high likelihood. But starting the Markov chain in an area of high likelihood enables to avoid high roughness of the likelihood function which could disrupt the learning of the standard deviations associated with proposal distributions of the parameter vector and alpha.

[9] Alpha and the standard deviations associated with proposal distributions of the parameter vector can not be estimated at the same time. In fact, standard deviations will grow up indefinitely because their augmentation will be compensated by increases of alpha.

[10] Alpha is updated with the formula as described in the equation 33.

3. The third step is to generate a Markov chain with our Gibbs sampler algorithms which allows to standard deviations associated with proposal distributions of the parameter vector to vary according to their learning rate. The alpha is fixed as its "optimal" values as defined in the step 2.

Once, the three-step estimation is performed, the Markov chain generate at the third step is at least a "good" estimation of the posterior distribution of the parameter vector in term of central tendencies (median or mean)[11]. Hence, the Markov chain generated in the third step is used in order to predict the parameter vector with central tendencies. Unfortunately, this method strongly depends on the initial standard deviations associated with proposal distributions of the parameter vector. Hence, these initial standard deviations have to be chosen consciously. A good choice is for instance using the fraction of the range of the prior distribution[12] as initial standard deviations.

[11]Unfortunately, other metrics as interquartile range or standard deviation have no sense because they depend on the initial standard deviations associated with proposal distributions of the parameter vector as defined in the second step of the estimation.

[12]For the case of an uniform prior.

Chapter 3

Empirical estimation

1 Estimation of a two-type adaptive belief model in laboratory conditions

This section aims to see the extent to which the NPSMLE method and the Bayesian inference are able to estimate the true value of the parameter vector $\boldsymbol{\theta}$ in a laboratory environment.

1.1 Model to estimate

The ABS can be summarized in three nested equations as follows :

$$Ry_t = \sum_{h=1}^{H} n_{h,t-1} f_{h,t} + \varepsilon_t = \sum_{h=1}^{H} n_{h,t-1}(g_h y_{t-1} + b_h) + \varepsilon_t \tag{36}$$

$$n_{h,t-1} = \frac{exp(\beta U_{h,t-1})}{\sum_{h=1}^{H}(exp(\beta U_{h,t-1}))} \tag{37}$$

$$\begin{aligned} U_{h,t-1} &= (y_{t-1} - Ry_{t-2})\frac{f_{h,t-2} - Ry_{t-2}}{\alpha\sigma^2} \\ &= (y_{t-1} - Ry_{t-2})\frac{g_h y_{t-3} + b_h - Ry_{t-2}}{\alpha\sigma^2} \end{aligned} \tag{38}$$

Where y_t is the deviation which is computed as follows : $y_t = p_t - p_t^*$. Where p_t and p_t^* denote respectively the observed and the fundamental price of the stocks. Moreover, r stands for a constant risk-free interest rate and R is fixed gross rate, R=1+r. The forecasting rule f_h stands for the expectation of the agent type $h \in \{1, ..., H\}$ about the deviation of the observed price from its fundamental value. The forecasting rule is supposed to be linear and with only one lag in its simplest form[1], it can be written as follows :

$$f_{h,t} = E_{h,t-1}[y_t] = g_h y_{t-1} + b_h \tag{39}$$

The equation 37 defines how to compute the market fraction n_h of trader type $h \in \{1, ..., H\}$. n_h is computed with a multinomial logit regression, an interesting property of this method is that the fraction of all types of traders is summed up to 1, $\sum_{h=1}^{H} n_h = 1$. Where $U_{h,t-1}$ and β denoted respectively the profitability measure for the trading strategy $h \in \{1, ..., H\}$ at time t-1 and the intensity of choice that captures how fast traders are willing to switch between strategies based on historic profitability.

Finally, equation 38 derives a profitability measure $U_{h,t-1}$ based on past strategies performance. Where α and σ^2 denote respectively the risk aversion coefficient and the beliefs about the conditional variance of excess returns. α and σ^2 are supposed constant over time and type of traders.

Variables which do not impact the dynamic of the time series are fixed. The daily constant gross rate R in the equation 36 is set to 1.0001 which represents a 2.5% annual risk-free interest rate (Kukacka and Barunik, 2017). The risk aversion coefficient α and the beliefs about the conditional variance of excess returns σ^2 in the equation 37 is arbitrary set to 1. Both are only scale factor parameters which do not affect the relative proportion of $U_{h,t}$, therefore, has no influence on the dynamics of the time series. The

[1] The forecasting rule can have more than one lag. For instance, the forecasting rule can take into account L lagged deviation, $f_{h,t} = f_h(y_{t-1}, ..., y_{t-L})$.

length (n) of the time series is fixed to 1100 through this section and the first 100 periods are removed to ensure that the dynamic is properly established (Kukacka and Barunik, 2017).

The noise term in the equation 36 is crucial in the ABS because it denotes markets uncertainty and unpredictable events which arise in the reality. We assume that this noise term is Normally distributed, which is a classic assumption thought the literature. Hence, the Normal distribution is used to generate the N IID errors terms $\{\varepsilon_i\}_{i=1}^{N}$ in the NPSMLE method. As empirically demonstrated by Kukacka and Barunik (2017), a large range of standard deviations σ_ε (from 10^{-8} to 2) associated with the Normal distribution enables to retrieve a good approximation of the true intensity of choice in laboratory conditions. Both ε_t and $\{\varepsilon_i\}_{i=1}^{N}$ are generated by the Normal distribution associated with standard deviations σ_ε (from 10^{-8} to 2). The NPSMLE method is able to retrieve a good approximation of the true parameters if the standard deviations of the noise term are properly parameterized.

The section 1.2 and 1.3 will respectively use the NPSMLE and the Bayesian inference methods to retrieve the true values of the parameter vector $\boldsymbol{\theta}$ in cases of a two-type ABS (4 parameters).

Two-type design adaptive belief system (4 parameters)

The only model to be estimated is a two-type design with two different types of traders in the artificial market. The first type of trader is the classic fundamentalist with fixed behavioural parameters as $b_{fund} = g_{fund} = 0$. The second trader is the chartist with $g_{chartist} \geq 0$ and $b_{chartist}$. Moreover, the intensity of choice is set to a value greater or equal to 0 as $\beta \geq 0$. The last parameter is the standard deviation of the noise term σ_ε, this parameter is called noise intensity. The parameter vector in the two-type design is written as follows : $\boldsymbol{\theta} = (g_{chartist}, b_{chartist}, \beta, \sigma_\varepsilon)^T$. The true value of the parameters are randomly drawn from a Normal distribution as described by the

	Trend chartist	Bias chartist	Beta	Noise term
Mean	1.0	0.10	0.10	1.0
SD	0.1	0.01	0.01	0.1

Table 1: Table shows the values of the mean and the standard deviation of the Normal distribution which generate randomly the true value of the parameter vector.

table 1.

1.2 NonParametric Simulated Maximum Likelihood Estimator

This section will describe the performance of the NPSMLE method to estimate the true values of the parameter vector $\boldsymbol{\theta}$ when the parameter space is explored by a grid search and a genetic algorithm.

Grid search

The table 2 depicts the value of each parameter tested in the grid search, 6 values are tested for each parameter. Hence, the number of combinations of parameter vector tested is equal to $6^4 = 1296$. The standard deviations of the observed time series $\hat{\sigma}_\varepsilon$ is an (upward bias) estimator of the noise intensity in the equation 36. The minimum and maximum of the noise intensity tested in the grid search are derived from the (upward bias) estimator as follows :

$$\begin{aligned} Min(\sigma_\varepsilon) &= \hat{\sigma}_\varepsilon - \frac{\hat{\sigma}_\varepsilon}{3} \\ Max(\sigma_\varepsilon) &= \hat{\sigma}_\varepsilon + \frac{\hat{\sigma}_\varepsilon}{4} \end{aligned} \tag{40}$$

The table 3 shows the true value of the parameter vector and the estimated value by the grid search as described by the table 2.

	Trend chartist	Bias chartist	Beta	Noise intensity
Minimum	0.0	-0.20	0.0	0.789
Maximum	2.5	0.20	1.0	1.480
Interval	0.5	0.08	0.2	0.138

Table 2: Value of the parameters tested in the grid search.

	Trend chartist	Bias chartist	Beta	Noise intensity
True value	0.985	0.104	0.085	1.018
Estimated value	1.000	0.200	0.000	0.927

Table 3: Table which shows the true value of the parameter vector and the value estimated with the grid search.

Genetic algorithm

The 'genalg' package is used to preform the optimization with a genetic algorithm. The table 4 shows the boundaries of the parameter space tested in the genetic algorithm. The boundaries of the noise intensity are computed as the equation 40. The mutation rate and the population size are respectively set to 0.4 and 10. The number of iterations (generations) is fixed to 100. The fitness score is the log-likelihood. The table 5 shows the true value of the parameter vector and the estimated values by the genetic algorithm.

1.3 Bayesian inference

This section will describe the result obtained with the Bayesian approach as described in the section 4 to estimate the true value of the parameter vector $\boldsymbol{\theta}$ when the parameter space is explored by the Gibbs sampler algorithm as described in the section 3.3.2 of the theory and methodology chapter.

The parameters of the Gibbs algorithm are fixed to 2, 10 and 0.25 respectively for

	Trend chartist	Bias chartist	Beta	Noise intensity
Minimum	0.0	-0.2	0	0.904
Maximum	2.5	0.2	1	1.696

Table 4: Boundaries of the parameter space of the genetic algorithm.

	Trend chartist	Bias chartist	Beta	Noise intensity
True value	1.007	0.096	0.100	1.191
Estimated value	0.947	0.131	0.089	1.141

Table 5: This table shows the true value of the parameter vector and the value estimated with the genetic algorithm.

the block size (B), the learning rate of alpha (γ_α) and the "optimal" acceptance rate ($\bar{A}$)[2]. The length of the Markov chain is set to 500. The initial value of α is set to 300.

The table 8 denotes the prior distribution, the proposal distribution and the learning rate of the standard deviations associated with proposal distributions of the parameter vector $\boldsymbol{\theta}$. We make the hypothesis that the parameters are independent for simplicity. The prior distributions of the chartist's trend and bias and the intensity of choice (β) are subjective prior information which can come from others hypothetical studies[3]. The prior of the noise intensity is computed as the equation 40. Standard deviations of the proposal distribution and the learning rate of the standard deviations associated with the proposal distribution are both computed as a fraction of the range of the prior distribution as follows :

$$\sigma_{\theta_j} = \frac{Max(\theta_j) - Min(\theta_j)}{20} \tag{41}$$

[2]The value of the "optimal" acceptance rate $\bar{A}$ is arbitrary. But this value has to be quite low in order to discriminate between good and bad candidate. In fact, there is no theory which can fix $\bar{A}$, therefore this value has to be determined logically.

[3]We are in laboratory conditions, therefore priors are invented.

	Trend chartist	Bias chartist	Beta	Noise term
Initial state	0.947	0.131	0.089	1.141

Table 6: Initial state of the Markov chain

$$\gamma_{\sigma_{\theta_j}} = \frac{Max(\theta_j) - Min(\theta_j)}{1000} \tag{42}$$

The first step of our simulated Bayesian estimator is to find the starting state of the Markov chain associated with the Gibbs sampler algorithm. This parameter vector $\boldsymbol{\theta}$ is found with the same genetic algorithm as described in the section 1.2. It should be noticed that the true parameter vector is the same as the genetic algorithm (section 1.2), therefore, the parameter vector estimated by the GA can be used as the initial state of the Markov chain (see table 5). The table 6 shows the initial parameter vector (state of the Markov chain) as estimated by the GA.

The "optimal value of alpha in order to accept 25% of the candidates as computed by the second step of our Bayesian method is equal to 282.53.

The third step is to generate a Markov chain which allows standard deviations associated with proposal distributions of the parameter vector to change according to their learning rate as defined in the table 8 and with an alpha fixed to 282.53.

Once, the three-step estimation is performed, the Markov chain generated at the third step is at least a "good" estimation of the posterior distribution of the parameter vector in term of central tendency (median or mean)[4]. The figure 6 shows the Markov

[4]Unfortunately, other metrics as interquartile range or standard deviation have no sense because they depend on the initial standard deviations associated with proposal distributions of the parameter

chain for each of the 6 combinations of pair of behavioral parameters. Moreover, the table 16 denotes the descriptive statistic associated with the Markov chain. The figure 7 denotes the bivariate posterior distribution associated with each of the 3 combinations of pair of behavioural parameters. Moreover, the figure 8 shows the univariate posterior distribution associate with each parameter. The following table 7 shows the true and the estimated values of the parameter vector.

	Trend chartist	Bias chartist	Beta	Noise intensity
True value	1.007	0.096	0.100	1.191
Estimated value	0.894	0.145	0.092	1.144

Table 7: Table which shows the true value of the parameter vector and the value estimated with our simulated Bayesian estimator.

We can see that the Gibbs sampler estimated the true value of the parameter vector with a high accuracy. This is encouraging for the estimation on real financial data.

	Trend chartist	Bias chartist	Beta	Noise intensity
Prior distribution (Uniform distribution)				
Minimum	0.0000	-1.000	-1.000	0.9117
Maximum	2.5000	1.000	1.000	1.7094
Proposal distribution (Normal distribution)				
Mean	0.0000	0.000	0.000	0.0000
SD	0.1250	0.100	0.100	0.0399
Learning rate of the SD of the proposal distribution				
Learning rate	0.0025	0.002	0.002	0.0008

Table 8: Prior distribution, proposal distribution and the learning rate used in the Gibbs sampler algorithm

vector as defined in the second step of the estimation.

1.4 Compare NPSMLE and Bayesian inference

This section aims to compare frequentist and Bayesian approaches in term accuracy (bias and variance) of their associated estimator.

The GA and the simulated Bayesian estimator will be used 10 times with different generated time series. All parameters set in the section 1.2 and 1.3 stay the same. The true parameter vector, the bias and variance of the simulated Bayesian estimator and GA are reported in the table 9. We can see that the simulated Bayesian estimator performs better than the GA on average in term of bias and variance. Although, the GA performs better in term of variance for the trend chartist parameters. This bias and variance analysis is poor because the size of the sample is only 10[5], hence the performance of both methods could be due to randomness. Nevertheless, we can say at least that both methods perform well in laboratory conditions. It should be noticed that the variance of the simulated Bayesian estimator is certainly proportional to the length of the Markov chain. Unfortunately, this method is time consuming for classic computer. This method requires cloud computing methods to reach very high level of accuracy. Moreover, this simulated Bayesian estimator will certainly perform very well in high dimension relatively to the GA.

2 Bayesian estimation of real financial data

This section aims to estimate the parameter vector $\boldsymbol{\theta}$ on real financial data in a Bayesian framework.

2.1 Fundamental price approximation

As mentioned in the section 1, the adaptive belief system is formulated in term of deviation from a fundamental price. Hence, the estimation of the fundamental price is

[5]The time needed to perform the simulated Bayesian estimator is around 4 hours

	Trend Chartist	Bias Chartist	Beta	Noise term
True value	1.0070	0.0960	0.1000	1.1910
Bayesian inference				
Mean	1.0100	0.0887	0.0635	1.1425
Median	0.9291	0.0709	0.0878	1.1391
Standard deviations	0.1338	0.0458	0.0682	0.0114
Genetic algorithm				
Mean	0.9855	0.0324	0.1520	1.1557
Median	0.9590	0.0428	0.1204	1.1634
Standard deviations	0.0846	0.1042	0.1176	0.0289

Table 9: True value of the parameter vector and the mean, median and SD of the Gibbs sampler and GA estimator.

a crucial step in order to compute properly the deviation y_t of the observed price p_t form its fundamental price p_t^*.

The centred moving average will be used through this master thesis as approximation of the fundamental price as motivated by Kukacka and Barunik (2017). As stated by Kukacka and Barunik (2017, p.34) :

> "The centred MA is therefore suggested to reduce the delay in information flow. Moreover, the centred MA incorporates a convenient property—that the price by definition converges to it—which is precisely the type of feature one would expect from the fundamental value."

Unfortunately, the centred MA is only a proxy for the true fundamental price which is kept unknown. Hence, it is important to be aware that this poor approximation may generate biases in the empirical estimation.

We use the centred MA with a window length of 61 days to approximate the fundamental price. Kukacka and Barunik (2017) have made robustness check with various ranges of window length (from one month to two years) and obtained comparable result.

2.2 Data set

The real financial data estimated in this section will be based on the S&P 500 index. All the data are retrieved from the Yahoo Finance website. We use the daily close price from 12/01/2016 to 12/31/2019 to create our data set in normal economic conditions. The length of the time series is 1000 as the laboratory estimation. We also use the daily close price from 01/03/2007 to 12/31/2010 to create our data set in time of financial crisis. The length of the time series is 1008. Finally, we have created a data set where the volatility of the deviation of the S&P index from its fundamental value is high. This data set used the daily close price from 06/02/2008 to 03/31/2009, the length of the time series is 210. The deviation y_t of the S&P 500 index are computed as described in the above section. Moreover, the fundamental price is computed with the centred MA61.

The table 10 shows the descriptive statistic associated with the deviation of the S&P 500 index for the three data sets.

	Mean	Median	SD	Min	Max	Skewness	Kurtosis
Normal conditions	-1.68	2.05	43.69	-280.99	149.00	-0.81	3.43
Financial crisis	-0.25	3.39	33.21	-145.17	112.99	-0.65	1.53
High volatility of the deviation	-5.01	0.96	49.13	-145.17	112.99	-0.50	0.19

Table 10: Descriptive statistic associated with the deviation of the S&P 500 index for the three data sets.

The figures 9,10 and 11 show each two plots; the first shows the actual index and the centred MA61 fundamental index and the second shows the deviation y_t of the index for respectively the data set in normal economic conditions, in financial crisis and in high volatility of the deviation.

2.3 Estimation of a two-type design ABS in normal economic conditions

This section will estimate the two-type design ABS as described in the section 1.1 on the data set in normal economic conditions with our simulated Bayesian estimator. A Markov chain will be constructed with the Gibbs sampler algorithm as described in the section 3.3.2 of the theory and methodology chapter.

The parameters of the Gibbs algorithm are fixed to 2, 5 and 0.25 respectively for the block size (B), the learning rate (γ_α) and the "optimal" acceptance rate ($\bar{A}$). The length of the Markov chain is set to 2000. The initial value of α is set to 60.

The table 11 denotes the prior distribution, the proposal distribution and the learning rate of the standard deviations associated with proposal distributions of the parameter vector $\boldsymbol{\theta}$. The prior distribution of the noise intensity is computed with the equation 40[6]. Standard deviations of proposal distributions and their associated learning rates are both computed as a fraction of the range of their prior distribution as described in equations 41 and 42.

The first step of our simulated Bayesian estimator is to find the starting state of the Markov chain associated with the Gibbs sampler algorithm. This parameter vector $\boldsymbol{\theta}$ is found with the same genetic algorithm as described in the section 1.2 except that the boundaries of the parameters are not the same. The table 12 shows the boundaries

[6]Its should be noticed that the minimum is set to 0 in order to not introduce bias in the estimation. In fact, there are exogenous factors such news, which can strongly affect the dynamic of time series and therefore their volatility.

	Trend chartist	Bias chartist	Beta	Noise intensity
Prior distribution (Uniform distribution)				
Minimum	0.0000	-2.000	-1.000	0.0000
Maximum	2.5000	2.000	1.000	54.6107
Proposal distribution (Normal distribution)				
Mean	0.0000	0.000	0.000	0.0000
SD	0.1250	0.200	0.100	2.7305
Learning rate of the SD of the proposal distribution				
Learning rate	0.0025	0.004	0.002	0.0546

Table 11: Prior distribution, proposal distribution and the learning rate used in the Gibbs sampler algorithm

	Trend chartist	Bias chartist	Beta	Noise intensity
Minimum	0.0	-2	-1	0.000
Maximum	2.5	2	1	54.611

Table 12: Boundaries of the parameter space of the genetic algorithm.

	Trend chartist	Bias chartist	Beta	Noise term
Initial state	1.69	0.07	0	18.33

Table 13: Initial state of the Markov chain

of the parameter space tested in the genetic algorithm. The table 13 shows the initial parameter vector (state of the Markov chain) as estimated by the GA, $\hat{\boldsymbol{\theta}}_{GA}$.

The "optimal value of alpha in order to accept 25% of the candidates as computed by the second step of our Bayesian method is equal to 53.2.

The third step is to generate a Markov chain which allows to standard deviations associated with proposal distributions of the parameter vector, to change according to their learning rate as defined in the table 11 and with an alpha fixed to 53.2.

The Markov chain generated in the third step is used to derive all the following tables and figures. The figure 12 shows the Markov chain for each of the 6 combinations of pair of behavioural parameters. Moreover, the table 17 denotes the descriptive statistic associated with the Markov chain. The figure 13 denoted the bivariate posterior distribution associated with each of the 3 combinations of pair of behavioural parameters. Moreover, the figure 14 shows the univariate posterior distribution associates to each parameter of the parameter vector.

2.4 Estimation of a two-type design ABS in time of financial crisis

This section will estimate the two-type design ABS as described in the section 1.1 on the data sets in time of financial crisis and in time of high volatility of the deviation with our simulated Bayesian estimator. All parameters of the Gibbs sampler described in the above section (2.3) stay the same because we do not have any more information available.

The first step of our simulated Bayesian estimator is to find the starting state of the Markov chain associated with the Gibbs sampler algorithm. This parameter vector $\boldsymbol{\theta}$ is found with the same genetic algorithm as described in the section 1.2 except that the boundaries of the parameters are not the same. The table 12 shows the boundaries of the parameter space tested in the genetic algorithm. The table 14 shows the initial parameter vector (state of the Markov chain) as estimated by the GA, $\hat{\boldsymbol{\theta}}_{GA}$.

The "optimal value of alpha in order to accept 25% of the candidates as computed by the second step of our Bayesian method is equal to 68.77.

	Trend chartist	Bias chartist	Beta	Noise term
Initial state	1.713	-0.299	-0.003	15.091

Table 14: Initial state of the Markov chain

The third step is to generate a Markov chain which allows standard deviations associated with proposal distributions of the parameter vector, to change according to their learning rate as defined in the table 11 and with an alpha fixed to 68.77.

The Markov chain generated in the third step is used to derive all the following tables and figures. The table 18 denotes the descriptive statistic associated with the Markov chain. The figure 16 denotes the bivariate posterior distribution associated with each of the 3 combinations of pair of behavioural parameters. Moreover, the figure 15 shows the univariate posterior distribution associates to each parameter of the parameter vector.

For the sake of simplicity only the result of our simulated Bayesian estimator is reported for the data set where the volatility of the deviation of the S&P index is high. Indeed, all parameters of the the Gibbs sampler stay the same unlike that the uniform distribution of the chartist's bias parameters is extended from -10 to 10 instead of -2 to 2. In fact, we suspect that the behaviour of the chartist could be bias during this period. The figure 17 shows the univariate posterior distribution associates to each parameter of the parameter vector as estimated by our simulated Bayesian estimator.

2.5 Interpretations

The table 15 shows the mean, the median and the SD of the parameters associated with their Markov chain as generated by the two last sections (2.3 and 2.4).

We can see that the chartist's trend parameter is largely statistically significant and greater than 1 which means that the trend following strategies is dominant over the

	Mean	Median	SD
Estimated parameters in normal conditions			
Trend chartist	1.457	1.455	0.151
Bias chartist	-0.097	-0.085	1.070
Beta	0.002	0.001	0.012
Noise intensity	19.843	19.635	2.350
Log-likelihood	-4.417	-4.412	0.031
Estimated parameters in time of financial crisis			
Trend chartist	1.370	1.364	0.148
Bias chartist	-0.188	-0.316	1.083
Beta	0.004	0.002	0.051
Noise intensity	16.754	16.528	1.632
Log-likelihood	-4.298	-4.293	0.029
Estimated parameters during a period where the volatility of the deviation is high			
Trend chartist	1.267	1.269	0.167
Bias chartist	-0.731	-0.363	4.623
Beta	0.013	0.008	0.029
Noise intensity	28.336	28.129	2.704
Log-likelihood	-4.709	-4.705	0.026

Table 15: This table shows the mean, the median and the SD of the estimated parameters and the average log-likelihood of their associated Markov chain.

contrarian strategies. But the chartist's bias parameter and the intensity of choice are both statistically insignificant[7]. In fact, as mentioned by Kukacka and Barunik (2017)

[7]We obtain very similar estimation of parameters than Kukacka and Barunik (2017) with their NPSMLE. It should be noted that Kukacka and Barunik (2017) have made robustness check on the

there is no reason why the trend following strategy could be bias in the long run[8]. The most interesting result is the fact that the intensity of choice is statistically insignificant because it suggest there is no switching in strategies within financial markets. As quoted by Boswijk et al. (2007, p.18):

> "We emphasize, however, that it is a common result in switching-type regression models that the parameter in the transition function is hardly significant due to the small variation of the fraction n_t caused by large changes in the β^*. As suggested by Teräsvirta (1994), this should not be worrying as long as there is significant heterogeneity in the estimated regimes."

The non-significant intensity of choice does not mean there is no switching in strategies but, more there are not occurring systematically (Kukacka and Barunik, 2017). A consequence of the insignificant of the intensity of choice is that the proportions of fundamentalists and chartists stay around 50%[9]. To conclude, the chartist's trend parameter is significant, therefore it proves there is an heterogeneity in strategies in financial markets. There are at least two type of traders in the S&P500 index ; a strong pure trend chaser because $g_{chartist} > R$ and $b_{chartist} = 0$ and a fundamentalist because $g_f = 0$ and $b_f = 0$.

Moreover, the chartist's trend parameter and the noise intensity are different in all three data sets[10]. It suggest that the agents' behaviours are not the same in different economic situations. Indeed, the chartist's trend parameter is almost 0.2 lower on the data set where the volatility of the deviation is high than on the data set in normal

length of the centred MA, the size of the bandwidth associated with the kernel density estimator, the frequencies of the data set and they have obtained similar results. Moreover, they have checked the smoothness condition and the unique maxima of the simulated log-likelihood.

[8] The estimated value of the chartist' bias is equal to -0.731 for the period where the volatility of the deviation is high. But this value is statistically insignificant because it has relatively a tiny impact on the dynamic of the time series relatively to the trend and the noise intensity which are respectively equal to 1.267 and 28.336.

[9] $n_{h,t-1} = \frac{exp(0 \times U_{h,t-1})}{\sum_{h=1}^{H}(exp(0 \times U_{h,t-1}))} = \frac{1}{2} = 0.5$, if there are two types of traders.

[10] It should be noted these estimations are replicated only one time because our simulated Bayesian estimator is very computational time consuming.

conditions. Moreover, the estimated noise intensity on the high volatility data set is significantly higher than on the two others data sets. It means that period of high volatility are not due to endogenous factors, as high intensity of choice or strong pure trend chaser which are the destabilizing force of the market. Hence, this additional volatility can only be explained by exogenous factors which is captured in the adaptive belief system by the noise intensity. Indeed, the noise intensity parameters represents market unpredictable events or uncertainty about economic fundamentals. This suggest that in period of high volatility, the dynamic of the price is deeply impacted by exogenous factors, like economic, financial or company news. Another and more likely scenario is that the two-type ABS that we have estimated is too restrictive to explain the price dynamic accurately. This is the reason why many other HAMs and ACFMs are developed in the academic literature[11]. The density of a two-type ABS is too simple to stands for complex markets as in finance. Hence, models with more complexity in term of heterogeneity, learning and interaction will be able to explain markets dynamic more accurately.

[11] The most famous HAMs and ACFMs are explained in the literature review of this thesis, but there are much more HAMs and ACFMs available in the up to date academic literature.

Conclusion

The dominant paradigm in finance, the traditional framework has received considerable numbers of criticisms from the academic literature due to its number of shortcomings. In fact, unrealistic rationality, puzzles (unable to explain financial stylized facts) and too strong assumptions overall, make this paradigm futureless. This is the reason why behavioural models are appealing because there are much closer to the reality. Unfortunately, to build more realistic models, they have to be more complex, which leads to models without closed-form of their likelihood function. We have followed the approach of Kukacka and Barunik (2017) in order to compute the simulated likelihood function of ACFMs with the NPSMLE. This master thesis embedded the NPSMLE into a Bayesian framework as introduced by Grazzini et al. (2017), but their Bayesian method used a very limited sampling method. This thesis also proposes an innovative way to estimate the posterior distribution of the parameter vector of ACFMs with an adaptive Gibbs sampler which takes into account the shape of the likelihood function when it is computed with the NPSMLE. We provided an empirical comparison in term of bias-variance between our Simulated Bayesian estimator and the NPSMLE to retrieve the true parameter vector of a two-type ABS in laboratory conditions.

Based on all theory and methodology that we have summarized in the chapter 2, we were able to derive an adaptive Gibbs sampler that can be applied on any ABMs where its likelihood function is computed with the NPSMLE. Once, this new adaptive Gibbs sampler was constructed, it could be used to build our simulated Bayesian estimator.

Our simulated Bayesian estimator is a three-step optimization process which is able to generate of Markov chain which stands for the posterior distribution of the parameter vector of ACFMs.

Then, the NPSMLE and our simulated Bayesian estimator can be applied on a two-type ABS in laboratory conditions where the true parameter vector is known. Firstly, the NPSMLE is used with a grid search and a genetic algorithm to explore the parameter space in order to find the maximum likelihood of the parameter vector. We have found that the grid search is too simple to generate a good estimation of the true parameter vector because this method does not explore the parameter space accurately. However the genetic algorithm is able to retrieve with a good accuracy the true value of parameters when it is properly parametrized. Finally, we have found that our simulated Bayesian estimator is able to retrieve the true value of parameters with high accuracy as well. Moreover, our simulated Bayesian estimator outperforms on average the NPSMLE when the parameter space is explored with a genetic algorithm in term of bias and variance. Moreover, the performance of our simulated Bayesian estimator is certainly proportional to the length of the Markov chain generated by the Gibbs sampler.

Once, our method is tested in laboratory conditions, we were able to estimate a two-type ABS on real financial data. We have estimated a two-type ABS on the S&P500 index in normal conditions, during the 2008 financial crisis and in period of high volatility during the 2008 financial crisis. We have seen there are at least two type of traders in the S&P500 index[12] ; a strong pure trend chaser because $g_{chartist} > R$ and $b_{chartist} = 0$ and a fundamentalist because $g_f = 0$ and $b_f = 0$. Moreover, the additional volatility in period of financial crisis with high volatility can not be explained by endogenous factors, as high intensity of choice or strong pure trend chaser which are the destabilizing force of the market. Hence, this additional volatility can only be explained by exogenous

[12] We have the "same" estimated parameters than Kukacka and Barunik (2017).

factors which is captured in the adaptive belief system by the noise intensity. This suggest that in period of high volatility, the dynamic of the price is deeply impacted by exogenous factors. A more likely scenario is that the two-type ABS that we have estimated is too simple to explain stocks price dynamic accurately.

There is one limitation on our empirical estimation chapter, indeed, the computational time needed in order to achieve a good accuracy with our simulated Bayesian method is very high. Therefore, we were only able to estimate a two-type ABS. This method required cloud computing methods to reach a very high level of accuracy or to move to model with more parameters. It should be noted that our adaptive Gibbs sampler is built with the intention to perform well in high dimension, therefore, further empirical estimations have to be made on ACFMs or HAMs with more parameters. Moreover, this method needs further exploration in laboratory conditions in order to assess more precisely their level of accuracy in term of bias and variance. (Robustness checks are necessary in order to assess the extent which autocorrelation, different size of samples and bandwidth affect accuracy of both approaches. Finally, different priors are interesting to test in order to understand the properties of our Bayesian estimator in different contexts.)

This master thesis contributes to fill the gap between Bayesian and frequentist approaches for indirect estimation. More precisely, the main contribution is the creation of an adaptive Gibbs sampler which is able to take into account the shape of the likelihood function when it is computed with simulated methods in order to sample the posterior distribution in a proper way. It seems to us that our simulated Bayesian method outperforms the NPSMLE on average on laboratory conditions.

Bibliography

Alfarano, S., Lux, T., Wagner, F., 2008. Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach. Journal of Economic Dynamics and Control, Applications of statistical physics in economics and finance 32, 101–136. https://doi.org/10.1016/j.jedc.2006.12.014

Allen, H., Taylor, M.P., 1990. Charts, Noise and Fundamentals in the London Foreign Exchange Market. The Economic Journal 100, 49–59. https://doi.org/10.2307/2234183

Amilon, H., 2008. Estimation of an adaptive stock market model with heterogeneous agents. Journal of Empirical Finance 15, 342–362. https://doi.org/10.1016/j.jempfin.2006.06.007

Andrieu, C., de Freitas, N., Doucet, A., Jordan, M.I., 2003. An Introduction to MCMC for Machine Learning. Machine Learning 50, 5–43. https://doi.org/10.1023/A:1020281327116

Andrieu, C., Thoms, J., 2008. A tutorial on adaptive MCMC. Stat Comput 18, 343–373. https://doi.org/10.1007/s11222-008-9110-y

Arthur, W.B., Holland, J.H., LeBaron, B., Palmer, R.G., Tayler, P., 1996. Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. https://doi.org/10.2139/ssrn.2252

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343. https://doi.org/10.1016/S0304-405X(98)00027-0

Barberis, N., Thaler, R., 2003. Chapter 18 A survey of behavioral finance, in: Handbook of the Economics of Finance, Financial Markets and Asset Pricing 1, 1053–1128. https://doi.org/10.1016/S1574-0102(03)01027-6

Barde, S., 2017. A Practical, Accurate, Information Criterion for Nth Order Markov Processes. Comput Econ 50, 281–324. https://doi.org/10.1007/s10614-016-9617-9

Béreau, S., 2014. The Empirics of Agent-Based Models : The Econometrics of Networks. Ghent University.

Bianchi, C., Cirillo, P., Gallegati, M., Vagliasindi, P.A., 2007. Validating and Calibrating Agent-Based Models: A Case Study. Comput Econ 30, 245–264. https://doi.org/10.1007/s10614-007-9097-z

Boswijk, H.P., Hommes, C.H., Manzan, S., 2007. Behavioral heterogeneity in stock prices. Journal of Economic Dynamics and Control, Tenth Workshop on Economic Heterogeneous Interacting Agents 31, 1938–1970. https://doi.org/10.1016/j.jedc.2007.01.001

Bramoullé, Y., Fortin, B., Durlauf, S., Blume, L., 2010. Social Networks: Econometrics. https://doi.org/10.1057/978-1-349-95121-5_2902-1

Brock, W., Hommes, C., 1997. A Rational Route to Randomness. Econometrica 65, 1059–1096. https://doi.org/10.2307/2171879

Brock, W.A., Hommes, C.H., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22, 1235–1274. https://doi.org/10.1016/S0165-1889(98)00011-6

Brock, W.A., Hommes, C.H., Wagener, F.O.O., 2005. Evolutionary dynamics in markets with many trader types. Journal of Mathematical Economics, Special Issue on Evolutionary Finance 41, 7–42. https://doi.org/10.1016/j.jmateco.2004.02.002

Calvez, B., Hutzler, G., 2006. Automatic Tuning of Agent-Based Models Using Genetic Algorithms, in: Sichman, J.S., Antunes, L. (Eds.), Multi-Agent-Based Simulation VI, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 41–57. https://doi.org/10.1007/11734680_4

Chen, S.-H., Chang, C.-L., Du, Y.-R., 2012. Agent-based economic models and econometrics. The Knowledge Engineering Review 27, 187–219. https://doi.org/10.1017/S0269888912000136

Chen, Z., Lux, T., 2018. Estimation of Sentiment Effects in Financial Markets: A Simulated Method of Moments Approach. Comput Econ 52, 711–744. https://doi.org/10.1007/s10614-016-9638-4

Chiarella, C., 1992. The dynamics of speculative behaviour. Annals of Operations Research 37, 101–123. https://doi.org/10.1007/BF02071051

Chiarella, C., He, X.-Z., 2002. Heterogeneous Beliefs, Risk and Learning in a Simple Asset Pricing Model. Computational Economics 19, 95–132. https://doi.org/10.1023/A:1014957310778

Colander, D., Holt, R.P.F., Rosser, J.B., 2004. The Changing Face of Economics: Conversations with Cutting Edge Economists. University of Michigan Press, Ann Arbor.

Coqueret, G., 2017. Empirical properties of a heterogeneous agent model in large dimensions. Journal of Economic Dynamics and Control 77, 180–201. https://doi.org/10.1016/j.jedc.2017.02.003

Cont, R., 2007. Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models, in: Teyssière, G., Kirman, A.P. (Eds.), Long Memory in Economics. Springer, Berlin, Heidelberg, 289–309. https://doi.org/10.1007/978-3-540-34625-8_10

Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1, 223–236. https://doi.org/10.1080/713665670

Day, R.H., Huang, W., 1990. Bulls, bears and market sheep. Journal of Economic Behavior & Organization 14, 299–329. https://doi.org/10.1016/0167-2681(90)90061-H

De Grauwe, P., Dewachter, H., 1993. A chaotic model of the exchange rate: The role of fundamentalists and chartists. Open Economies Review 4, 351–379. https://doi.org/10.1007/BF01011136

De Grauwe, P., Dewachter, H., Embrechts, M., 1993. Exchange Rate Theory: Chaotic Models of Foreign Exchange Markets. Blackwell Pub, Oxford, UK ; Cambridge, Mass., USA.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990a. Noise Trader Risk in Financial Markets. Journal of Political Economy 98, 703–738. https://doi.org/10.1086/261703

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990b. Positive Feedback Investment Strategies and Destabilizing Rational Speculation. The Journal of Finance 45, 379–395. https://doi.org/10.2307/2328662

Deneubourg, J.L., Aron, S., Goss, S., Pasteels, J.M., 1987. Error, communication and learning in ant societies. European Journal of Operational Research, Modelling Complex Systems I 30, 168–172. https://doi.org/10.1016/0377-2217(87)90093-2

Diks, C., van der Weide, R., 2005. Herding, a-synchronous updating and heterogeneity in memory in a CBS. Journal of Economic Dynamics and Control, Bounded rationality, heterogeneity and market dynamics 29, 741–763. https://doi.org/10.1016/j.jedc.2003.12.004

Duffy, J., 2006. Chapter 19 Agent-Based Models and Human Subject Experiments, in: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics 2, 949–1011. https://doi.org/10.1016/S1574-0021(05)02019-8

Duong, T., 2007. ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. Journal of Statistical Software 21, 1–16. https://doi.org/10.18637/jss.v021.i07

Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., Roventini, A., 2019. Validation of Agent-Based Models in Economics and Finance, in: Beisbart, C., Saam, N.J. (Eds.), Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives, Simulation Foundations, Methods and Applications. International Publishing, Cham, 763–787. https://doi.org/10.1007/978-3-319-70766-2_31

Fama, E.F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417. https://doi.org/10.2307/2325486

Föllmer, H., 1974. Random economies with many interacting agents. Journal of Mathematical Economics 1, 51–62. https://doi.org/10.1016/0304-4068(74)90035-4

Franke, R., 2009. Applying the method of simulated moments to estimate a small agent-based asset pricing model. Journal of Empirical Finance 16, 804–815. https://doi.org/10.1016/j.jempfin.2009.06.006

Franke, R., Westerhoff, F., 2012. Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. Journal of Economic Dynamics and Control, Quantifying and Understanding Dysfunctions in Financial Markets 36, 1193–1211. https://doi.org/10.1016/j.jedc.2011.10.004

Frankel, J.A., Froot, K.A., 1990. Chartists, Fundamentalists, and Trading in the Foreign Exchange Market. American Economic Review 80, 181–85.

Gilbert, N., 2008. Agent-Based Models. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America. https://doi.org/10.4135/9781412983259

Gilli, M., Winker, P., 2003. A global optimization heuristic for estimating agent based models. Computational Statistics & Data Analysis, Computational Econometrics 42, 299–312. https://doi.org/10.1016/S0167-9473(02)00214-1

Gilli, M., Winker, P., 2001. Indirect Estimation of the Parameters of Agent Based Models of Financial Markets. Social Science Research Network, Rochester, NY.

Gouriéroux, C., Monfort, A., 1997. Simulation-based Econometric Methods. Oxford University Press, Oxford; New York.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 85–118. https://doi.org/10.1002/jae.3950080507

Grazzini, J., Richiardi, M., 2015. Estimation of ergodic agent-based models by simulated minimum distance. Journal of Economic Dynamics and Control 51, 148–165. https://doi.org/10.1016/j.jedc.2014.10.006

Grazzini, J., Richiardi, M.G., Tsionas, M., 2017. Bayesian estimation of agent-based models. Journal of Economic Dynamics and Control 77, 26–47. https://doi.org/10.1016/j.jedc.2017.01.014

Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50, 1029–1054. https://doi.org/10.2307/1912775

Hastings, W.K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 57, 97–109. https://doi.org/10.2307/2334940

Hurlin, C., 2007. Macro-Econométrie : Méthodes des moments. Université d'Orléans.

Hommes, C., 2011. The heterogeneous expectations hypothesis: Some evidence from the lab. Journal of Economic Dynamics and Control 35, 1–24. https://doi.org/10.1016/j.jedc.2010.10.003

Hommes, C.H., 2006. Chapter 23 Heterogeneous Agent Models in Economics and Finance, in: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics 2, 1109–1186. https://doi.org/10.1016/S1574-0021(05)02023-X

Hommes, C.H., 2002. Modeling the stylized facts in finance through simple nonlinear adaptive systems. Proceedings of the National Academy of Sciences 99, 7221–7228. https://doi.org/10.1073/pnas.082080399

Kirman, A., 1993. Ants, Rationality, and Recruitment. The Quarterly Journal of Economics 108, 137–156. https://doi.org/10.2307/2118498

Kirman, A., 1991. Epidemics of opinion and speculative bubbles in financial markets, in: Money and Financial Markets. Cambridge: Blackwell, 354-368.

Kristensen, D., Shin, Y., 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood. Journal of Econometrics 167, 76–94. https://doi.org/10.1016/j.jeconom.2011.09.042

Kukacka, J., Barunik, J., 2017. Estimation of financial agent-based models with simulated maximum likelihood. Journal of Economic Dynamics and Control 85, 21–45. https://doi.org/10.1016/j.jedc.2017.09.006

Lamperti, F., 2018. Empirical validation of simulated models through the GSL-div: an illustrative application. Journal of Economic Interaction and Coordination 13, 143–171. https://doi.org/10.1007/s11403-017-0206-3

LeBaron, B., 2006. Chapter 24 Agent-based Computational Finance, in: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics 2, 1187–1233. https://doi.org/10.1016/S1574-0021(05)02024-1

LeBaron, B., 2000. Agent-based computational finance: Suggested readings and early research. Journal of Economic Dynamics and Control 24, 679–702. https://doi.org/10.1016/S0165-1889(99)00022-6

Lo, A.W., 2004. The Adaptive Markets Hypothesis. The Journal of Portfolio Management 30, 15–29. https://doi.org/10.3905/jpm.2004.442611

Lux, T., 2009. CHAPTER 3 - Stochastic Behavioral Asset-Pricing Models and the Stylized Facts, in: Hens, T., Schenk-Hoppé, K.R. (Eds.), Handbook of Financial Markets: Dynamics and Evolution, Handbooks in Finance. North-Holland, San Diego, 161–215. https://doi.org/10.1016/B978-012374258-2.50007-5

Lux, T., 1998. The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. Journal of Economic Behavior & Organization 33, 143–165. https://doi.org/10.1016/S0167-2681(97)00088-7

Lux, T., 1995. Herd Behaviour, Bubbles and Crashes. The Economic Journal 105, 881–896. https://doi.org/10.2307/2235156

Lux, T., Zwinkels, R.C.J., 2018. Chapter 8 - Empirical Validation of Agent-Based Models. Handbook of Computational Economics 4, 437–488. https://doi.org/10.1016/bs.hescom.2018.02.003

Macal, C.M., North, M.J., 2010. Tutorial on agent-based modelling and simulation. Journal of Simulation 4, 151–162. https://doi.org/10.1057/jos.2010.3

Marks, R.E., 2013. Validation and model selection: Three similarity measures compared. https://doi.org/10.7564/13-COEC10

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, 1087–1092. https://doi.org/10.1063/1.1699114

Palmer, R.G., Brian Arthur, W., Holland, J.H., LeBaron, B., Tayler, P., 1994. Artificial economic life: a simple model of a stock market. Physica D: Nonlinear Phenomena 75, 264–274. https://doi.org/10.1016/0167-2789(94)90287-9

Pasteels, J.M., Deneubourg, J.L., Goss, S., 1987. Self-organization mechanisms in ant societies. I. Trail recruitment to newly discovered food sources, in: Pasteels, J.M., Deneubourg, J.L. (Eds), From individual to collective behavior in social insects 155–175.

Platt, D., 2020. A comparison of economic agent-based model calibration methods. Journal of Economic Dynamics and Control 113, 103859. https://doi.org/10.1016/j.jedc.2020.103859

Rabin, M., 2002. Inference by Believers in the Law of Small Numbers. The Quarterly Journal of Economics 117, 775–816.

Resnik, P., Hardisty, E., 2010. Gibbs Sampling for the Uninitiated (No. CS-TR-4956). Maryland univ College park inst for advanced computer studies.

Sethi, R., 1996. Endogenous regime switching in speculative markets. Structural Change and Economic Dynamics 7, 99–118. https://doi.org/10.1016/0954-349X(95)00040-T

Shiller, R.J., Fischer, S., Friedman, B.M., 1984. Stock Prices and Social Dynamics. Brookings Papers on Economic Activity 1984, 457–510. https://doi.org/10.2307/2534436

Shleifer, A., 2000. Inefficient Markets: An Introduction to Behavioral Finance. Oxford University Press.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman and Hall, London; New York.

Simon, H.A., 1997. Models of Bounded Rationality, Vol. 3: Empirically Grounded Economic Reason. The MIT Press, Cambridge, Mass.

Simon, H.A., 1957. Models of man; social and rational, Models of man; social and rational. Wiley, Oxford, England.

Simon, H.A., 1955. A Behavioral Model of Rational Choice. The Quarterly Journal of Economics 69, 99–118. https://doi.org/10.2307/1884852

Teräsvirta, T., 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. Journal of the American Statistical Association 89, 208–218. https://doi.org/10.1080/01621459.1994.10476462

Tesfatsion, L., 2006. Chapter 16 Agent-Based Computational Economics: A Constructive Approach to

Economic Theory, in: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics. Elsevier, 831–880. https://doi.org/10.1016/S1574-0021(05)02016-2

Tesfatsion, L., 2003. Agent-based computational economics: modeling economies as complex adaptive systems. Information Sciences 149, 262–268. https://doi.org/10.1016/S0020-0255(02)00280-3

Tseng, K.C., 2006. Behavioral Finance, Bounded Rationality, Neuro-Finance, and Traditional Finance. Investment Management and Financial Innovations 3, 7-18.

Vigfusson, R., 1997. Switching Between Chartists and Fundamentalists: A Markov Regime-Switching Approach. International Journal of Finance & Economics 2, 291–305. https://doi.org/10.2139/ssrn.50562

Wang, L., Ahn, K., Kim, C., Ha, C., 2018. Agent-based models in financial market studies. Journal of Physics: Conference Series 1039. https://doi.org/10.1088/1742-6596/1039/1/012022

Winker, P., Gilli, M., Jeleskovic, V., 2007. An objective function for simulation based inference on exchange rate data. J Econ Interac Coord 2, 125–145. https://doi.org/10.1007/s11403-007-0020-4

Zeeman, E.C., 1974. On the unstable behaviour of stock exchanges. Journal of Mathematical Economics 1, 39–49. https://doi.org/10.1016/0304-4068(74)90034-2

Appendix A : Supplementary tables

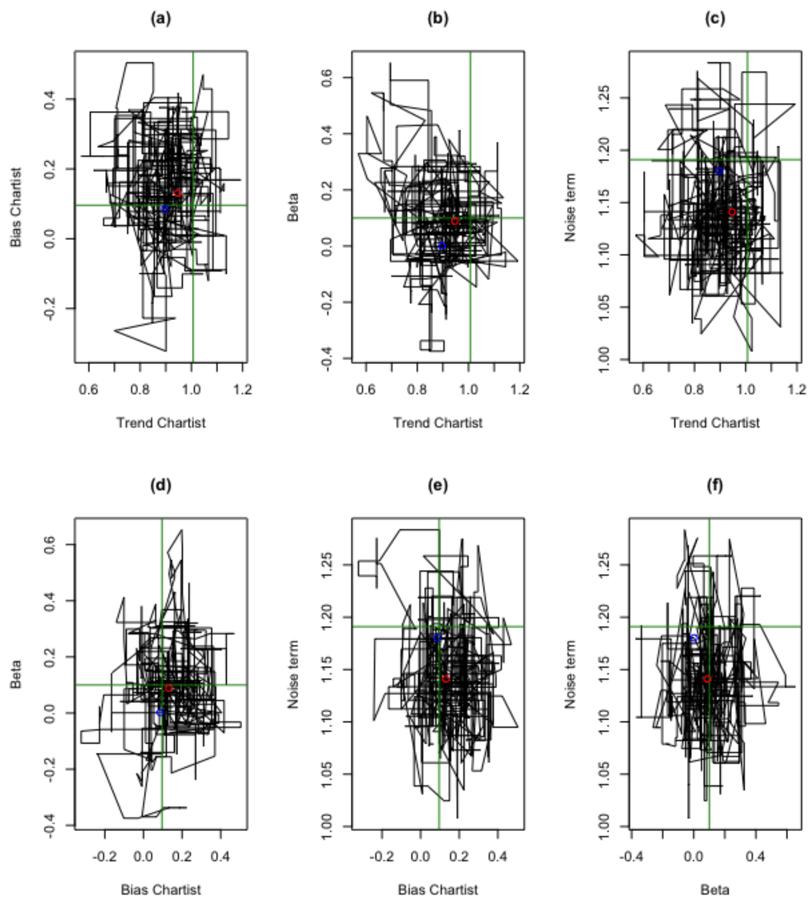

Figure 6: Markov chain generated by the Gibbs sampler algorithm. The figure (a),(b),(c),(d),(e) and (f) depict the Markov chain for the parameters ; trend chartist, bias chartist,the intensity of choice (beta) and the noise intensity. The red and the blue point shows the starting and ending point of the Markov chain respectively. Green lines show the true value of the parameters.

	mean	median	sd	min	max	skew	kurtosis
Trend chartist	0.89	0.90	0.11	0.57	1.19	-0.14	-0.30
Bias chartist	0.14	0.15	0.14	-0.32	0.50	-0.31	0.13
Beta	0.09	0.08	0.16	-0.37	0.65	0.20	0.62
Noise intensity	1.14	1.14	0.05	1.01	1.28	0.39	0.16
Log likelihood	-1.60	-1.60	0.01	-1.63	-1.58	-0.70	0.41
Acceptance rate	0.26	0.07	0.35	0.00	1.00	1.29	0.10
Learning rate of alpha	282.53	282.53	0.00	282.53	282.53	NaN	NaN
Learning rate trend	0.13	0.13	0.02	0.10	0.15	-0.12	-1.44
Learning rate of bias	0.13	0.12	0.01	0.10	0.15	-0.11	-0.60
Learning rate of beta	0.11	0.10	0.01	0.09	0.13	0.16	-0.47
Learning rate of noise intensity	0.05	0.05	0.00	0.04	0.05	-0.02	-0.82

Table 16: Descriptive statistic of the Markov chain.

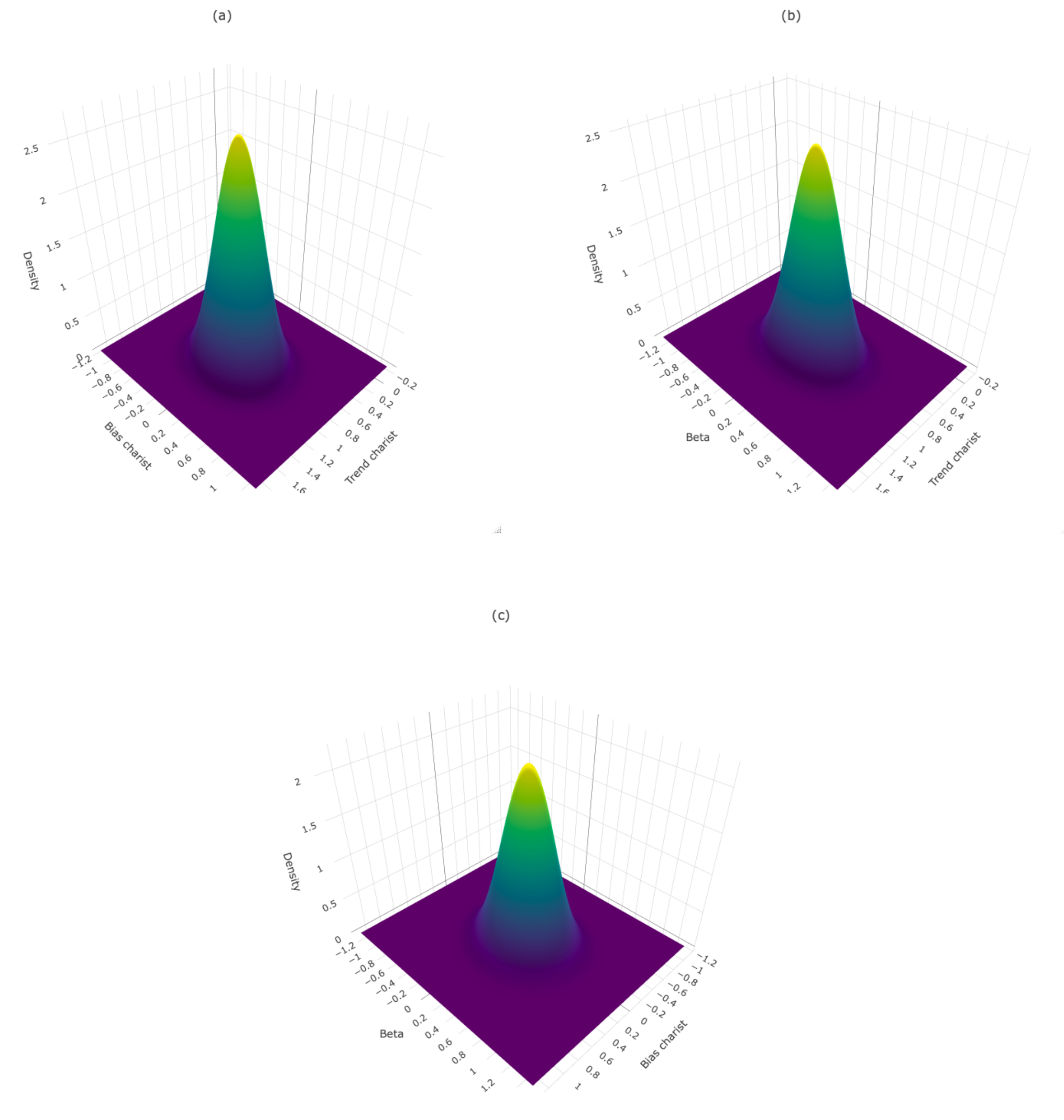

Figure 7: The plots (a),(b) and (c) depict the bivariate KDE posterior distribution for the three combinations of pair of parameters for the parameters ; trend chartist, bias chartist and the intensity of choice (beta).

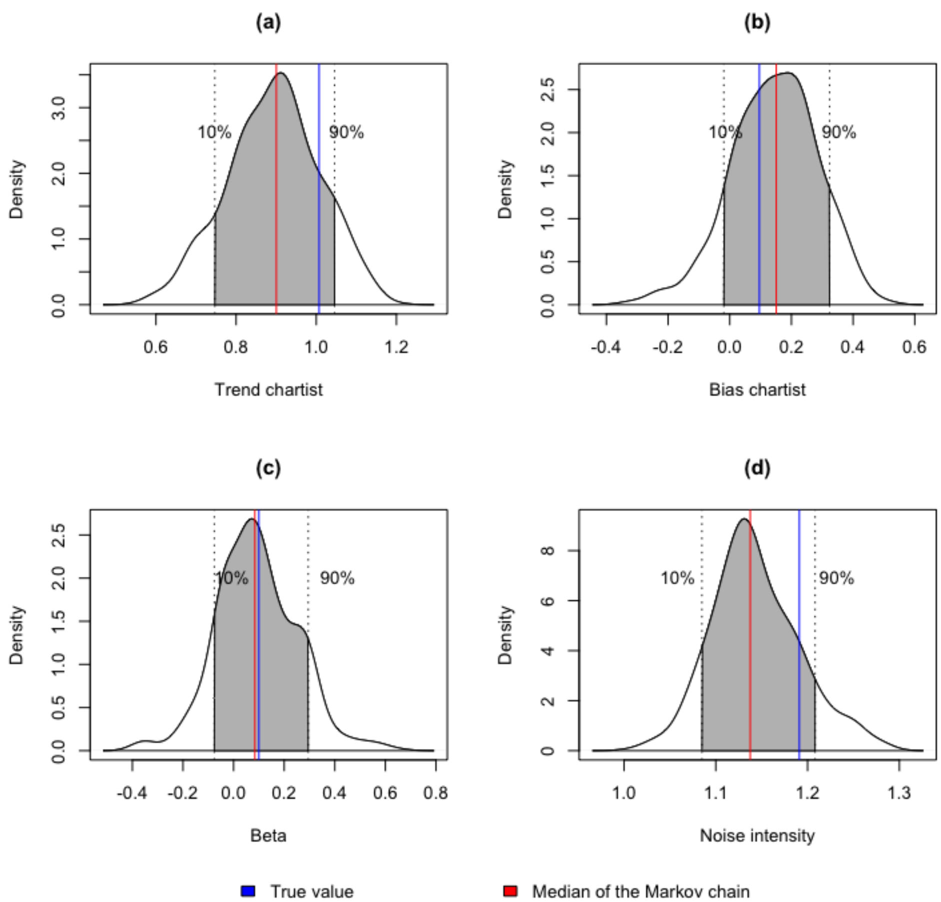

Figure 8: The figures (a),(b),(c) and (d) show respectively the KDE posterior density of the trend and bias parameters of the chartist, the intensity of choice and the noise intensity.

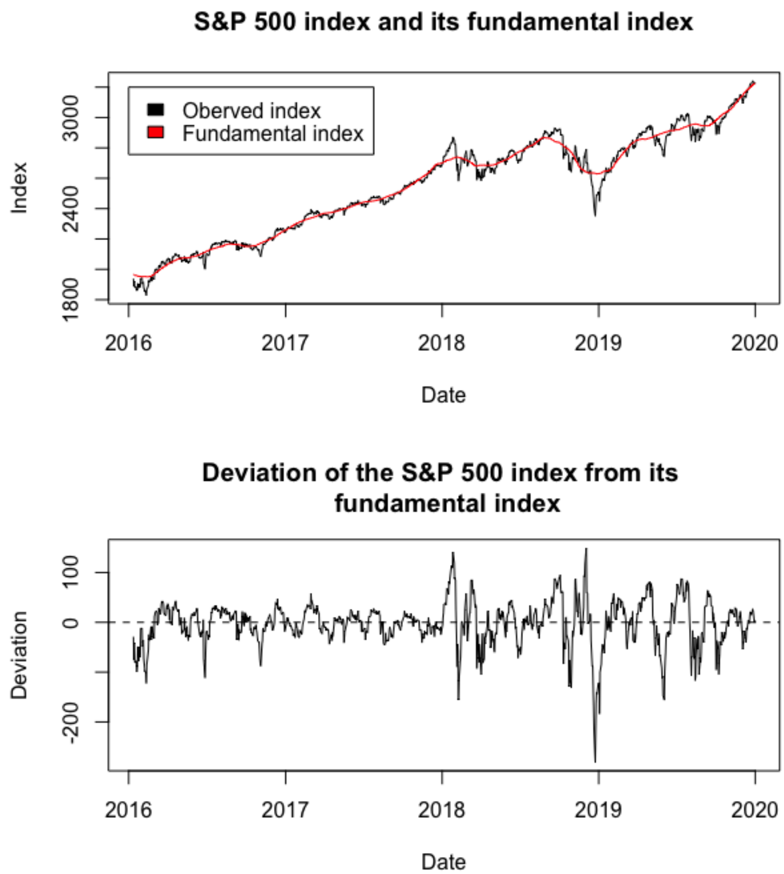

Figure 9: The plot on top shows the actual index and the MA61 fundamental index and the plot on bottom shows the deviation.

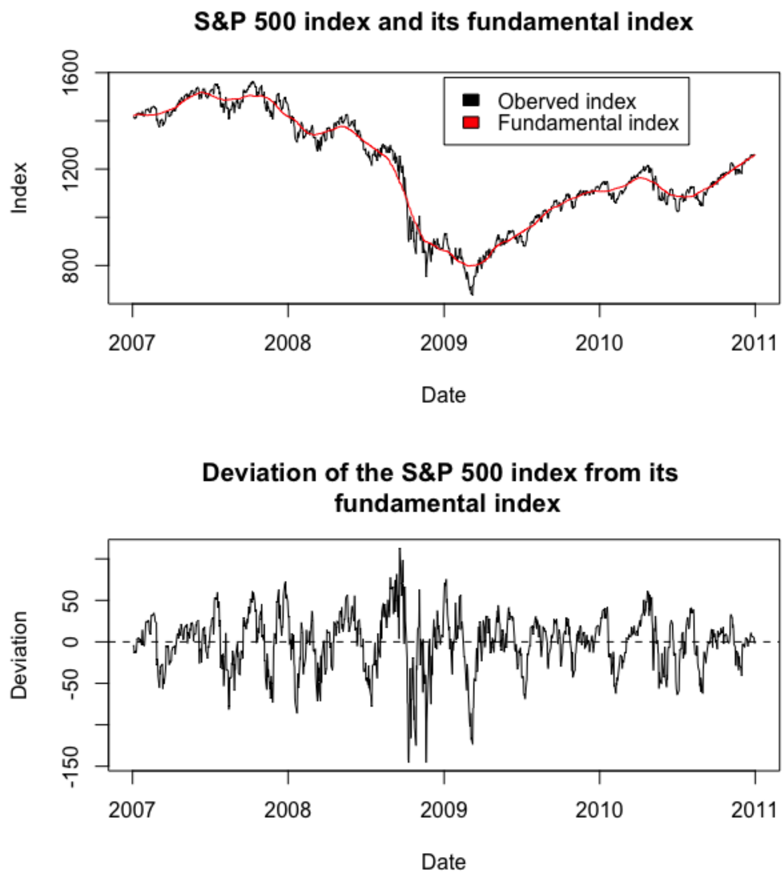

Figure 10: The plot on top shows the actual index and the MA61 fundamental index and the plot on bottom shows the deviation.

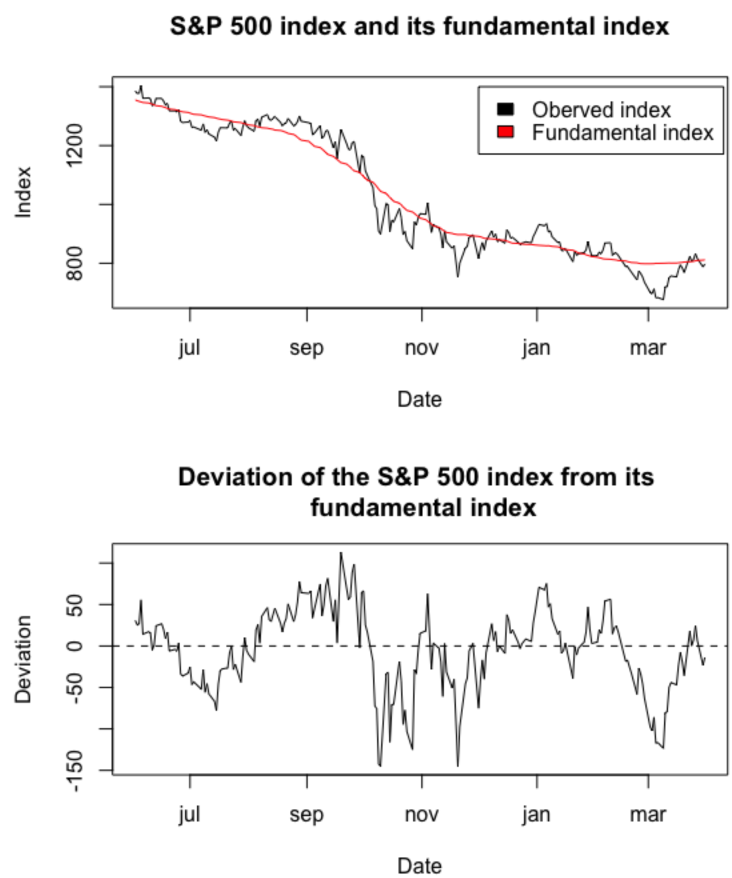

Figure 11: The plot on top shows the actual index and the MA61 fundamental index and the plot on bottom shows the deviation.

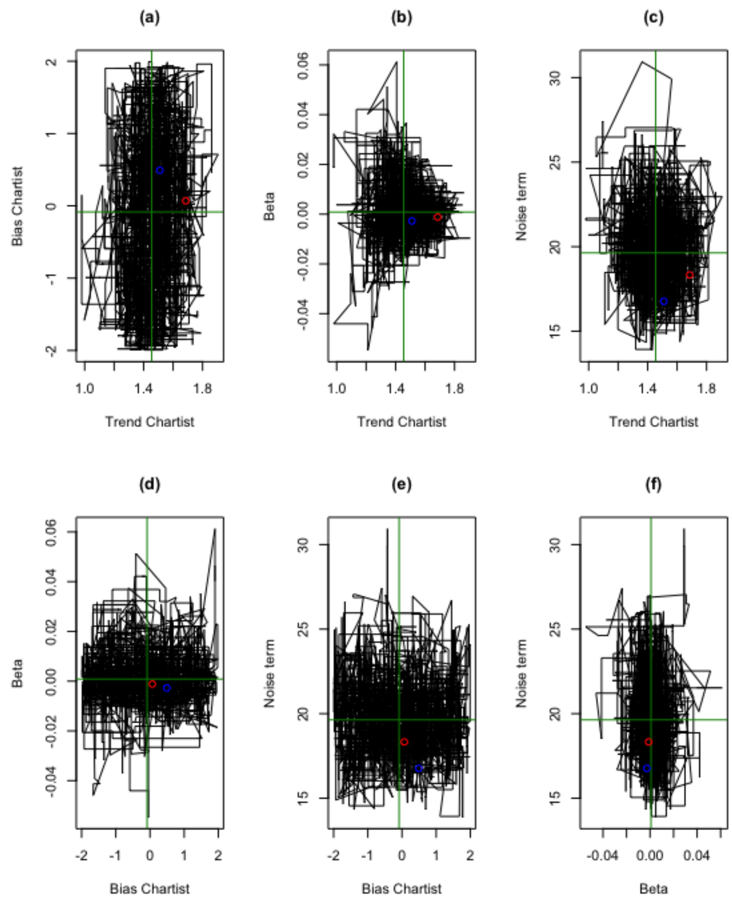

Figure 12: Markov chain generated by the Gibbs sampler algorithm. The figure (a),(b),(c),(d),(e) and (f) depict the Markov chain for the parameters ; trend chartist, bias chartist, the intensity of choice (beta) and the noise intensity. The red and the blue point show the starting and ending point of the Markov chain respectively. Green lines show the median of the Markov chain associated with each parameter.

	mean	median	sd	min	max	skew	kurtosis
Trend chartist	1.457	1.455	0.151	0.980	1.905	-0.055	-0.102
Bias chartist	-0.097	-0.085	1.070	-1.997	1.991	0.032	-1.069
Beta	0.002	0.001	0.012	-0.055	0.061	0.200	2.046
Noise intensity	19.843	19.635	2.350	13.874	30.931	0.483	0.390
Log-likelihood	-4.452	-4.447	0.031	-4.609	-4.365	-0.852	1.355
Acceptance rate	0.268	0.073	0.358	0.000	1.000	1.190	-0.156
alpha	53.201	53.201	0.000	53.201	53.201	NaN	NaN
SD of the proposal distribution of trend chartist	0.210	0.215	0.033	0.124	0.259	-0.609	-0.721
SD of the proposal distribution of bias chartist	0.673	0.722	0.258	0.199	1.094	-0.100	-1.245
SD of the proposal distribution of beta	0.022	0.019	0.014	0.000	0.100	3.120	11.780
SD of the proposal distribution of noise intensity	3.812	3.917	0.561	2.567	4.905	-0.498	-0.673

Table 17: Descriptive statistic associated with the Markov chain.

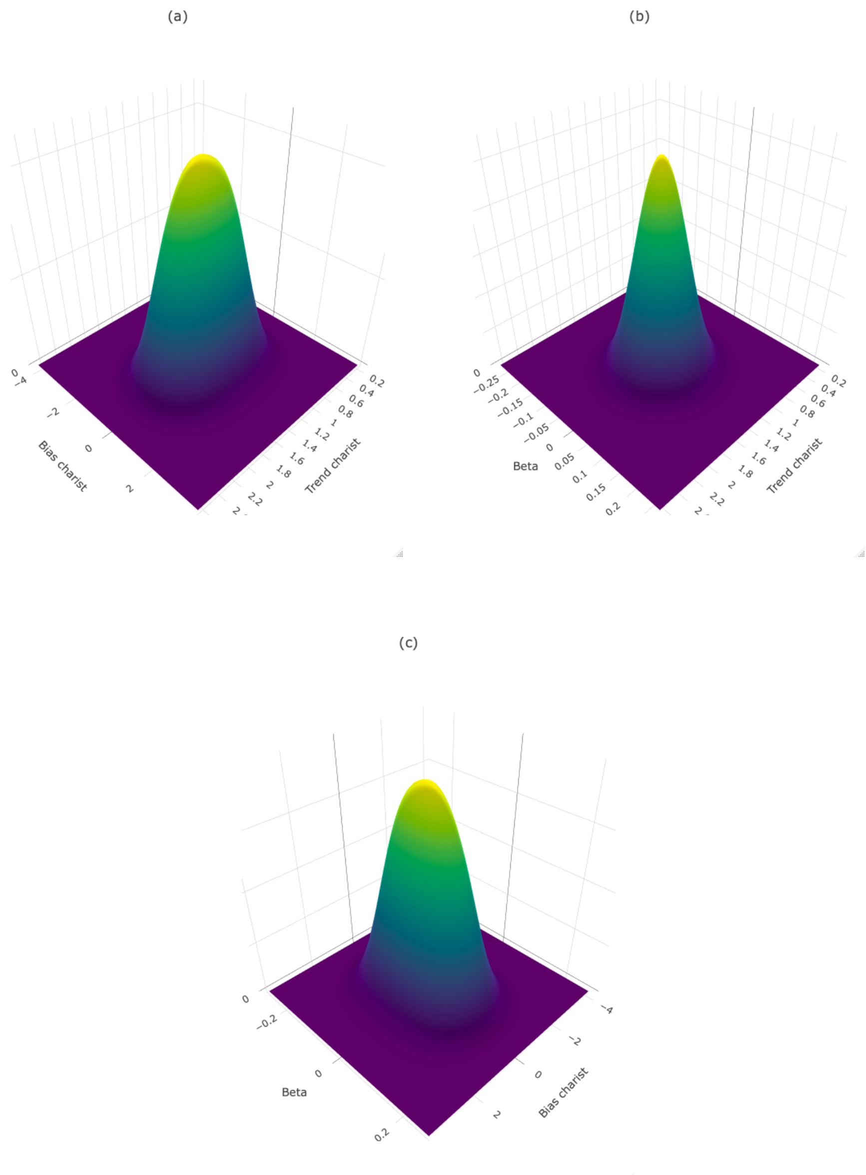

Figure 13: The plots (a),(b) and (c) depict the bivariate KDE posterior distribution for the three combinations of pair of parameters for the parameters ; trend chartist, bias chartist and the intensity of choice (beta).

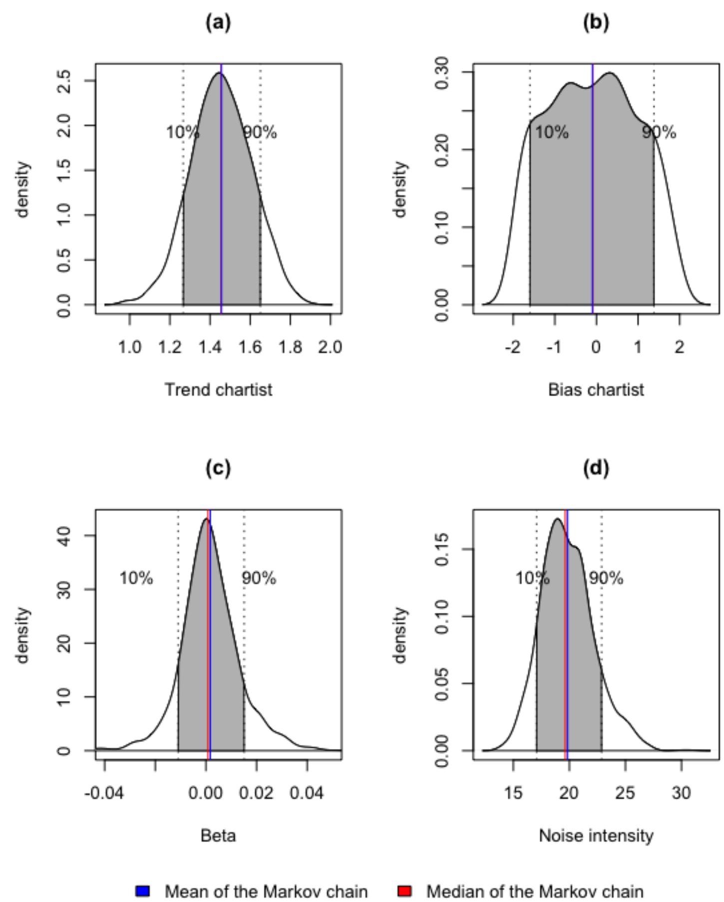

Figure 14: The figures (a),(b),(c) and (d) show, respectively, the KDE posterior density of the trend and bias parameters of the chartist, the intensity of choice and the noise intensity.

	mean	median	sd	min	max	skew	kurtosis
Trend chartist	1.370	1.364	0.148	0.872	1.833	-0.143	0.053
Bias chartist	-0.188	-0.316	1.083	-1.995	1.991	0.312	-0.964
Beta	0.004	0.002	0.051	-0.193	0.346	2.306	15.769
Noise intensity	16.754	16.528	1.632	12.514	22.668	0.380	0.028
Log-likelihood	-4.332	-4.327	0.030	-4.483	-4.266	-0.898	1.248
Acceptance rate	0.256	0.071	0.348	0.000	1.000	1.283	0.112
alpha	68.770	68.770	0.000	68.770	68.770	Inf	NaN
SD of the proposal distribution of trend chartist	0.153	0.157	0.022	0.082	0.195	-0.741	0.260
SD of the proposal distribution of bias chartist	0.557	0.568	0.169	0.196	0.798	-0.449	-0.924
SD of the proposal distribution of beta	0.040	0.042	0.019	0.000	0.100	0.218	0.803
SD of the proposal distribution of noise intensity	2.076	2.186	0.374	0.635	2.527	-1.733	2.825

Table 18: Descriptive statistic associated with the Markov chain.

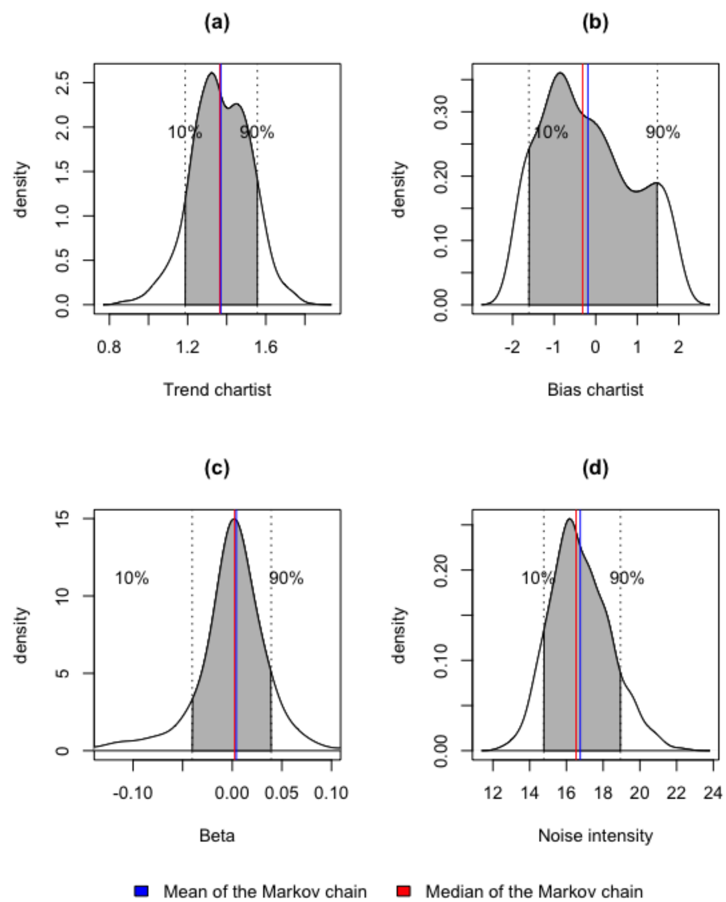

Figure 15: The figures (a),(b),(c) and (d) show, respectively, the KDE posterior density of the trend and bias parameters of the chartist, the intensity of choice and the noise intensity.

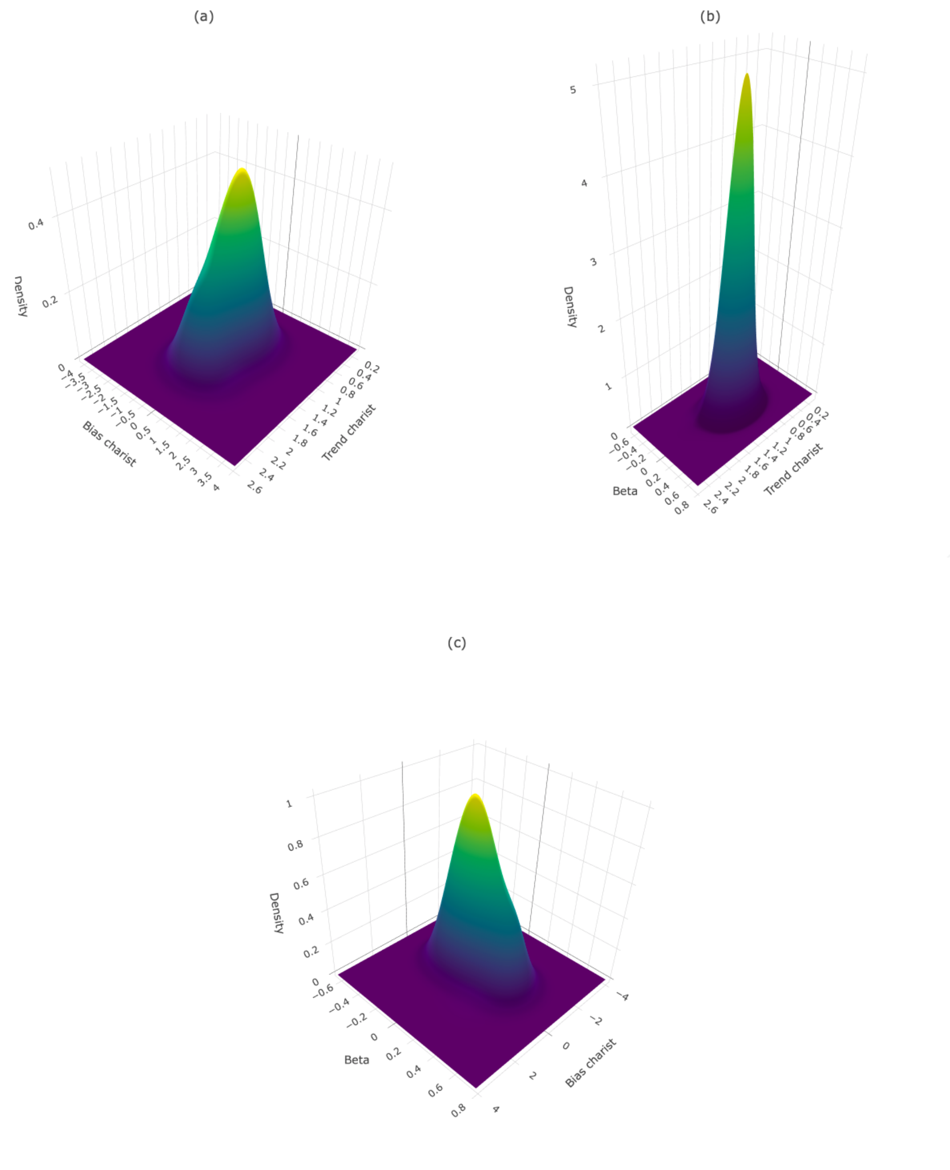

Figure 16: The plots (a),(b) and (c) depict the bivariate KDE posterior distribution for the three combinations of pair of parameters for the parameters ; trend chartist, bias chartist and the intensity of choice (beta).

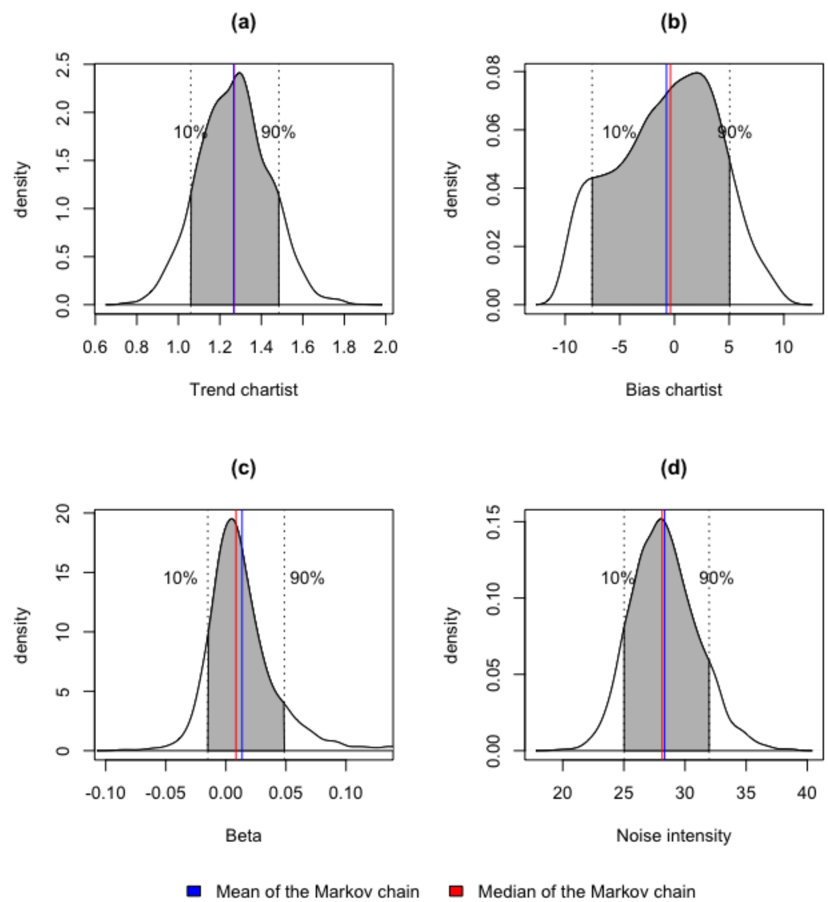

Figure 17: The figures (a),(b),(c) and (d) show, respectively, the KDE posterior density of the trend and bias parameters of the chartist, the intensity of choice and the noise intensity.

Appendix B : additional theory

1 Kernel density estimation

The Kernel Density Estimation (KDE) is a non-parametric density estimator and is considered as an extension of the histogram for continuous variable.

1.1 Univariate KDE

For a univariate random sample $\boldsymbol{X} = (X_1, ..., X_n)^T$, the univariate KDE can be written as follows :

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{43}$$

Where n is the number of observations, K is the kernel or the weight function and h is the smoothing parameter ("bandwidths") of the kernel. The kernel has the three following properties :

1. $K(x) \geq 0$ for all x
2. $\int K(x)dx = 1$
3. $\int xK(x)dx = 0$

The kernel function places more weight on x which are closer to observation X_i. Some common kernel are the Uniform, triangle, epanechnikov or Gaussian kernel.

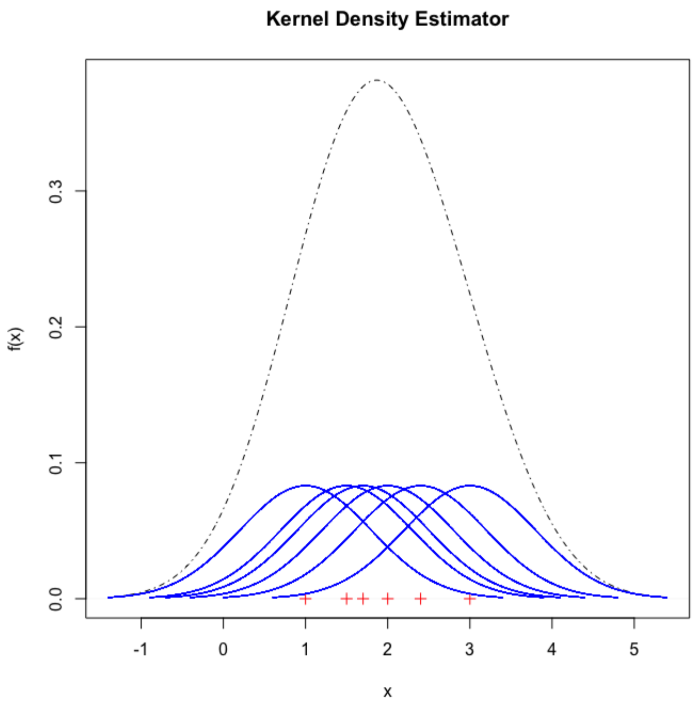

Figure 18: KD estimation of the density of six points with a Gaussian kernel and a bandwidth equal to 0.8. The black dashed line depicts the non-parametric density of the six points and the six blue lines show the Gaussian kernel associated with the six red points.

Throughout this thesis, the gaussian kernel and the Silverman's rule of thumb for finding an approximation of the optimal bandwidth will be used. The rule of thumb (Silverman, 1986) can be written as follows :

$$h = \left(\frac{4}{3n}\right)^{1/5} \hat{\sigma} \approx 1.06\hat{\sigma}n^{-1/5} \tag{44}$$

Where $\hat{\sigma}$ is the standard deviation of the sample, n the size of the sample and h the bandwidth.

1.2 Multivariate KDE

This part came from Duong (2007). For a d-variate random sample $\boldsymbol{X} = (\boldsymbol{X_1}, ..., \boldsymbol{X_n})^T$, the multivariate KDE can be written as follows :

$$\hat{f}(\boldsymbol{x}; \boldsymbol{H}) = \frac{1}{n} \sum_{i=1}^{N} K_H(\boldsymbol{x} - \boldsymbol{X}_i) \tag{45}$$

Where $\boldsymbol{x} = (x_1, ..., x_d)^T$ and $\boldsymbol{X}_i = (X_{i,1}, ..., X_{i,d})^T, i = 1, ..., n$. $K(\boldsymbol{x})$ is the multivariate kernel. The multivariate kernel has the three following properties :

1. $K(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$
2. $\int_{\mathbb{R}^d} K(\boldsymbol{x}) d\boldsymbol{x} = 1$
3. $\int_{\mathbb{R}^d} \boldsymbol{x} K(\boldsymbol{x}) d\boldsymbol{x} = 0$

The multivariate kernel can be for instance a multivariate Normal distribution[13], as $K(\boldsymbol{x}) = (2\pi)^{-d/2} exp(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x})$.
$\boldsymbol{H}$ is the bandwidth matrix which is symmetric and positive-definite, as $\boldsymbol{H} = diag(h_1^2, ..., h_d^2)$ and $K_{\boldsymbol{H}}(\boldsymbol{x}) = |\boldsymbol{H}|^{-1/2} K(\boldsymbol{H}^{-1/2}\boldsymbol{x})$.

The multivariate Gaussian kernel and the Silverman's rule of thumb for the bandwidth will be used in this thesis. The rule of thumb (Silverman, 1986) can be written as follows :

$$\boldsymbol{H}_{i,i}^{1/2} = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \sigma_i \tag{46}$$

Where σ_i is the standard deviation of the ith variable.

2 Bayesian statistic and maximum likelihood estimation

This section aims to summarize the general idea behind Maximum Likelihood Estimation (MLE) and Bayesian statistic. Bayesian statistic and MLE are introduced by Hartig et al. (2011, p.817) as follows :

[13]The package 'ks' enables to use kernel smoothers for univariate and multivariate data. See : https://www.rdocumentation.org/packages/ks/versions/1.10.7

> "The key idea underlying both Bayesian inference and maximum likelihood estimation is that the support given a parameter θ by the data D is proportional to $p(D|\theta)$, the probability that D would be observed given $M(\theta)$. [...] The word proportional is crucial both for Bayesian and likelihood-based inference, the value $p(D|\theta)$ carries no absolute information about the support for a particular parameter, but is only used to compare parameters by their probability of producing the observed data given the model M."

Maximum likelihood estimation

First of all, the likelihood function can be defined as follows :

$$\mathcal{L}(\theta) \propto p(D|\theta) \tag{47}$$

Where $\mathcal{L}(\theta)$ is the likelihood of the data given the parameter θ and $p(D|\theta)$ is the likelihood function. The goal of MLE is to find the parameters θ which maximise the likelihood function. The log-likelihood function can be written as follows :

$$\ell(\theta) \propto log\, p(D|\theta) \tag{48}$$

Bayesian statistic

Bayesian statistic can be seen as an extension of the likelihood function by adding prior information to the likelihood. Bayes' rule is the conditional probability density function $p(\theta|D)$ of the parameters θ conditionally to the data D. The Bayes' rule can be written as follows:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \tag{49}$$

Where $p(\theta|D)$ is called the posterior distribution and it depends on the probability of observing the data given a value of the parameter θ ($p(D|\theta)$) and the prior $p(\theta)$. The prior can be interpreted as additional knowledge about the parameter values before observing data D. This information can come from prior study or domain knowledge for instance. In the case where no initial knowledge is available, a non-informative prior can express this ignorance. An important conclusion on the existing literature is that non-information prior is not especially a uniform distribution (Hartig et al., 2011). In practice, only the numerator of the equation 49 matter and $p(D|\theta)$ is replaced by the likelihood (or the log-likelihood).

$$p(\theta|D) \propto \mathcal{L}(\theta)p(\theta) \tag{50}$$

The main difference between Bayesian statistic and MLE is that the former aims to compute a posterior distribution which is a conditional probability density function and the latter aims to find the maximum value of the likelihood. Moreover, the Bayesian statistic has the advantage to add prior information if there are available.