Quantitative Analyses on Non-Linearities in Financial Markets

Alessia Cafferata

Department of Economics (DIEC)

University of Genova

Supervisor

Marina Resta, University of Genoa
Fabio Tramontana, Cattholic University of Sacred Heart

In partial fulfillment of the requirements for the degree of

PhD program in Economics

January 31, 2019

Abstract

> "*The brief market plunge was just a small indicator of how complex and chaotic, in the formal sense, these systems have become. Our financial system is so complicated and so interactive [...]. What happened in the stock market is just a little example of how things can cascade or how technology can interact with market panic*" (Ben Bernanke, IHT, May 17, 2010)

One of the most important issues in economics is modeling and forecasting the fluctuations that characterize both financial and real markets, such as interest rates, commodities and stock prices, output growth, unemployment, or exchange rate. There are mainly two opposite views concerning these economic fluctuations. According to the first one, which was the predominant thought in the 1930s, the economic system is mainly linear and stable, only randomly hit by exogenous shocks. Ragnar Frisch, Eugen Slutsky and Jan Tinbergen, to cite a few, are important exponents of this view, and they demonstrated that the fluctuations observed in the real business cycle may be produced in a stable linear system subject to an external sequence of random shocks. This view has been criticized starting from the 1940s and the 1950s, since it was not able to provide a strong economic explanation of observed fluctuations. Richard Goodwin, John

Hicks and Nicholas Kaldor introduced a *nonlinear view* of the economy, showing that even in absence of external shocks, fluctuations might arise. The economists then suggested an alternative within the exogenous approach, at first by using the stochastic real business cycle models (Finn E. Kidland and Edward C. Prescott, 1982) and, more recently, by the adoption of the New Keynesian Dynamic Stochastic General Equilibrium (DSGE) models, very adopted from the most important institutions and central banks. These models, however, have also been criticized for the assumption of the *rationality* of agents' behaviour, since rational expectations have been found to be systematically wrong in the business cycle. *Expectations* are of fundamental importance in economics and finance, since the agents' decisions about the future depends upon their expectations and their beliefs. It is in fact very unlikely that agents are perfect foresighters with rational expectations in a complex world, characterized by an irregular pattern of prices and quantities dealt in financial markets, in which sophisticated financial instruments are widespread.

In the first chapter of this dissertation, I will face the *machine learning* technique, which is a nonlinear tool used for a better fitting, forecasting and clustering of different financial time series and existing information in financial markets. In particular, I will present a collection of three different applications of these techniques, adapted from three different joint works:

- *"Yield curve estimation under extreme conditions: do RBF networks perform better?*, joint with Pier Giuseppe Giribone, Marco Neffelli, Marina Resta, published Anna Esposito, Marcos Faundez-Zanuy, Carlo Francesco Morabito, Eros Pasero Edrs, Multidisci-

plinary Approaches to Neural Computing/Vol. 69/ WIRN 2017 and Chapter 22 in book "Neural Advances in Processing Nonlinear Dynamic Signals", Springer;

- *Interest rates term structure models and their impact on actuarial forecasting*, joint with Pier Giuseppe Giribone and Marina Resta, presented at XVIII Quantitative Finance Workshop, University of Roma 3, January 2018;
- *Applications of Kohonen Maps in financial markets: design of an automatic system for the detection of pricing anomalies*, joint with Pier Giuseppe Giribone and published on Risk Management Magazine, 3-2017.

In the second chapter, I will present the study *A financial market model with confirmation bias*, in which nonlinearity is present as a result of the formation of heterogeneous expectations. This work is joint with Fabio Tramontana and it has been presented during the X MDEF (Dynamic Models in Economics and Finance) Workshop at University of Urbino Carlo Bo.

Finally, the third chapter is a rielaboration of another joint paper, *"The effects of negative nominal risk rates on the pricing of American Calls: some theoretical and numerical insights"*, with Pier Giuseppe Giribone and Marina Resta, published on Modern Economy 8(7), July 2017, pp 878-887. The problem of quantifying the value of early exercise in an option written on equity is a complex mathematical issue that deals with continuous optimal control. In order to solve the continuous dynamic optimization problem that involves high non linearity in the state variables, we have adopted a discretization scheme based

on a stochastic trinomial tree. This methodology reveals a higher reliability and flexibility than the traditional approaches based on approximated quasi-closed formulas in a context where financial markets are characterized by strong anomalies such as negative interest rates.

Contents

List of Figures

List of Tables

Chapter 1

Applications of machine learning techniques in finance

The first chapter of this thesis deals with machine learning and its applications in finance. More precisely, I focus on the revision of three joint works in which we have applied machine learning to three different issues. The first two studies concern the application of supervised learning techniques to test the capabilities of different techniques in modelling the yield curve. At first, we test the capability of Radial Basis Function (RBF) networks to fit the yield curve under extreme conditions, namely in case of either negative spot interest rates, or high volatility. In particular, we compare the performances of conventional parametric models (NelsonSiegel, Svensson and de RezendeFerreira) to those of RBF networks to fit the term structure. To such aim, we consider the Euro Swap EUR003M Euribor, and the USDollar Swap (USD003M) curves, on two different release dates: on December 30th 2004 and 2016, respectively, i.e. under very different market situations, and we examined the various ability of the abovecited methods in fitting them. In the second study, we examine the discrepancies arising from the use of alternative interest rates databases, and we test the capabilities of differ-

ent techniques in modelling the yield curve. To do so, we consider the different bootstrapping methods used to obtain the yield shape implied by the treasury bonds to assess the problem of the discrepancies in pricing. In particular, we compared three approaches: the parametric models of Nelson-Siegel, Svensson and De Rezende-Ferreira; the interest rates structure methods based on the Vasicek and the Cox, Ingersoll and Ross stochastic processes; and different Machine Learning models to show how these discrepancies can impact on financial valuations. Thus, we compare the impact of the two databases with respect to the calculation of annuity factors, an issue which is of great importance in the insurance field. We choose this benchmark as it represents the pricing base of life-insurance contracts. To such aim, we examine two popular yield curve datasets: the Daily Treasury Yield Curve, and the time series posted in the Federal Reserve Board. We calibrate all the models using panel methods, showing that significant discrepancies on the estimation of temporary life annuities arise. The results of these analyses show that while in general conventional methods fail in adapting to anomalies, such as negative interest rates or big humps, RBF nets provide excellent statistical performances, thus confirming to be a very flexible tool capable of adapting to every markets condition. Finally, the aim of the last study is to implement an automatic method for organizing and clustering the information observed on secondary markets, focusing the attention on the recognition of potential anomalies. To do so, we use an unsupervised neural network, designed to cluster the data, known as Kohonen Network or Self-Organizing Map (SOM). We find out it is able to highlight potential opportunities of trading.

1.1 Introduction

The algorithms based on artificial intelligence, on self-learning strategies and on numerical strategies of soft-computing inspired by natural processes, are gaining importance in finance, since they allow faster analysis on a huge quantity of financial data. These methods are even more used in several field of applications, starting from biology to engineering, cognitive sciences, medical diagnosis and, of course, finance. (J. Schmidhuder, 2015; G. Huang, 2015). Artificial intelligence aims to replicate tasks traditionally requiring human sophistication with computation tools. The first applications of Artificial Intelligence trace back to the Eighties, like for example the *Self-Organizing Maps*, developed by Teuvo Kohonen in 1982, but nowadays recent increase in computing power coupled with increase in the availability and quantity of data have resulted in a resurgence of interest in potential applications of artificial intelligence (European Joint Committee Discussion Paper on the Use of Big Data by Financial Institutions,2016). Machine Learning are non-linear structures of statistical data organized as modelling tools. They can be used to simulate and describe complex relationships between inputs and outputs that other analytic functions would not be able to model. They are data-driven techniques that receive external signals through an input layer of neurons which is connected with other several internal neurons, organized in different layers. Networks are trained by operating on the difference between the actual output and the desired output of the network- the prediction error; by iterating, the weights are modified until the output error reaches an acceptable level. These techniques can leverage the ability of computers to perform tasks, such as recognizing images and processing natural languages, by learning from experience. (Financial Stability Board, 2017). There exist different types of learning algorithms, which vary according to the level of intervention in labelling the data.

- *Unsupervised learning*: it aims to group and interpret data based only on input data. It detects patterns in the data by identifying clusters of observations that depend on similar underlying characteristics: for this reason, it is very good for clustering;

- *Supervised learning*: it develops predictive model based on both input and output data. The algorithm identifies a general rule of classification and it will use it to predict the labels for the remaining observations in the data set. This kind of algorithm is used for classification or regression. One important methodology of supervised learning is the *deep learning*, a method based on learning data representation, where algorithms work in layers inspired by the structure of the human brain and the biological nervous system. Its structure are called artificial neural networks. These algorithms can be used for supervised, unsupervised and reinforcement learning. It is applied in different field, as image recognition and natural language processing.

- *Reinforcement learning*: it is a mid-way between supervised and unsupervised learning. It concerns how software agents should take actions in a certain environment so as to maximize some notion of cumulative reward. It is very common in game theory and robotics;

Of course, there are things that machine learning cannot do, such as determining causality: machine learning algorithms are often used to identify patterns that are correlated with other events or patterns. These patterns, however, simply highlight correlations, some of which are a sort of "black box", unrecognizable to the human eye. Despite this, economists are increasingly using artificial intelligence and machine learning applications, just because they help to understand complex relationships, that would be difficult to analyze with other tools.

1.1.1 Machine learning in finance: some applications

As I have mentioned before, machine learning techniques can be addressed to different purposes, such as regression, clustering, classification, forecasting and outlier detection and data quality. In particular, supervised learning deals with classification and regression problems while unsupervised learning can be used to solve clustering, and, rarely, forecasting and association problems. We refer to a classification problem when the output variable is a category, such as "true" or "false". On the other hand, in a regression problem the output variable is a real value. The problem of time series prediction is related to classification and regression approaches. A clustering problem is when you want to discover a way to group together objects that are similar; an association rule learning problem is where you want to discover rules that describe large portions of your data. In finance, machine learning has been applied to several purposes: to support technical analysis (Giribone et al., 2017), to recognize anomalies on data quality, to rebuild volatility surfaces for pricing, by accurately computing the fair value of an option and also to credit risk valuation.

In this chapter I am going to present three different case studies in which my coauthors and I have applied the algorithms of machine learning to financial markets. The first two applications deal with the fitting of the yield curve with two different purposes, that is to test the capability of Radial Basis Function (RBF) networks to fit the yield curve under extreme conditions, namely in case of either negative spot interest rates, or high volatility and to examine the discrepancies from using alternative interest rates databases, and to test the capabilities of different techniques in modelling the yield curve. The last study implements an automatic method for organizing and clustering the information observed on secondary markets, focusing the attention on the recognition of potential anomalies. Given the subjects of these case studies, I am going to focus only on machine

learning for regression and fitting and on outlier detection and data quality.

1.2 Regression and fitting

In this secton I present two applications of machine learning for regression and fitting problems. The function of regression and fitting of data estimate the outcome with an infinite number of possible solutions. In finance, these regressive methods can be used for option pricing or, as in our case, for modelling the yield curve. As for the modellization of the yield curve, for example, the traditional parametric techniques have the disadvantage of requiring an ex-ante functional form to which the term-structure observed in financial markets, after the estimation of parameters, must be fitted (Gilli et al., 2006). These approaches are thus not always able to provide a good fitting, when for example the term structures of interest rates present some anomalies, as we can currently observe in financial markets: negative interest rates, illiquidity, high volatility (Giribone, 2017). This limit can be overcome with the implementation of a machine learning technique based on neural network and radial basis functions. It is possible to design the function of regression with different types of networks. The first typology of network that can be used in an *Artificial Neural Network*. A neural network is a parallel distributed system, composed by simple units, which is able to synthesize knowledge processing information present in external data. Moreover, it does not require any aprioristic assumption on the functional form of the regression; for this reason, it is largely used in different fields of science and engineering (Principe, 2000). There exist several possible classes of ANN, the most common in literature is the one that distinguishes between *feed-forward* and *feedback* networks, but to our aim, we only consider the first class of networks. Another technique that can be used for fitting is the *Radial Basis Function*: a network in

which the radial basis function (real-valued function ϕ whose value depends only on the distance from the origin) takes the role of the activation functions. We have used this kind of network to fit the yield curve.

1.2.1 The yield curve: theoretical background

The yield curve represents a relationship between the spot rates of zero coupon bonds and their respective maturities and provides a way of understanding whether the economy will be strong or weak. Understanding the evolution of the yield curve is therefore an important issue in finance, especially for assets pricing, financial risk management and portfolio allocation. During the past forty years considerable research efforts have been devoted to this task. The two main research tracks refer to equilibrium models, as pioneered by Vasicek (1977), and statistical models. Here we are mainly concerned with discussing those latter, as the underlying approach is strongly related to our research question: can Radial Basis Function nets provide a suitable environment to fit the yield curve under extreme conditions? Statistical contributions embrace a wide range of techniques, including the smoothed bootstrap by Bliss and Fama (1987), and the parametric approaches by Nelson and Siegel (1987), Svensson (1996), and de Rezende and Ferreira (2011). These techniques deal with in–sample estimation of the yield curve, while the forecasting issue has been addressed in a more recent literature track. Diebold and Li (2006), pioneered the field with a dynamic version of the Nelson–Siegel model (NSm), and (2013) modified the NSm by way of a procedure based on ridge regression to avoid collinearity issues. Furthermore, approaches employing Machine Learning (ML) paradigms have been already explored by Ait Sahalia (1996), Cottrell Cottreli (1998) Tappinen (Tap1998) to mention a few, who proposed nonparametric models, with no restriction on the functional form of the process generating the structure of interest rates. More recently Bose et

al. (2006), Joseph et al. (2011), Rosadi et al (2011), Sambasivan and Das (2017) explained the behavior of the yield curve with various neural architectures, while Barunik and Malinska (2016) employed artificial neural networks to fit the term structures of crude oil future prices. However, so far very little attention has been devoted to the practice of fitting the yield curve in extreme conditions within the ML framework: to the best of our knowledge, Gogas et al.(2015) is the only attempt at forecasting recession from a variety of short (treasury bills) and long term interest rate bonds applying Support Vector Machines (1995) for classification. We therefore think that there is enough room for contributing with Radial Basis Function (RBF) Networks. Indeed RBF nets have been already widely employed in financial applications: for some examples one can refer to (2012)and to (2012); however, we intend to explore how much this technique can be effective in providing in–sample matching to the yield curve under conditions of stress, and we are going to compare RBF nets performances to those of traditional statistical models.

Now, I provide an overview on the estimation models generally employed to fit the yield curve: starting from the parametric models to the RBF.

1.2.1.1 Thirty years of parametric estimation models of the yield curve

We are mainly focused on the Nelson and Siegel model and on the extensions discussed by Svensson and by de Rezende and Ferreira: for other variants the reader can refer to (DePorter, 2007) Nelson and Siegel suggested to model the yield curve in the following way:

$$y_{NS}(t,\beta,\tau) = \beta_0 + \beta_1 \frac{\tau\left[1-\exp(-t/\tau)\right]}{t} + \beta_2 \frac{\tau\left[1-\exp(-t/\tau)\right] - \exp(t/\tau)}{t} \quad (1.1)$$

where the dependent variable y_{NS} represents the zero rate to be determined, t is the time to maturity, $\beta = [\beta_0\, \beta_1\, \beta_2]'$ is the parameters vector, with β_0 representing the impact of long–run yield levels, β_1 and β_2 expressing the short–term and the mid–term components, respectively; finally τ is the decay factor. By properly estimating the parameters value, (1.1) allows to explain the different shapes the yield curve can assume: flat, humped or S–shaped.

The extension suggested by Svensson in 1994 introduced the possibility to model a second hump in the yield curve:

$$\begin{aligned} y_{SV}(t,\beta,\tau) &= \beta_0 + \beta_1 \frac{\tau_1\,[1-\exp(-t/\tau_1)]}{t} + \beta_2 \frac{\tau_1\,[1-\exp(-t/\tau_1)] + -\exp(t/\tau_1)}{t} + \\ &\quad + \beta_3 \frac{\tau_2\,[1-\exp(-t/\tau_2)] - \exp(t/\tau_2)}{t} \end{aligned}$$

where $\beta = [\beta_0\, \beta_1\, \beta_2\, \beta_3]'$, with $\beta_0, \beta_1, \beta_2$ likewise in (1.1), while β_3 is the parameter associated to the second hump. Moreover, we now have: $\tau = [\tau_1\, \tau_2]'$ representing the decay factors associated to the three earlier parameters (τ_1) and to β_3 (τ_2), respectively.

Finally, de Rezende and Ferreira in 2011 discussed a five parameters extension of the Nelson–Siegel model, allowing to insert a third hump in the yield curve:

$$\begin{aligned} y_{dRF}(t,\beta,\tau) &= \beta_0 + \beta_1 \frac{\tau_1\,[1-\exp(-t/\tau_1)]}{t} + \beta_2 \frac{\tau_1\,[1-\exp(-t/\tau_1)] - \exp\left(t/\tau_1\right)}{t} + \\ &\quad + \beta_3 \frac{\tau_2\,[1-\exp(-t/\tau_2)] - \exp(t/\tau_2)}{t} + \beta_4 \frac{\tau_3\,[1-\exp(-t/\tau_3)] - \exp\left(t/\tau_3\right)}{t} \end{aligned}$$

where $\beta = [\beta_0\, \beta_1\, \beta_2\, \beta_3\, \beta_4]'$, with β_4 being the parameter associated to the third hump, and $\tau_3 \in \tau = [\tau_1\, \tau_2\, \tau_3]'$ representing the decay factor associated to β_4.

Clearly both the Svensson (SV) and de Rezende–Ferreira (dRF) variants of

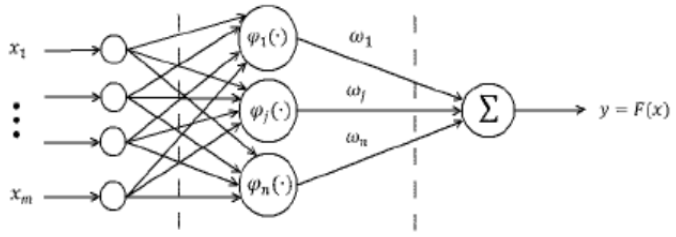

Figure 1.1: The traditional Radial Basis Function Network

the original Nelson–Siegel (NS) model are more complex to manage than the NS model, but they generally improve the desired fitting, by rising the number of parameters the related SSE, and MSE decrease and R^2 increases. In all the examined cases, the estimation of parameters can be performed by way of quasi–Newton methods like the Broyden–Fletcher–Goldfarb–Shanno algorithm –BFGS– [**?**] or with an optimization heuristic, as in (Gilli, 2010). However, while this latter solution seems to be capable of reliably solving the models, it fails (likewise the BFGS) in assuring the stability of estimated parameters under certain conditions, namely under small perturbations of the data. This motivated us to explore a non–parametric alternative, represented by RBF networks.

1.2.2 Radial Basis Function Networks

Radial Basis Function Networks –RBF– (Broom, 1998) are a kind of neural architecture generally organized into three layers, as illustrated in Figure 1.

Here $x_1, x_2, \ldots, x_m$ represent the components of the input vector $\mathbf{x}$ that are transmitted to the first layer nodes (in the same number as the input elements), by realizing a linear transformation. The signal then moves to the hidden layer where it is interpreted by a number of radial functions $\phi_j(\cdot)$, $j = 1, \ldots, n$, whose

number n is decided by the user, being:

$$\phi_j(\mathbf{x}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{c}_j||}{r_j}\right),\ j = 1, \ldots, n \tag{1.2}$$

where $\mathbf{c}_j$ and r_j are the center and the radius of the function, respectively. The characteristic feature of those functions $\phi_j(\cdot)$ is that their response decreases monotonically with distance from a central point. Finally the output neuron generates a weighted sum of the information processed by the hidden layer units:

$$F(\mathbf{x}) = \sum_{\mathbf{j=1}}^{\mathbf{n}} \omega_{\mathbf{j}} \phi_{\mathbf{j}}(\mathbf{x}) \tag{1.3}$$

The output signal $F(x)$ is then compared to the observed value, and the weights ω_j $(j = 1, \ldots, n)$ are adjusted accordingly, by way of an iterative process, until a stopping criterion is reached. It is a common practice to initialize the number n of nodes in the hidden layer to a small value, iteratively inserting an additional node if the desired tolerance is not fulfilled.

1.2.2.1 Application to the Euro Swap EUR003M Euribor, and the USDollar Swap curves: Simulation settings and results discussion

The goal of our work is to assess the capability of RBF nets to fit the yield curve in situations of stress, likewise in case of extreme humps or when interest rates turn negative. To such aim, we collected end–of–month data from Bloomberg sheets containing the bid and ask par rates for the Euro Swap – EUR003M Euribor, and the USDollar Swap –USD003M Curves, both observed on two different release dates, on December 30th 2004 and 2016, respectively. The average between bid and ask par rates was computed for each tenor and employed to derive the zero rates for each curve. We were therefore able to manage four curves whose name

Table 1.1: Yield curves employed in this work.

ID	Extended name	Inception Date	Length
EUR03MOld	3–Months Euribor	12/30/2004	22
EUR03M	3–Months Euribor	12/30/2016	22
USD03MOld	3–Months USDollar Swap	12/30/2004	22
USD03M	3–Months USDollar Swap	12/30/2016	22

is provided in Table 1.1.

Our choice can be easily motivated: the credit crunch in 2007–2008 and the Eurozone sovereign debt crisis in 2009–2012 have changed the fixed income market, fostering the emergence of the s.c. multiple curve issue [**?**] and altering traditional connections between interest rates and zero coupon bond prices [**?**]; our rationale is then to consider curves in both pre–crisis and crisis times to check the different fitting ability of conventional interpolation techniques against RBF nets. Figure 1.2 shows the dynamics of the yield curves under examination.

At first glance, the 3 Months Euribor Curves (both Eur003MOld and Eur003M) appear a bit more tricky to fit than the 3 Months USD Swap curve: starting from the graphs in the left–hand side, in fact, the USD003MOld curve is sensitively smoother than the EUR003MOld; moving to the right–hand side of Figure 1.2, the actual EUR003M profile shows singularities and slowdowns to negative values, while the USD003M is still quite flat, apart from a hump at short maturities. Figures 3-6 show the graphical comparison among the interpolations obtained with the various methods. Figures are self–explaining: the parametric methods work well, at the same level of the RBF net in interpolating the EUR003M old yield curve; the performance, however, is declining, at least for what is concerning the de Rezende–Ferreira approximation model, in the fitting of the USD003MOld; this is probably due to the known problems (already discussed in Sec. 1.2.2) of precision in the parameters estimation procedure. Things become worse when we

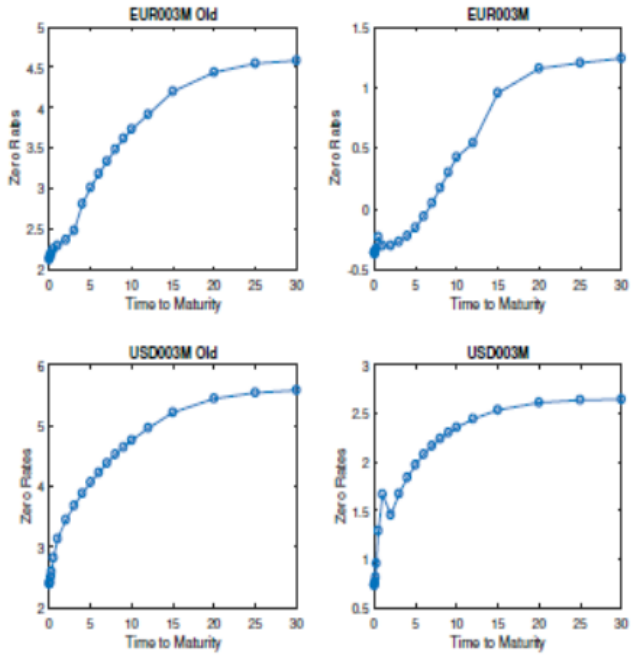

Figure 1.2: The dynamics of the examined yield curves: time to maturity is on the x–axis, while the value of zero rates appears on the y–axis. On top: the 3 Months Euribor Curve (release date Dec. 2004 on the left, release date Dec. 2016 on the right), on the bottom: the 3 Months USD Swap Curve (release date Dec. 2004 on the left, release date Dec. 2016 on the right)

turn to examine the estimation performed on actual data. In this case, the fitting to the observed yield curves is quite poor for all the examined parametric techniques; on the contrary the RBF net performances remain stable. This evidence is also confirmed by analysing the main statistics of the various methods, given in Table 1.2 for the parametric techniques, including parameters estimation, and in Table 1.3 for the RBF network.

The values in Tables 1.2–1.3 support the graphical evidence: quite surprisingly the de Rezende–Ferreira model is the worst, in terms of R^2 and RMSE, in three out of four cases (namely, in approximating the EUR003M, USD003MOld and USD003M curves). The remaining parametric techniques (Nelson–Siegel and

Table 1.2: Estimated coefficients for the parametric techniques

	Eur003M Swap Curve					
Coeff.	NS_{2004}	SV_{2004}	dRF_{2004}	NS_{2016}	SV_{2016}	dRF_{2016}
β_0	0.051205	0.049531	0.023256	0.019413	0.017525	0.023256
β_1	-0.029522	-0.028132	-0.024378	-0.022607	-0.021132	-0.024378
β_2	-0.029323	0.00111	-17.834098	-0.029211	-0.006591	-17.83410
β_3	–	-0.05842	17.874810	–	-0.053531	17.87481
β_4	–	–	-0.157251	–	–	-0.157251
τ_1	2.471681	0.740510	0.230981	3.4671849	0.556850	0.230980
τ_2	–	1.687435	0.232039	–	2.3897270	0.232039
τ_3	–	–	1.048890	–	–	1.048890
R^2	0.998042	0.997884	0.998594	0.986832	0.992417	0.314388
$RMSE$	0.000423	0.000467	0.000408	0.000701	0.000564	0.005736
	USD003M Swap Curve					
Coeff.	NS_{2004}	SV_{2004}	dRF_{2004}	NS_{2016}	SV_{2016}	dRF_{2016}
β_0	0.060161	0.0606401	0.023256	0.028773	0.027744	0.023256
β_1	-0.035384	-0.0368601	-0.024378	-0.020578	-0.020762	-0.024378
β_2	0.003668	-0.0145134	-17.83410	0.002111	0.105636	-17.834098
β_3	–	-0.0512392	17.87481	–	-0.122490	17.874810
β_4	–	–	-0.157251	–	–	-0.157251
τ_1	4.034936	0.5146503	0.230981	2.641094	0.583050	0.230981
τ_2	–	2.0924553	0.232039		0.704717	0.232039
τ_3	–	–	1.048890	–	–	1.048890
R^2	0.994607	0.9996403	N.A.	0.965927	0.981953	N.A.
$RMSE$	0.000905	0.0002480	0.1355922	0.001413	0.001091	0.173339

Table 1.3: RBF nets settings for the observed yield curves

	Eur003MOld	Eur003M	USD003MOld	USD003M
Max Nr Neur.	100	100	100	100
RMSE	7.32293E-11	1.30206E-10	9.38584E-11	1.96893E-10

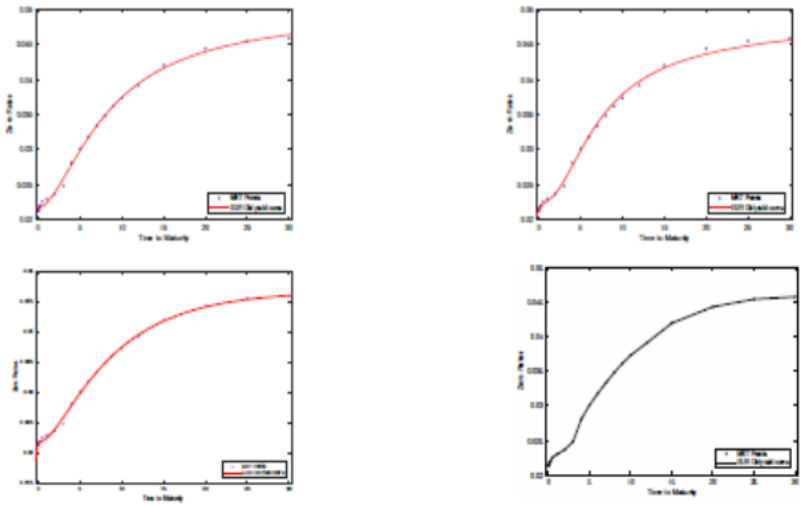

Figure 1.3: From top to bottom and from left to right: interpolation of the EUR003M Old yield curve with the Nelson–Siegel (NS), Svensson (SV), de Rezende–Ferreira (dRF) models and with the RBF Network

Svensson) maintain satisfying values of the R^2, but the related RMSE is sensitively higher than in the case of RBF net interpolation.

The results thus confirm that RBF nets can reach very satisfying results to manage anomalies such as extreme humps or negative interest rates.

1.2.2.2 Application to actuarial forecasting

Interest rates term structures are of fundamental importance for pricing financial instruments and/or insurance contracts: it represents a basic tool for both scholars and practitioners. Due to its importance, the task has gained a considerable interest from the researchers. The studies about this issue can be gathered into two branches: the first one refers to equilibrium models, as pioneered by Vasicek () while the second one refers to statistical models. A huge amount of contributions belongs to this latter field, but in this paper, we focus on the most famous contributions of the smoothed bootstrap by Bliss and Fama (1987), and

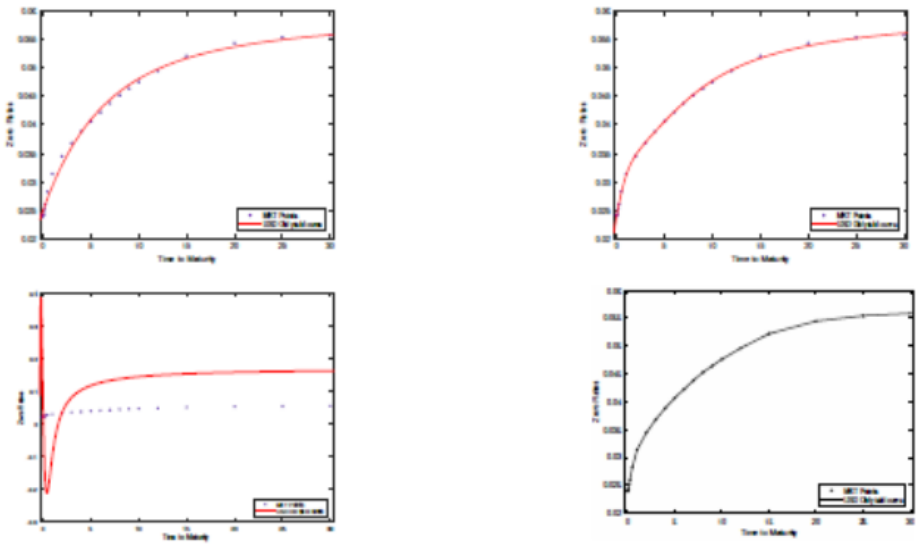

Figure 1.4: From top to bottom and from left to right: interpolation of the USD003M Old yield curve with the Nelson–Siegel (NS), Svensson (SV), de Rezende–Ferreira (dRF) models and with the RBF Network

on the parametric approaches by Nelson and Siegel (1987), Svensson (1996) and De Rezende and Ferreira (2013). Recently, approaches based on Machine Learning paradigms have also been deepened: this is the case of Ait Sahalia (1996), Cottrell et al. (1998) and Tappinen (1998), who all proposed nonparametric models in which no restriction on the functional form of the process generating the structure of the interest rates are present; Bose et al. (2006), Joseph et al (2011), Rosadi et al (2011), Sambasivan and Das (2017) implemented various neural architectures to illustrate the behavior of the yield curve; Barunik and Malinska (2016) fitted the term structure of the crude oil future prices with artificial neural networks; Gogas et al. (2015) gave an attempt to forecast recession from a variety of short and long term interest rates bond applying the Support Vector Machines for classification. However, some problems might arise since the zero-coupon interest rates can be retrieved by different database providers. For

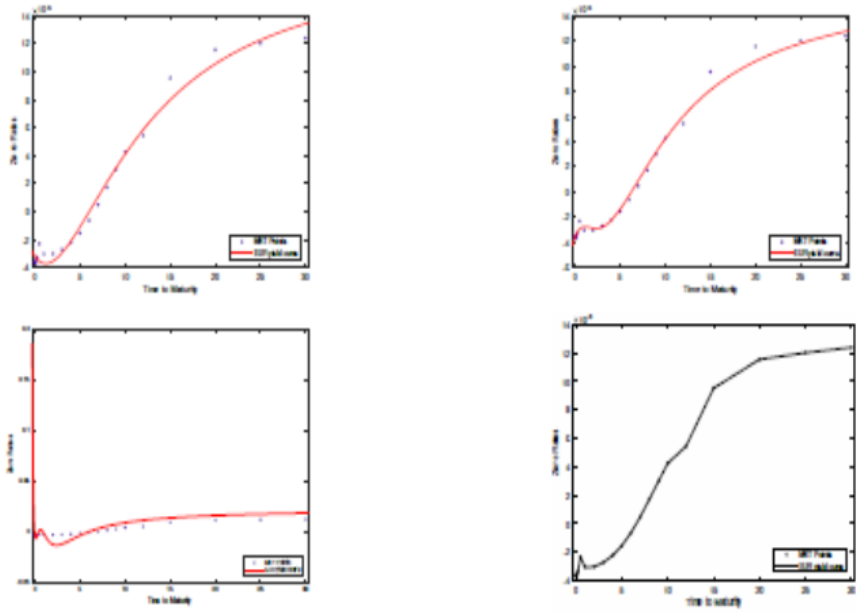

Figure 1.5: From top to bottom and from left to right: interpolation of the EUR003M yield curve with the NS, SV, dRF models and with the RBF Network.

different reasons, more than one yield curve dataset has become popular, but they can differ considerably from each other. Thus, an analyst must take into account how much these essential measures- together with the price of the instrument- can be affected by an inaccurate process of yield curve modeling. There exists two sources of discrepancies: first, the different models and numerical techniques used to estimate the zero-coupon rates; second, as the information is generally retrieved from a database provider, the same maturity can match to different risk-free debt instruments quotes. The first point has been investigated in several papers, such as Bliss, 1996 and Bolder et al., 2012, while DAmato et al. (2016) have faced the problem as a whole. Starting from this study, we examine the discrepancies arising from the use of alternative interest rate databases, and to test the capabilities of different techniques in modelling the yield curve. However, our approach goes beyond DAmato et al. (2016), since we also fit the interest

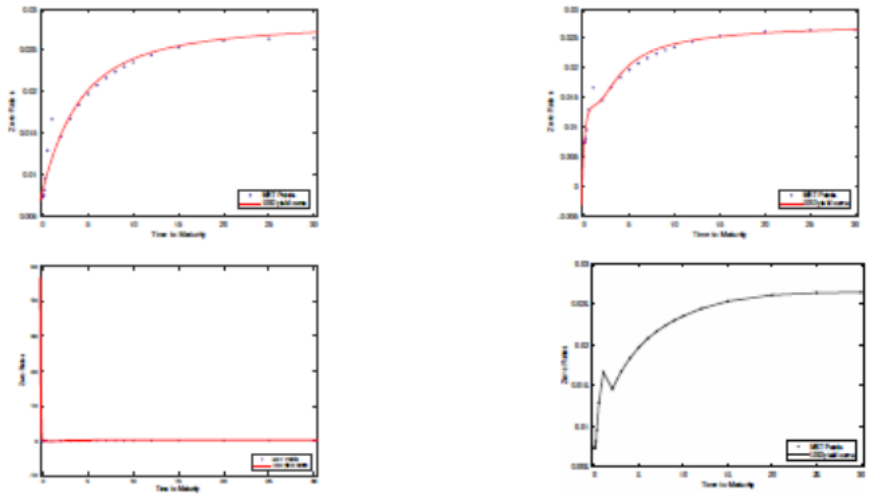

Figure 1.6: From top to bottom and from left to right: interpolation of the USD003M yield curve with the with the NS, SV, dRF models and with the RBF Network

rate term structure by a set of models that allows a more robust analysis of the zero-rates curve and the discount factors, considering the parametric approaches by Nelson and Siegel (1987), Svensson (1996) and De Rezende and Ferreira (2013) and also adding a battery of Machine Learning Methods. These methods are even more used in several fields of applications, starting from biology to engineering, cognitive sciences, medical diagnosis and, of course, finance. (J. Schmidhuder, 2015; G. Huang, 2015). In practical terms, Machine Learning methods are non-linear structures of statistical data organised as modelling tools. They can be used to simulate and describe complex relationships between inputs and outputs that other analytical functions would not be able to model. They are data-driven techniques that receive external signals through an input layer of neurons which is connected with several other internal neurons, organized in different layers. Networks are trained by operating on the difference between the actual output and the desired output of the network- the prediction error; by iterating, the

weights are modified until the output error reaches an acceptable level. In this study, we compare the performance of conventional models used for deriving the calibration of the term-structure models. We have considered three approaches: the parametric models of Nelson-Siegel, Svensson and De Rezende-Ferreira; the interest rate structure methods based on the Vasicek and the Cox, Ingersoll and Ross stochastic processes; and the battery of Machine Learning Models. Finally, in the third part of the paper, we show how these discrepancies can impact on financial valuations. Thus, in this last part, we compare the impact of the two databases with respect to the calculation of annuity factors, issue that is of great importance in the insurance field. We choose this benchmark as it represents the pricing base of life-insurance contracts. To do so, we examine two popular yield curve datasets: the Daily Treasury Yield Curve2, and the time series posted in the Federal Reserve Board3. We calibrate all the models using panel methods. We apply this method to different maturities: 3- and 6-months and 1-, 2-, 3-, 5-, 7-, 10-, 20-, 30-years. In addition to the parametric estimation models of Nelson and Siegel, Svensson and De Rezende and Ferreira I have explained before, in this study we have also employed the stochastic representations of the short-term interest rate. The most widespread models who belong to this category are the Vasicek model (1977) and the Cox, Ingersoll and Ross model (1985). These models describe the evolution of the interest rates. According to Vasicek(1977), the interest rate is supposed to follow a differential stochastic equation:

$$dr_t = \kappa \left[\theta - r_t\right] dt + \sigma dW_t \tag{1.4}$$

$$\text{Zero-coupon bond: } P(t,T) = A(t,T) \cdot \exp\left[-B(t,T)r(t)\right]$$

$$A(t,T) = \exp\left[\left(\theta - \frac{\sigma^2}{2\kappa^2}\right)\left[B(t,T) - (T-t)\right] - \frac{\sigma^2}{4\kappa}B(t,T)^2\right]$$

$$B(t,T) = \frac{1}{\kappa}\left[1 - \exp\left[-\beta(T-t)\right]\right]$$

Where W_t is a Wiener process, σ determines the volatility of the interest rate, r_t is the interest rate and k, θ are positive parameters that characterize the dynamic of the equation. In particular, θ represents the long-run level of the process.

The CoxIngersollRoss model (or CIR model) is a short rate model which models interest rate movements as driven by only one source of market risk:

$$dr_t = \kappa\left[\theta - r_t\right]dt + \sigma\sqrt{r_t}dW_t \tag{1.5}$$

$$\text{Zero-coupon bond: } P(t,T) = A(t,T)\cdot\exp\left[-B(t,T)r(t)\right]$$

$$A(t,T) = \left(\frac{2h\cdot\exp\left[\frac{(\kappa+h)(T-t)}{2}\right]}{2h+(\kappa+h)\left(\exp\left[h(T-t)-1\right]\right)}\right)^{\frac{2\kappa\theta}{\sigma^2}}$$

$$B(t,T) = \frac{2\cdot\left[\exp\left(h(T-t)-1\right)\right]}{2h+(\kappa+h)(\exp\left(h(T-t)-1\right))}$$

$$h = \sqrt{\kappa^2+2\sigma^2}$$

To **calibrate** the models, we have looked at Optimization techniques which are often very important in mathematical finance, especially in cases in which an exact calibration of the target value is hardly possible. In this case, indeed, it is not possible to solve directly the relevant equations (4) and (5). We therefore have to opt for an optimization method thanks to which the deviation between model and target values of certain state variables becomes minimal. In particular, to calibrate the parameters of the stochastic models, we resort to a deterministic expression that determines the fair value of a financial instrument which follows the supposed dynamic and which is also function of the parameters composing

the reference differential stochastic equation. The parameters of the stochastic model are then calibrated through a maximum likelihood method. As for Vasicek, we have a set of closed formula which allow a direct tuning of the parameters θ, σ and k, starting from the rates zero rates.

$$\hat{\theta} = \frac{S_y S_{xx} - S_x S_{xy}}{n\left(S_{xx} - S_{xy}\right) - \left(S_x^2 - S_x S_y\right)} \tag{1.6}$$

$$\hat{\kappa} = -\frac{1}{\delta} \ln\left(\frac{S_{xy} - \theta S_x - \theta S_y + n\theta^2}{S_{xx} - 2\theta S_x + n\theta^2}\right) \tag{1.7}$$

$$\hat{\sigma}^2 = \frac{2\kappa\left[S_{yy} - 2\alpha S_{xy} + \alpha^2 S_{xx} - 2\theta(1-\alpha)\left(S_y - \alpha S_x\right) + n\theta^2\left(1-\alpha\right)^2\right]}{n\left(1-\alpha^2\right)} \tag{1.8}$$

Where:

$\alpha = \exp\left(-\kappa\delta\right)$, $S_x = \sum_{i=1}^{n} r_{i-1}$, $S_y = \sum_{i=1}^{n} r_i$, $S_{xx} = \sum_{i=1}^{n} r_{i-1}^2$,$S_{yy} = \sum_{i=1}^{n} r_i^2$, $S_{xy} = \sum_{i=1}^{n} r_{i-1} r_i$ and δ is the time period to which the zero rate refers to. As for CIR, we suggest to use Ordinary Least Squares (OLS). For performing OLS, we transform (5) into:

$$\frac{r_{t+\Delta t} - r_t}{\sqrt{r_t}} = \frac{\kappa\mu\Delta t}{\sqrt{r_t}} - \kappa\sqrt{r_t}\Delta t + \sigma\epsilon_t \tag{1.9}$$

Finally, as for the **Machine Learning Method**, in addition to the RBF, we have considered other techniques:

- Feed Forward Artificial Neural Networks. Feed-forward Artificial Neural Networks are the simplest type of ANN: connections between the units, i.e. information, just moves in one direction, forward, from the input knots through the output knots. They are just composed by one input layer,

whose knots are used to receive information from the data, by one hidden layer and by one output layer. They use back propagation algorithm to update weights. These networks are very useful in function approximation when one only knows a set of inputs and outputs. They are static networks. As we told before, the focus here is the fitting of the interest rates term structure, so only one output node is needed. Otherwise, as in Zhang 2012, *the size of the output layer is usually determined by the nature of the problem*. Here, however, there is no feedback from the network output.

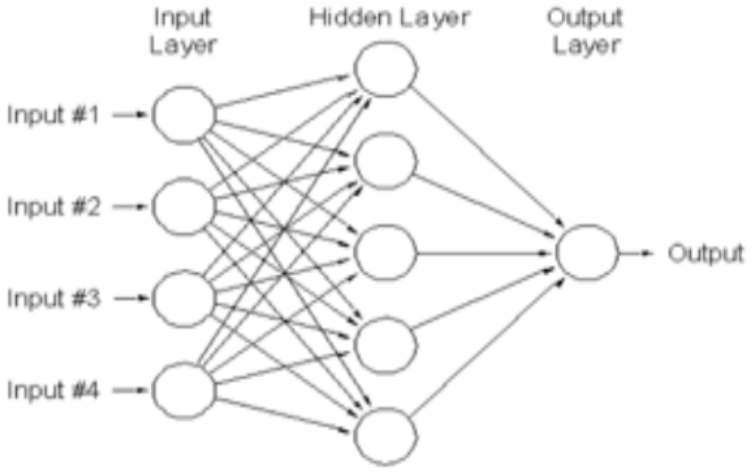

Figure 1.7: A feed forward Artificial Neural Network

- Cascade Forward Artificial Neural Networks. Cascade-forward Artificial Neural Networks are really similar to Feed-forward but they include a weight connection from the input to each layer and from each layer to successive layers. The main characteristic of this network is that each layer is connected to all the previous ones. For their more complex structure, they are able to learn complex relationships more quickly (Badde et al., 2013).

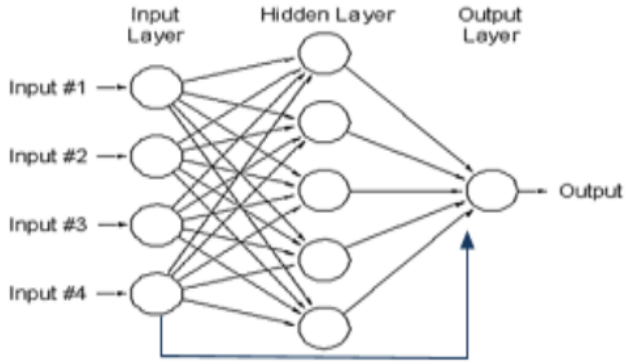

Figure 1.8: A Cascade forward Artificial Neural Network

- Deep Learning Method. The deep learning method is the natural extension of Artificial Neural Networks. It is a technique of Machine Learning which implements a deep neural network, that is an ANN with more than one hidden layers. The most critical point for an efficient design of a network characterized by more than two hidden layers, is its associated training rule. The difficulties associated to this aspect are essentially three: vanishing gradient, overfitting and huge computational load. When we talk about Vanishing gradient, we mean the problem linked to the back propagation algorithm: the error observed in the output neurons vanishes when they cover backwards the hidden layers. As a result, we obtain an ineffective upgrade of the weights of the neurons that are closer to the input layer. To avoid this risk, the solution is the adoption of a special activation function, which is sharper than the sigmoid function. In particular, the activation function that transmitted the error in the best way is the Rectified Linear Unit (ReLU). This is defined as:

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} = \max(0, x) \rightarrow f'(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1.10}$$

The sigmoid puts upper limits to the output associated to the neurons to the independent unit of the magnitude of the input signal. The ReLU does not have this upper bound, so it succeeds in transmitting the error in a more effective way to the inner layers of the network. The second critical point concerns the overfitting problem. Deep neural networks are for sure more vulnerable with respect to the other network architectures, since the model becomes more complex, and it includes more hidden layers- and then more weights. The more characteristic solution of this problem is called dropout. This technique consists in training only a randomly selected group of nodes instead of the net as a whole: a percentage determines how many nodes have to be chosen for each layer, while the remainders are deactivated. Since the neurons and the associated weights are continuously modified, we can avoid the problem of overfitting. The last problem of huge computational load is a typical technological issue: the training of deep learning networks requires a lot of time, since the number of weights that have to be calibrated increases together with the number of the hidden layers, leading to the need for a greater number of training data. The most recent hardware provides greater performances and the opportunity of parallel computing, allowing a huge spread of the techniques based on the deep learning.

We have done this analysis on the most retrieved yield curves, which came from two different daily datasets: the Daily Treasury Yield Curve Rate and one posted in the Federal Reserve website. The data of these two time series may differ substantially, because of the different models and the different techniques used to estimate the zero coupon rates (Bliss, 1996),(Bolder, 2012) and also because of the different prices used as inputs. We have thus examined the two datasets listed above. The first one is simultaneously posted in two different official websites: in the Daily Treasury Yield Curve Rate in the US Department of the Treasury

website, and also appears as U.S. government securities. We name this dataset DoT. The second dataset which we considered is posted in the Federal Re-serve Board. There are different kinds of discrepancies among the two datasets, but along the line of DAmato et al. (2016), we mainly focus on four of them:

- The fitting methods: the DoT interpolates among yields-to-maturity of observed Treasury securities with a quasi-cubic Hermite spline function, while the FRB follows a weighted Svensson model to fit the yield curve.

- The yields that are used as the input in the fitting process. The DoT in fact fits yields-to-maturity while FRB uses as dependent variable bond prices, since the DoT employs close of business bid yields-to-maturity but the RBF estimates are based on end-of-day prices, altohugh they do not specifiy which kind of prices they use. Grkaynak et al. ()

- The basket of assets from which the yield curve is estimated, since the two series consider different instruments with different maturities. This has different implications, the most important is the liquidity differences they lead to.

- The reported interest rates. The DoT provides par yields whereas the FRB shows spot or zero-coupon rates. DoT considers on-the-run bills and bonds, so that the resulting yield curve can be interpreted as a par yield curve instead of the desirable zero-coupon yield curve. (DAmato et al. 2016).

We are interested in applying our study to life annuity factors, a_x. They combine information on discount factors with survival probabilities. It is defined as:

$$a_{x,w} = \sum_{t=0}^{w-x} P_{x,t} \cdot DF_t = \sum_{t=0}^{w-x} P_{x,t} \cdot \frac{1}{(1+r)^t} \tag{1.11}$$

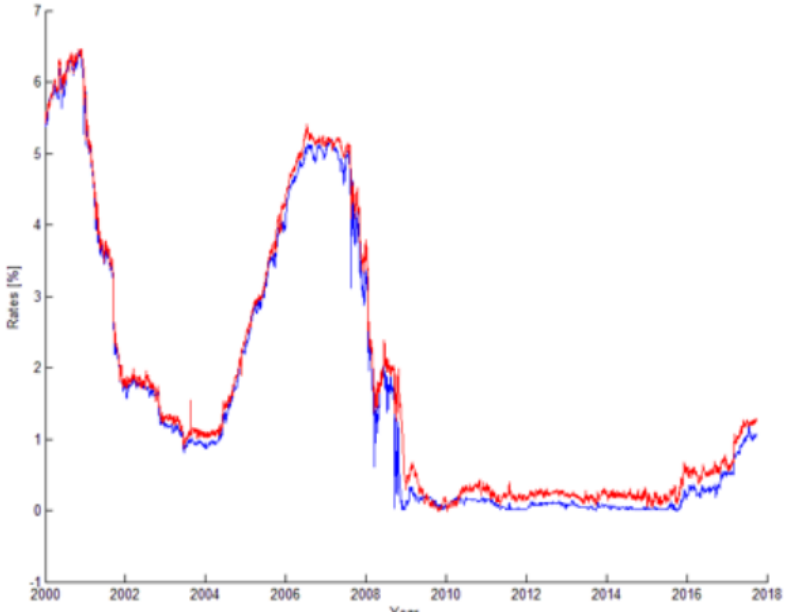

Figure 1.9: The historical time series of the three-months rate of the two datasets.

Where $P_(x,t) = p_x...p_(x+t-1)$ is the cumulative probability that a man aged x will survive to age $x+t$, $p_(x+t-1)$ indicates the conditional year-to-year survival probability for the person x and DF_t is the discount factor at time t. The survival probability is derived form a mortality table with ending age ω and it varies according to the sex of the person. As we can notice, life annuity factors depend on age, sex, interest rate and mortality patterns; for this reason, they are very useful for projecting pension liabilities.

As mentioned above, the aim of this study is to analyze the discrepancies that can arise from using alternative interest rate databases and to test the different capabilities of the existing techniques used to model the yield curve. To do so, we considered the two most popular yield curve datasets: the DoT and the FRB. The discrepancies between the two series can have effects on the interest rate structure model forecasting and they can also lead to practical consequences. To study the first group of consequences, we focused on the interest rate model parameters

Stochastic Process	k	θ	σ
Vasicek (DoT)	0.1451	0.0351	0.000162
Vasicek (FRB)	0.1608	0.0382	0.0028
CIR (DoT)	0.1210	0.0374	0.0053
CIR (FRB)	0.1259	0.0427	0.0046

Table 1.4: Estimated parameters for the stochastic processes

Model	β_0	β_1	β_2	β_3	β_4	τ_1	τ_2	τ_3	R^2
NSD	3.073	-2.065	0.0015	-	-	3.789	-	-	0.99
NSF	3.432	-2.23	-0.0016	-	-	6.075	-	-	0.99
SvD	2.939	-1.991	-168.8	167.5	-	1.463	1.456	-	0.99
SvF	4.664	-3.403	-2.555	-4.517	-	1.935	10.54	-	1
RFD	2.505	-1.573	-1.08	4.739	-2.862	0.1751	133.2	1.03	0.99
RFF	2.803	-7.101	8.654	-5.831	-0.087	0.0544	0.5275	0.67	n.d.*

Table 1.5: Estimated parameters for traditional parametric models.* Results caused by overfitting problems

employed on different datasets. We then consider the most popular interest rate structure models- i.e. the CIR and the Vasicek stochastic processes- the traditional parametric methods- Nelson and Siegel (), Svensson () and De Rezende and Ferreira () and the Machine Learning methodologies to check the importance of the choice of different datasets during the calibration of the parameters of a stochastic term structure model. The calibration of these two models have been performed on the two interest datasets mentioned above, using a panel method. Using the theoretical framework shown above, we move to the calibration of the parameters of the considered fitting models. In the three tables below, we show the estimated values referred to interest rate market term structures observed on September, 29th 2017.

ML technique	Num. hidden layers	Neurons for layers	Perf. Func.
RBF Network	1	121	$MSE \leq 10^{-6}$
Feedforward ANN	1	50	$MSE \leq 10^{-6}$
Cascade ANN	1	15	$SSE \leq 10^{-6}$
Deep Network	2	10	$SSE \leq 10^{-6}$

Table 1.6: Neural networks architecture.

1.2.2.3 Application to actuarial forecasting: Results discussion

After the estimation of the model parameters, we are able to model the different term structures. In the three panels below, we show the results; in particular, in the top part of all the panels we have plotted the zero rate curves, while in the bottom graphs the associated discount factors are pointed out. The left part of the panels is referred to the DoT database, while in the right part the data concerns the FRB time series. The blue dots represent the market zero rates, directly retrieved from the two databases. These dots are fitted by the reference regressive model. Going into further detail, the first panel shows the traditional parametric fitting model: the blue line designates the Svensson model, the red line illustrates the Nelson-Siegel model and the black line corresponds to De Rezende and Ferreira model. As discussed in section 2.1.1, we obtaine that the De Rezende and Ferreira model is not suitable for the representation of the FRB zero rate since an overfitting problem has occurred. In the second panel, we indicate with the blue line the term structure obtained from the CIR stochastic process and with the red line the curve from Vasicek approach. It is clear that the explanatory power is lower: this is caused by the fact that we have to calibrate only three parameters. On the other hand, in the third panel it is clearly shown that the machine learning techniques outperform the previous approaches. This result is fully confirmed by their excellent ability to fit (the associated R^2 is very close to 1) and the robustness associated to all four different approaches: independently

from the different architecture we obtain the same optimal fitting power.

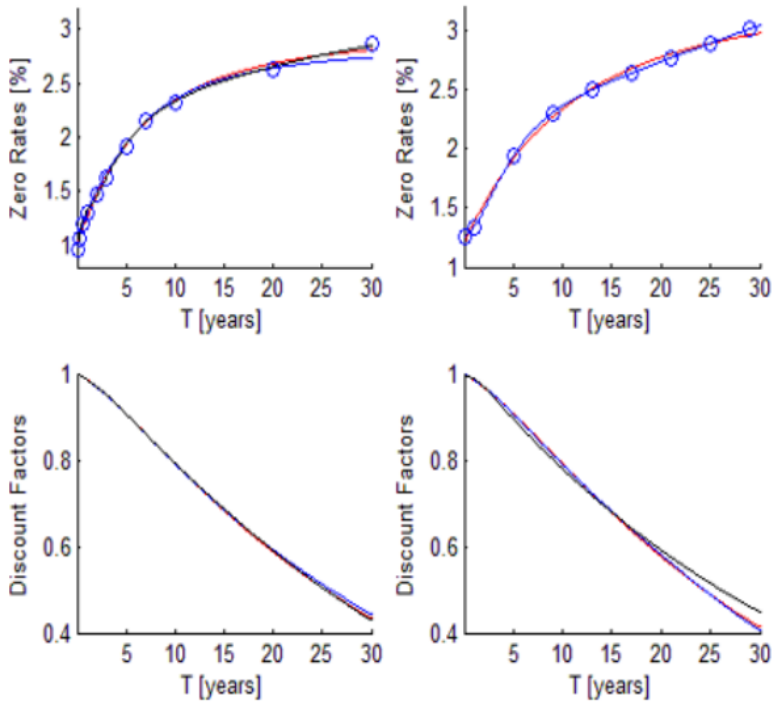

Figure 1.10: Interest rate parametric fitting model

This study is applied to the determination of the annuity factors. In particular, we estimate the annuity factors starting from the term structure fitting models widely discussed above, showing the discrepancies caused by the adoption of different modelling approaches and by the two different datasets. Figure 5 shows the annuity factors associated to a man of an age varying from 65 to 95 for the following ten years. The bundle of blue curves represents the models constructed on the RFB datasets, while the red bundle designates the models which uses the DoT as reference dataset. From this figure we can deduce that both the different modelling approaches and the different datasets are of extreme importance. The two bundles of curves in fact cross each other: this means that both a risk model and an error associated to two different datasets are signifi-

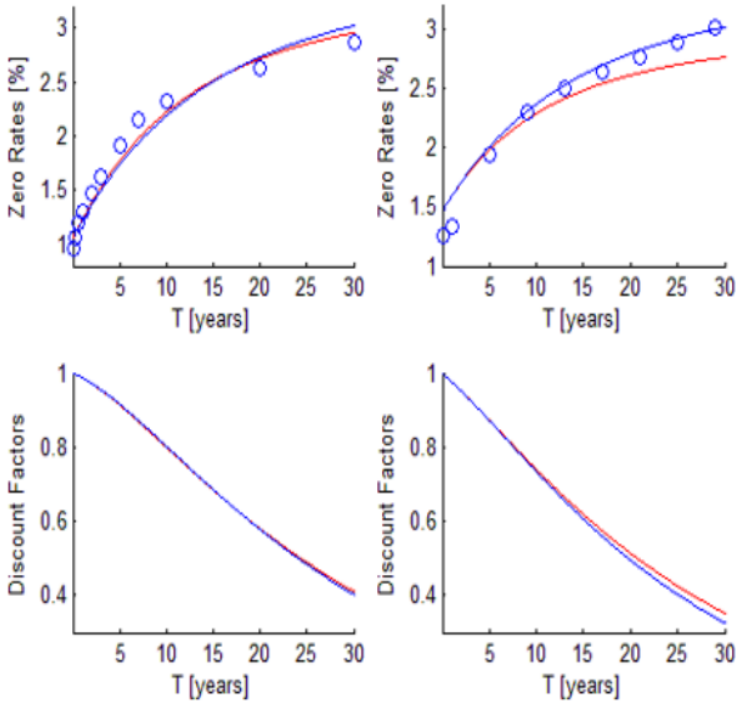

Figure 1.11: CIR and Vasicek interest rate model

cant in the determination of annuity factors. So, as our results have shown, the machine learning techniques gave the most satisfactory results in modeling the interest rate term structure.

1.3 Clustering

In this last section I am going to discuss an application of machine learning for clustering. This technique belongs to the unsupervised machine learning methods. Clustering allows to divide the population into a number of groups in such a way that each data points in the same groups have similar characteristics to other data points in the same group and dissimilar to the data points in other groups. The criterion is the similarity or the dissimilarity between objects. As

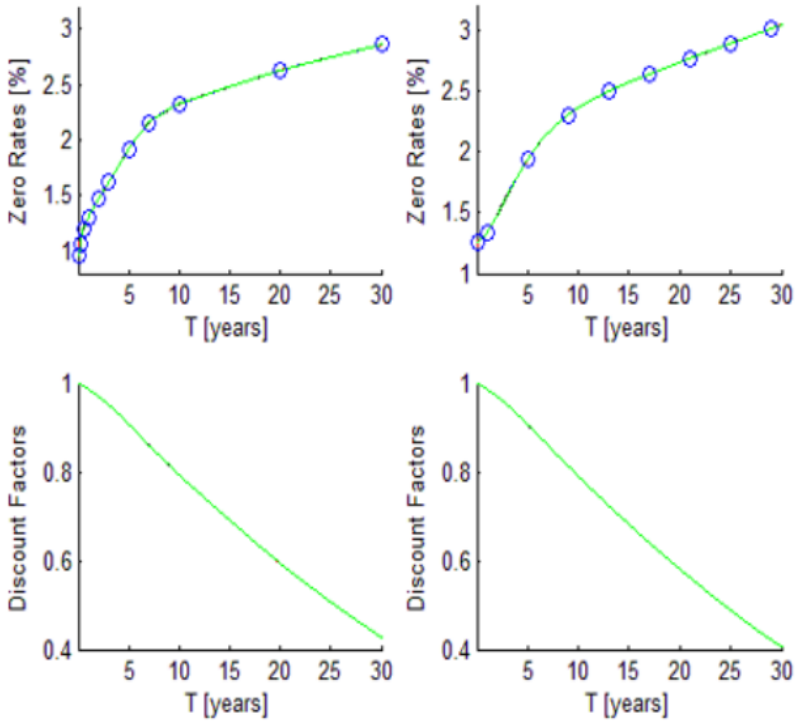

Figure 1.12: Interest rate fitting model via machine learning methods

Bezdek et al. (1983) suggest, in the excellent monograph of Duda and Hart (1973) many algorithms are discussed, each with its own mathematical clustering criterion for identifying "optimal" clusters. However, two popular techniques of unsupervised learning for clustering are k-means for clustering problems and fuzzy c-means clustering, in which each data point is allowed to belong to more than one cluster. K-means the clustering algorithm is the simplest unsupervised learning algorithm for this aim. It partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster. Fuzzy C-means clustering (J. C. Dunn, 1973, J.C. Bezdek 1981) is very similar to k-means technique and it is very used in pattern recognition. This procedure at first chooses a number of clusters, then randomly assigns

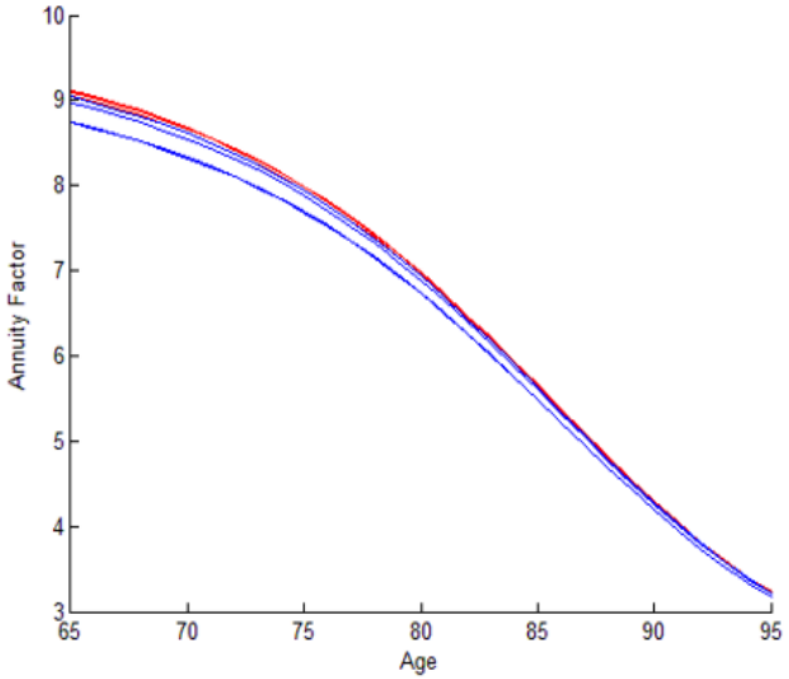

Figure 1.13: Annuity factor curves

coefficients to each data point for being in the clusters. These steps are repeated until the coefficients' change between two iterations is no more than ϵ, the given sensitivity threshold; the algorithm has thus converged. Finally, it computes the centroid for each cluster and the coefficients for each data point for belonging to that cluster. For each data point, compute its coefficients of being in the clusters. We have applied one of these automatic methods for organizing and clustering the information observed on secondary markets, focusing the attention on the recognition of potential anomalies, which may be potential trading opportunities. The methodology we used is constituted by a non-supervised neural network, designed to cluster the data, known as Kohonen Network or Self-Organizing Map (SOM). The aim of this study is to implement an automatic methodology which is able to organize and cluster information registered on secondary markets, in order to identify potential anomalies. A relevant problem for trading activities, in fact, is

to determine potential opportunities of arbitrage. Several traders quotes different prices associated to every single asset, subjecting the analysis of the comparison of the most convenient price to the single trader, who also has to consider the associated volumes. We suggest a methodology that may be considered a good tool to highlight interesting anomalies in price and, consequently, interesting trading opportunities. We use as algorithm to detect these potential trading opportunities a non-supervised neural network, aimed at data clustering, known as Kohonen Network or Self-Organizing Map (SOM). This kind of neural network has the advantage of not requiring to specify a-priori the number of partitions to implement, differently from other clustering methodologies, like K-means and Fuzzy C-means.

1.3.1 How a Self-Organizing Map works

During the Eighties, Teuvo Kohonen, developed the self-organizing feature maps, known in literature as self-organizing maps or SOM (T. Kohonen, 1982; 1989; 1998), while he was studying the biological functioning of how neural networks do organize and cluster the acquired information. The result of this construction is an Artificial Neural Network (ANN), characterized by the property that the ouput shows how the network works, that is how it organizes itself, learning completely in a non-supervised way. Differently from other types of ANN, as the feed-forward neural networks that are very good for fitting and data fitting, the SOM doesnt require to know which value the neuron calculates, but only which neurons actively intervene on the network. The analogy is very close to biology: in nature, neurons are connected with some muscles, and the first problem of our nerve centre is to know which muscle is active. In other words, we are not just interested about the correct output provided by the neuron, but also about understanding which neuron gives us the requested output. Neuro-scientists consider by far the

working principle of the SOM more linked to biology than to other paradigms inspired by human brain.

Moving to the **architecture** of a SOM, typically, the task of a SOM is to organize a complex input, characterized by high dimensions (N) in spatial areas, partitioned by grids, with lower dimensions (G dimensional) than the first one. This process of mapping allows to generate a representative map of the context of interest, highlighting the most relevant characteristics and reducing its complexity. ($G << N$). In order to generate the map, the SOM at first allocates the dots inside the N-dimensional reference vector space. During the training phase, the dots that compose the SOM change their position in the vector space and, consequently, the topology of the network, with the aim of recovering at best the positions taken by the initial dots perceived by its neurons. Going ahead with the analogy to nature, what we have just explained means that every neuron may be activated to interpret a precise position in the external environment: the greater the stimuli that come from a certain area, the greater will be the number of neurons activated to codify the perceived signal. This mechanism allows the brain to isolate the interesting areas in which the dots of the grid, which are neurons assigned by the coding of information, are more thickened with respect to the more marginal areas, in which less perceptron will be assigned. On short, we can say that we have two different areas in which SOM operates:

- The N- dimensional input space;
- The G- dimensional grid. Here the neurons that interpret the relationships of proximity that depend on the signal received from the environment (i.e. the spatial disposition of the knots within the input space) are disposed.

By modifying the neighborhood relationships we determine the network topology. In a monodimensional grid, the starting neurons are generally arranged in an

equi-distant way on a segment. On the other hand, a bidimensional grid may be broadly designed according to three different typologies:

- rectangular (the links between adjacent neurons form right angles that form a regular rectangular grid);
- exagonal (the links between adjacent neurons form sixty degrees angles that form a grid composed by equilateral triangles);
- random (the dots of the network are randomly generated within the vector space input).

The most common grids are the first two typologies, since the starting configuration is cleaner for the initial construction. In figure 14 the regularity in partitioning is shown: on the left hand side, a rectangular grid is represented (also known as matricial grid), on the right an exagonal grid (or beehive grid).

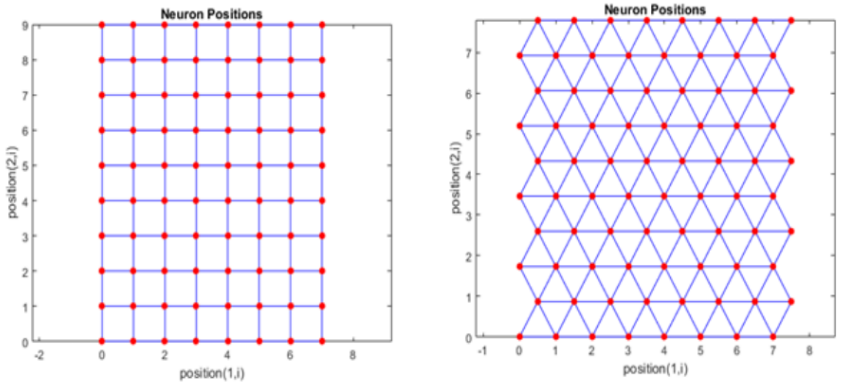

Figure 1.14: The most widespread network topologies for the architecture of a SOM

Topologies characterised by higher dimensions may be managed by a SOM, but they are not very used in practice because it is hard to intuitively view and interpret the complexity of the relationships of proximity. Similarly to the neurons of a RBF (Radial Basis Function) network, the basic unit of computation

of a Kohonen network (*SOM neurons*), k, does not recover a fixed position, c_k in the input space. A common characteristic of neural networks and, consequently, of SOMs (Calligaris, 2007), is the training process of their neurons, which is described by the following steps:

- Input: an arbitrary value p of the input space R^N;
- Calculation of the euclidean distance between every neuron k and p , that is the estimation of: $\|p - c_k\|$
- Activation of a neuron: the centre of the computational unit with the distance closer to the input, called i , becomes active, while all the others are inactive. This way of proceeding is called winner-takes-all scheme.
- Output: the procedure highlights which neuron becomes active.

1.3.2 The training algorithm of a SOM

The training of a SOM allows the map to correctly explore the domain; for this reason, it is considered the heart of techniques (Kriesel, 2005). Given its importance, it is good to deepen its procedural steps:

- Initialization. The network randomly distributes the centres $c_k \in R^N$;
- Creation of a pattern. A (stimulus), that is a dot p, is selected within the input vector space R^N;
- Measuring the distance. For every neuron k of the network, the euclidean distance is calculated: $\|p - c_k\|$;
- Winner-takes-all scheme. We determine the winner neuron i , that is the one closer to p. It thus must satisfy the following condition: $\|p - c_i\| \leq \|p - c_k\|$

- Adaptation of the centres The centres c_k move towards their reference vector space according to the *learning rule*:

$$\Delta c_k = \eta(t) \cdot h(i,k,t) \cdot (p - c_k) \tag{1.12}$$

$$c_k(t+1) = c_k(t) + \Delta c_k(t) \tag{1.13}$$

Where the variation $\Delta c_k(t)$ is just added to the current position of the centres c_k.

The last term of the equation shows that the change of neurons k position is proportional both to the distance p and to the time-variant learning coefficient $\eta_(t)$.

The topology of the net exerts its influence according the function $h(i,k,t)$, known as *topology function*: it is defined on the grid and it must be unimodal, this means it must have only one maximum, which is closer to the winner neuron i. One of the most common choices is to model h as a gaussian bell, since it is unimodal and its maximum is close to zero. In addition to that, its amplitude may be modified by the parameter σ, which may be used to reduce the distance with the passing of time: it is sufficient to link the timing dependence with σ, obtaining as final result a monotonic decreasing function, $\sigma_(t)$. The following topology function has the characteristics described above; for this reason, it is the most widespread in literature:

$$h(i,k,t) = \exp\left(-\frac{\|g_i - c_k\|^2}{2\sigma(t)^2}\right) \tag{1.14}$$

Where $\|g_i - c_k\|^2$ is the distance between the centre c_k and the i-th points on the grid. The learning rate is a decreasing monotonic function: it generally

assumes values that depend on the topology of the net and it generally belongs to the interval $0.01 \leq \eta \leq 0.6$ This is a necessary condition to a correct training of Kohonen networks, in addition to the check of the following inequalities: $hn \leq 1$

1.3.3 Validation of the algorithm on simulated data

We have written the SOM by using the numerical software Matlab. Before applying it to a real case, it is necessary to apply and validate the code in order to be confident about a correct functioning of the library developed. To this aim, we have implemented two tests, to check the correct fit of the artificial neural network to experimental data (Beale, Hagan and Demuth, 2014). We have randomly generated 5000 dots in the domain $[0; 1]^2$ and, starting from a SOM with an initial grid made by 100 knots randomly distributed (Random Network Topology) in a lager interval $[-0.5; 1.5]^2$, we have checked the correct mapping of simulated data. (Figura 15) (Beale, Hagan and Demuth, 2014) The expected result is that, iteration after iteration, the initial grid will slowly recover the more dots as possible in the space in which the generated dots have been placed. These dots have been created using a uniform distribution, so the map must not recognize any relevant pattern. (Figura 16). In other words, the expected result is that the SOM, by putting its neurons in the unitary domain, will uniformly recover it.

The figures show the correct training of the non-supervised neural network and, thus, the correct disposition of its neurons within the experimental domain.

The SOM can also be trained using a **gaussian distribution**. We have randomly generated 5000 dots spread in the domain according to a standard normal distribution, that is a gaussian with zero mean and unitary variance. The SOM we used for this test has an hex-grid topology, composed by 72 knots. (Beale, Hagan and Demuth, 2014)

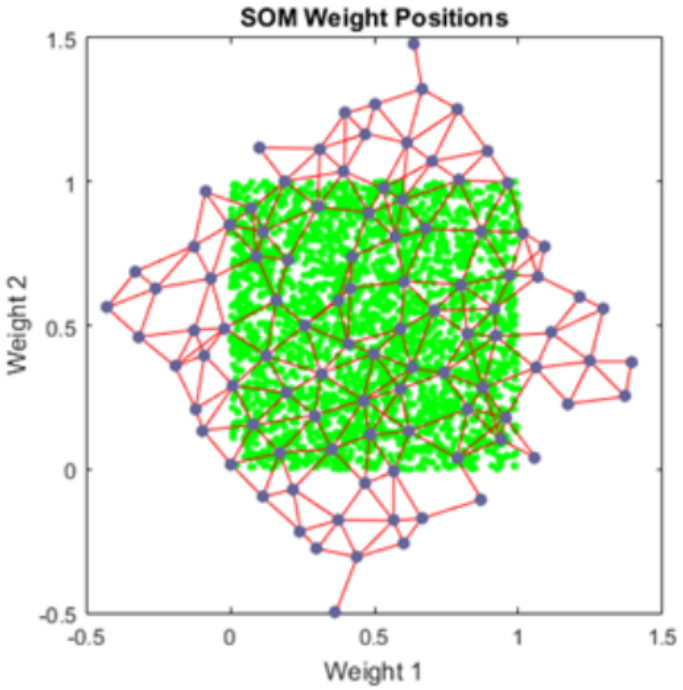

Figure 1.15: Initial disposition of dots and ok the knots of the net for the test shown in paragraph 3.1

After 5.000 epoches the SOM has correctly mapped the experimental domain, focusing its attention on a circumference of center (0;0) with a radius three times longer than the standard deviation (figure 17). We can thus conclude the algorithm has been able to detect the outliers, which are identified as the points outside the circumference with the center in the origin and the radius greater than 3σ.

1.3.4 Application of the soft-computing methodology in the financial field

After the theoretical validation of the model, we apply the Self-Organizing Maps as instrument for the automatic detection of the potential market anomalies. In

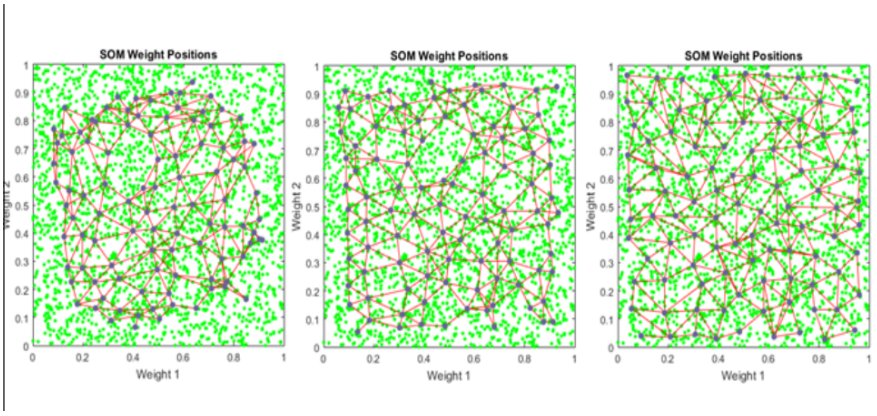

Figure 1.16: Learning of the SOM after 10, 100 e 1000 epoches respectively

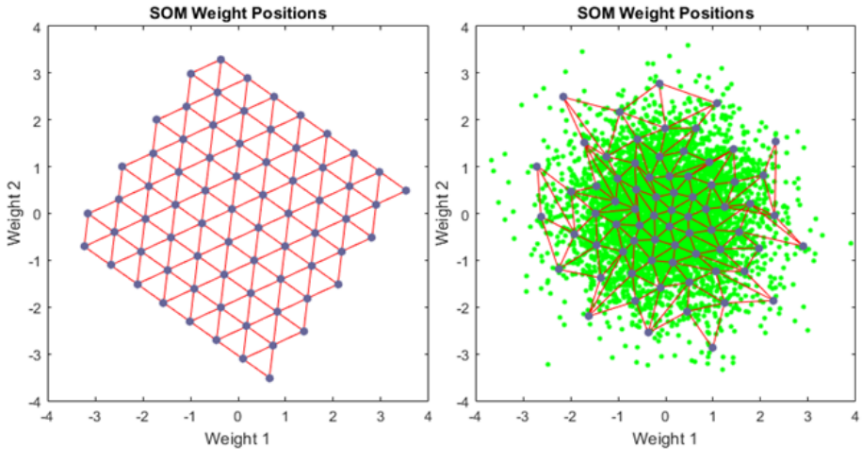

Figure 1.17: Initial and final of neurons constituting the SOM grid (Training of 5000 epoches

this context, the methodology has been applied in the secondary market of fixed-income securities: every transaction is characterized by a quantity to be sold or bought and a price at which the trader is willing to close the deal.

Putting these two fundamental variables on a Cartesian plane, the non-supervised neural network is able to separate the areas in which there is a greater density of points from the others.

Therefore the methodology can discriminate the market anomalies (i.e. prices and volumes higher or lower compared to the standard levels) treating them as

outliers in the reference plane: all the points not included inside the network could be potential financial anomalies, for which the automatic procedure is able to signal to the trader.

We apply the methodology in relation to two bonds listed on the market on 19th July 2016: the first one has a coupon of 1 3/8% and matures in 2026, while the second one has a coupon of 2 7/8% and matures in 2020. It is worth pointing out that the choice of the number of knots in the map (G) is an input for the algorithm, while the price/volume observations (N) are function of the available number of the market quotes. The greater the number of nodes, the smaller the ability of the network to isolate the outliers is. In fact if $N = G$, the knots of the map will allocate exactly on the points generated by the market observations.

On the other hand, a map with a very low number of neurons could give too many warnings. Experimentally, a good trade-off could be given by using $G = N7l$ with $l \in [3, 5]$. In these cases implementing a deterministic network can be useful for maintaining a good level of traceability and reproducibility of the intermediary step: in particular, we adopt a hexagonal map. Furthermore, the training phase has been carried out using 1000 epochs: this value represents a good compromise between the reliability of the results and computational time spent by the Kohonen map on updating the position of knots in function of the financial context.

1.3.5 Price/Volume anomalies detection for the first fixed-income instrument

Starting from a hexagonal map made by 9 knots, the algorithm proceeds to train the network with the aim of identifying the potential outliers in the reference plane (i.e. price/volume). Market data have been retrieved from the All Quotes (ALLQ) Bloomberg function. Figure 18 shows the Self-Organizing Maps (SOM)

before (on the left) and after (on the right) the training in the plane [Quantity (in millions of Euro); Price (expressed as a percentage)].

Blue points are the neurons of the ANN ($G = 9$), green points are the experimental market data ($N = 36$), red lines determine the spatial partitions outlined by the network. Points out of the map can be considered outliers and consequently potential market anomalies.

For instance, the procedure was able to highlight that for the same quantity of 1 million the security has received trading prices significantly different from the average (106.4% and quotes higher than 107.7%). The map has correctly excluded prices that have a volume equals to 2 millions, given that the centroid of the system is in the range between 0.5 and 1 million.

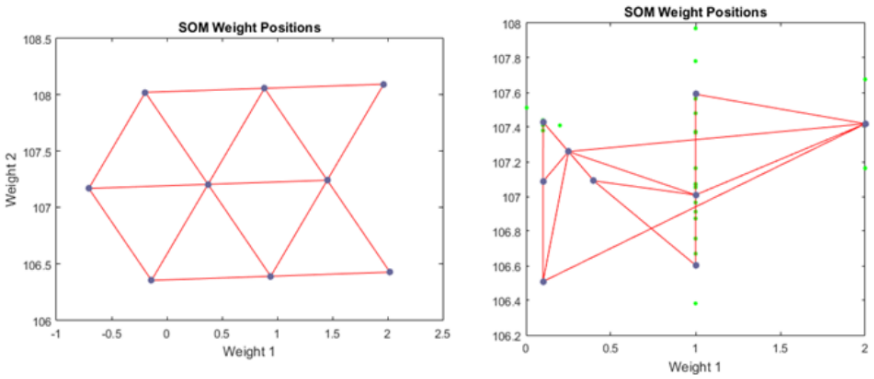

Figure 1.18: Detection of market anomalies for the first security

Similarly to the previous case, we implement a SOM which has a hexagonal topology with a number of neurons equal to $G = 9$ for $N = 42$ of experimental data (quotes). Figure 19 shows the right clustering of the reference space obtained after a training of 1000 epochs. Graphic conventions and unit of measures of the financial data are the same as the previous analysis. The spatial partition obtained by the SOM is able to find potential outliers: the two exchange proposals with a volume higher than one million and the deals with a quote higher than

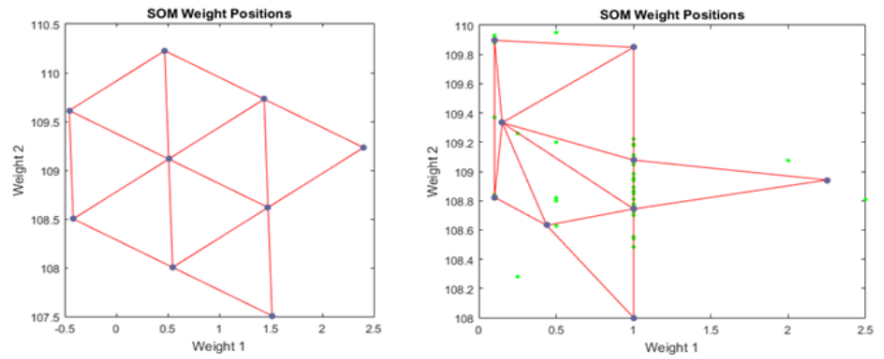

Figure 1.19: Detection of market anomalies for the second security

109.9% and lower than 108.4% at an amount lower than or equal to half a million.

1.4 Conclusions

In this chapter I have presented three applications of machine learning in financial markets, showing how they may be used for fitting and for clustering. In the first work I performed a comparison between parametric techniques and RBF networks to fit the yield curve under conditions of stress. The issue is of actual interest in this challenging time of high volatility and negative interest rates, because the yield curve is an important tool in finance, especially for assets pricing, financial risk management and portfolio allocation. We therefore investigated the capability of various methods to interpolate the yield curve under such extreme conditions of instability; to such aim, we considered the Euro Swap Euribor (EUR003M), and the USDollar Swap (USD003M) curves, on two different release dates (December 30th 2004 and 2016), corresponding to very different market situations, and we examined the various ability of the abovecited methods in fitting them. The results confirm that RBF nets can reach very satisfying results to manage anomalies such as extreme humps or negative interest rates. Besides, our

opinion is that the results pave the way to a research trail more focused on the use of Machine Learning methods to provide a integrated model of insample fitting and forecasting. In the second work we compared the performances of different techniques in modelling the yield curve and the discrepancies that might arise from using alternative interest rate databases, namely the DoT and the FRB. We then applied our analysis to life annuity factors, which are of extreme importance for the pension systems, since this topic includes relevant practical aspects of all the financial systemic risks, that is the main component in life contracts evaluations. By simulating all the categories of model that we considered on the two different time series, we found out that significant discrepancies on the estimation of temporary life annuities arise and that these discrepancies are also caused by a model risk. The results also show that the machine learning techniques gave the most satisfactory results in modeling the interest rate term structure. The third application I have presented is a cluster algorithm applied to secondary bond markets. We have used a Self Organizing Map and it has proved to be an efficient methodology both from a theoretical and a practical point of view. The Kohonen networks, differently from other clustering algorithms as K-means and Fuzzy C-means, do not need to have an aprioristic specification of the number of cluster to use. Thanks to this feature, SOMs are a flexible tool for the organization of financial information. It will be very useful to apply this methodology in other financial sectors in which the correct clustering of information is of fundamental importance, as behavioural finance, experimental finance and algo-trading systems. networks to fit the yield curve under conditions of stress. The issue is of actual interest in this challenging time of high volatility and negative interest rates, because the yield curve is an important tool in finance, especially for assets pricing, financial risk management and portfolio allocation. We therefore investigated the capability of various methods to interpolate the yield curve under

such extreme conditions of instability; to such aim, we considered the Euro Swap Euribor (EUR003M), and the USDollar Swap (USD003M) curves, on two different release dates (December 30th 2004 and 2016), corresponding to very different market situations, and we examined the various ability of the abovecited methods in fitting them. The results confirm that RBF nets can reach very satisfying results to manage anomalies such as extreme humps or negative interest rates. Besides, our opinion is that the results pave the way to a research trail more focused on the use of Machine Learning methods to provide a integrated model of insample fitting and forecasting. In the second work we compared the performances of different techniques in modelling the yield curve and the discrepancies that might arise from using alternative interest rate databases, namely the DoT and the FRB. We then applied our analysis to life annuity factors, which are of extreme importance for the pension systems, since this topic includes relevant practical aspects of all the financial systemic risks, that is the main component in life contracts evaluations. By simulating all the categories of model that we considered on the two different time series, we found out that significant discrepancies on the estimation of temporary life annuities arise and that these discrepancies are also caused by a model risk. The results also show that the machine learning techniques gave the most satisfactory results in modeling the interest rate term structure. The third application I have presented is a cluster algorithm applied to secondary bond markets. We have used a Self Organizing Map and it has proved to be an efficient methodology both from a theoretical and a practical point of view. The Kohonen networks, differently from other clustering algorithms as K-means and Fuzzy C-means, do not need to have an aprioristic specification of the number of cluster to use. Thanks to this feature, SOMs are a flexible tool for the organization of financial information. It will be very useful to apply this methodology in other financial sectors in which the correct clustering of informa-

tion is of fundamental importance, as behavioural finance, experimental finance and algo-trading systems.

Chapter 2

A financial market model with confirmation bias

We simulate a theoretical model in which different groups of traders coexist. In particular, we focus our attention on the role of confirmation bias during the formation of heterogeneous expectations. We are in a market in which one single risky asset can be traded. We have a group of fundamentalist traders, who believe that the price of the asset will follow its fundamental value and two groups of chartist traders, who have different expectations. They then form the expected price in two different ways, according to their initial guess: optimistic or pessimistic. We have expressed the demand form of the fundamentalists as a cubic function, as in Day and Huang (1990), to introduce an increasing level of aggressiveness of the fundamentalists as the price deviates from the fundamental value, while the two groups of chartist traders anchor their future expectations on the present ones. However, they adjust their expectations through social comparison, since they are affected by confirmation bias. Confirmation bias is a psychological phenomenon that describes the attitude of people to seek or to interpret the existing evidences in order to confirm their prior beliefs or expectations. This

term has been used in the existing psychological literature to define a wide variety of phenomena, but in financial markets it refers to the unaware habit to select evidences, or to give more weight to information that support one position to go to the detriment of another one. Several studies have provided theoretical and experimental evidences and various researches have shown that this bias also works in practical situations. Confirmation bias then involves people who wish to defend beliefs that they are willing to preserve. We have measured the strength of confirmation bias to understand its importance on the formation of fluctuations in the price pattern,concluding that the more the strength of confirmation bias increases, the more the different expectations are far. Using a mixture of theoretical and numerical tools, we have also found out that when the confirmation bias is strongly present, the price of the asset is far from its fundamental value for longer periods, leading these fluctuations to have an impact on financial markets. Our analysis has been computed through Montecarlo simulations over 1000 periods.

2.1 Introduction

> "*Once a human intellect has adopted an opinion (either as something it likes or as something generally accepted), it draws everything else in to confirm and support it*" Lord Francis Bacon (1620)

This quote from Lord Francis Bacon is an excellent definition of what nowadays psychologists call *Confirmation (or Confirmatory) bias*. Human beings display the tendency to select the information they receive and to give more importance to new evidences confirming their prior beliefs and less (or even nothing at all) to those against them. This is an error in the updating of beliefs in the light of new evidences, that should respect the so-called Bayesian updating to

be considered rational. Initial beliefs should only be a starting point, becoming less and less relevant as the number of evidences accumulates (a process known as washing-out the priors). Instead, according to the definition of confirmation bias, our first impressions will influence how we update our beliefs about some uncertain thing, distorting our reasoning process.

The first attempt to collect and review evidence of confirmation bias in several contexts can be probably found in Nickerson (1998). He identified confirmation bias in medicine, judicial reasoning and other fields, but he did not talk about finance and about its influence in the formation of asset prices. Quite surprisingly, only recently confirmation bias has been empirically found in financial markets, for instance by Park et al. (2010) in the field (in South Korea) or by Bìsiere et al. (2014) in the lab. The role of such a bias in fact in the persistence of prices different from their fundamental value is intuitive. The more an asset price increases, the more bullish traders are reinforced in their beliefs, making their assumptions correct with their own behavior. The same happens for bearish traders when prices drop. Confirmation bias may thus play a role in the emergence and/or other features (duration, frequency, etc...) of speculative bubbles and other important stylized facts in financial markets.

Besides experiments and empirical evidences of confirmation bias, there is also a branch of research dealing with the formalization of a beliefs updating consistent with this bias. The most known attempt is the one made by Rabin and Schrag (1999), who corrected the classical Bayesian updating mechanism, proving that under certain circumstances almost all agents may come to believe in a wrong hypothesis. Among the huge amount of follow-up papers, only recently Pouget et al. (2017) adapted and applied to financial markets the Rabin and Schrag mechanism, replicating several stylized facts such as excess volatility, excess volume and momentum. In addition to them, also Charness and Dave (2017)

adapted the Rabin and Schrag mechanism to test background strategies to correct or mitigate the negative effects of the bias, and formerly Bowden (2015) to an agent-based framework, finding that the bias may have ambiguous consequences on the volatility and kurtosis. To our knowledge the only paper where the confirmation bias is modeled in an alternative way with respect to the one proposed by Rabin and Schrag is Aldashev et al. (2011), who introduce confirmation bias in the way social interaction takes place, and combine it with adaptive expectations, showing that the two biases may both increase or decrease informational efficiency.

A first goal of the present work is to build a simple financial market model where traders (or a portion of them) are endowed with confirmation bias in order to better understand how this bias influences the characteristics of financial bubbles and other stylized facts. In this sense the work is similar to Bowden (2015) but differently from him, we introduce a different, simpler, mechanism to insert confirmation bias, that leads to a piecewise definition of the dynamical system regulating the time evolution of asset price and beliefs. The model is at least partially analytically tractable in its fully deterministic version while by introducing some noise we replicate important features of real financial markets.

Our model is inserted into the so-called Heterogeneous Agents Models (HAM) literature, that has proved to be a quite good framework to introduce and study the effects of the behavioral biases of traders (see Chiarella et al. 2009, Hommes and Wagener 2009, Lux 2009 and Westerhoff 2009 for complete surveys). Usually the interactions between fundamental and technical trading rules are able to produce the features of financial markets that are hardly reconcilable with the assumption of rationality of traders and efficiency of the markets. Both laboratory experiments (Hommes, 2011) and other empirical evidences (like those surveyed by Menkhoff and Taylor, 2007) support the hypothesis that traders rely on simple

rules. Starting from the pioneering contribution of Day & Huang (1990), several papers have proved that complicated price dynamics can be obtained by simple deterministic models (see for instance Kirman 1991, De Grauwe et al. 1993, Lux 1995, Brock and Hommes 1998, Chiarella et al. 2002 and Westerhoff 2004). In this work we move from the subdivision of technical traders in pessimistic (or bearish) and optimistic (or bullish), similarly to Lux (1995, 1998) and Lux and Marchesi (1999, 2000), and we introduce confirmation bias in the way they interpret and use current information. A positive trend will be considered by bullish traders and ignored by pessimistic ones (if it is not too positive) and, at the opposite, a negative price trend will be considered by pessimistic traders and ignored by optimistic traders (if it is not too negative). The dynamical systems arising in HAM are usually smooth, but nonlinear. The nonlinearity may originate from the trading rule of the traders or from the switching mechanism from one strategy to another one, in evolutive models. Only a few papers in the HAM literature deal with discontinuous (or at least not differentiable) dynamical systems. Among them we have Huang and Day 1993, Huang et al. 2010, Tramontana et al. 2010, 2011. Our model is discontinuous because of the intrinsic dichotomy of the confirmation bias itself. Traders affected by this bias display a sort of cognitive dissonance, alternating a certain behavior when they face evidence supporting their current hypothesis and a totally different one when dealing with evidence against what they actually believe.

In order to mimic more qualitative features of financial markets (such as excess volatility, fat tails, volatility clustering, etc...) some researchers study a stochastic version of a deterministic HAM, obtaining quite interesting results (see Lux and Marchesi 1999, Westerhoff and Dieci 2006, Gaunersdorfer and Hommes 2007 or He and Li 2007). In the present work we also present a stochastic version of the model, replicating some features of real financial markets. So, a second goal of

the model is to try to identify the degree of confirmation bias that better permits to replicate such stylized facts.

2.2 The model

Our financial market model consists of a standard building block which formalizes the behaviour of a market maker and of three different groups of traders. One novelty we have introduced in our model is that, in addition to a group of *fundamentalist* traders, i.e. investors who believe that the price of the asset will follow its fundamental, we consider two different groups of *chartist* traders. These last two groups differ according to the way they form the expected price or, more precisely, according to their initial guess: they can be optimistic or pessimistic. As fundamentalists believe that the price of the asset will be close to its fundamental, they buy an undervalued asset and they sell an overvalued one. On the other hand, the other two groups of traders consider the difference between the current price and the expected price in the next period and they adjust their expectation through social comparison, affected by confirmation bias. We have thus studied the behaviour of the dynamic system regulating the asset price with a piecewise-defined map, which is helpful to understand and to test the different properties of our model.

2.2.1 Our model's building block

We start our model adopting a price adjustment mechanism with market maker, in a market with only a single risky asset. Following Day and Huang (1990), in our frame the first player we have is the market maker who, among other things,

quotes the price according to the following equation:

$$P_{t+1} = P_t + \alpha D_t$$

where D_t is the total excess demand at time t made up by the excess demands of the different kinds of traders. The primary function of the market maker, in fact, is to mediate transactions out of equilibrium, when demand exceeds the supply (as in this case), or vice versa. He acts by providing or absorbing liquidity, according to whether there is a positive or negative excess of demand. We have a classic group of traders called *fundamentalists*, who behave as follows:

$$D_t^f = f(F - P_t)^3$$

where $f \geq 0$ is their speed of reaction. They believe the price of the asset must follow its fundamental value, so they buy an asset when it is undervalued, (the price is below its fundamental value), driven by a higher probability of a capital gain, and they sell an overvalued one (the price is higher than the fundamental), because in this case it is more likely to incur in a capital loss. We also consider two different groups of *chartist* traders, who have different expectations about the future price of the asset: they can therefore be optimistic or pessimistic. In particular, they consider the difference between the current price and the expected price in the next period according to this rule:

$$D_t^c = c(EP_{t+1} - P_t)$$

Again, $c \geq 0$ is a positive reaction parameter. Optimists and pessimists adjust their expectations through social comparison, affected by confirmation bias. They thus tend to give different importance to facts that support their

initial guess rather than news that go in the opposite direction.

The first group of optimist traders think the price is going to increase:

$$EP_{t+1}^{opt} > P_t,$$

More precisely, they think the price is going to grow at an endogenous growth rate σ_{t+1}:

$$P_{t+1} = P_t \left(1 + \sigma_{t+1}\right)$$

They then adjust the expected growth rate by considering the last price variation $(\frac{P_t - P_{t-1}}{P_{t-1}})$ and according to three different rules.

Optimists Rule 1: If the last variation has been positive ($P_t > P_{t-1}$), so the initial prediction about the future price has been confirmed, the optimists adjust their beliefs as follows:

$$\sigma_{t+1} = \min\left[\beta_1 \sigma_t + (1 - \beta_1)\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right), 0.05\right]$$

where $0 \leq \beta_1 \leq 1$ measures the anchoring to the previous belief. We assume that increasing expectations cannot excess 5%.

Optimists Rule 2: When the last variation has been negative, but not excessively negative ($P_t < P_{t-1}$ and $P_t - P_{-1} \geq -w$, with $w > 0$), optimists basically ignore the signal and slightly adjust the belief of the previous period:

$$\sigma_{t+1} = k\sigma_t$$

with $0 << k < 1$.k is a minimum adjustment that occurs when the signal associated to the opposite event is not strong.

The positive parameter w measures the strength of the Confirmation Bias, so how much investors are anchored to their beliefs. This parameter is very

important for our model because it quantifies the role of confirmation bias on the formation of fluctuations in the price pattern. As we are going to demonstrate in section 2.2, the more the strength of confirmation bias increases, the more the different expectations are far.

Optimists Rule 3: Finally, if the last price variation has been negative and greater than the threshold w, the investors of this group reduce their growth expectations in this way:

$$\sigma_{t+1} = \max\left[\beta_2\sigma_t + (1-\beta_2)\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right), 0\right]$$

Here, as mentioned before, $0 \leq \beta_2 \leq 1$ measures the anchoring to the previous belief. It is greater than β_1 due to the loss aversion: the signal is in fact stronger when the market signal is opposite. We avoid a negative value of σ: optimists cannot radically change their initial expectations and become pessimists.

Pessimists, of course, have an original expectation that involves a decreasing price:

$$EP_{t+1}^{pes} < P_t,$$

Again, they think the price is going to drop at an endogenous growth rate γ_{t+1} as follows:

$$P_{t+1} = P_t\left(1 - \gamma_{t+1}\right)$$

. Similar to optimists, they adjust the expected growth rate by considering the last realized price variation, according to the following rules.

Pessimists Rule 1 If the last price variation has effectively been negative ($P_{t-1} > P_t$), the pessimists adjust their beliefs in this way:

$$\gamma_{t+1} = \min\left[\beta_1\gamma_t + (1-\beta_1)\left(\frac{P_{t-1} - P_t}{P_{t-1}}\right), 0.05\right]$$

Here, again $0 \leq \beta_1 \leq 1$ measures the anchoring to the previous belief and, again, negative expectations cannot cross the 5% level.

Pessimists Rule 2 If the last variation has been positive, but not excessively positive ($P_{t-1} < P_t$ and $P_t - P_{-1} \leq w$, with $w > 0$), pessimists ignore the signal and they adjust the belief of the previous period:

$$\gamma_{t+1} = k\gamma_t$$

The positive parameter w has the same meaning already underscored.

Pessimists Rule 3. If the last price variation has been positive and greater than the threshold w, pessimists cannot totally ignore the signal and then they adjust their negative expectations according to this rule:

$$\gamma_{t+1} = \max\left[\beta_2\gamma_t + (1-\beta_2)\left(\frac{P_{t-1}-P_t}{P_{t-1}}\right), 0\right]$$

Again, $0 \leq \beta_1 \leq \beta_2 \leq 1$ measures the anchoring to the previous belief. As in the previous case, it is stronger when the market signal is opposite (denoting some kind of loss aversion). We avoid a negative value of γ because pessimists cannot become optimists.

If we put everything together and we consider two potentially different reactivities (c_1 for optimists and c_2 for pessimists), we obtain the following dynamic

system regulating the asset price:

$$T: \begin{cases} P_{t+1} = P_t + \alpha \left[f(F - P_t)^3 + c_1 P_t \sigma_{t+1} - c_2 \gamma_{t+1} P_t \right] \\ z_{t+1} = P_t \\ \sigma_{t+1} = \begin{cases} \min \left[\beta_1 \sigma_t + (1-\beta_1) \left(\frac{P_t - z_t}{z_t} \right), 0.05 \right] & \text{if } P_t - z_t \geq 0 \\ k\sigma_t & \text{if } -w \leq P_t - z_t < 0 \\ \max \left[\beta_2 \sigma_t + (1-\beta_2) \left(\frac{P_t - z_t}{z_t} \right), 0 \right] & \text{if } P_t - z_t < -w \end{cases} \\ \gamma_{t+1} = \begin{cases} \min \left[\beta_1 \gamma_t + (1-\beta_1) \left(\frac{z_t - P_t}{z_t} \right), 0.05 \right] & \text{if } P_t - z_t \leq 0 \\ k\gamma_t & \text{if } 0 < P_t - z_t \leq w \\ \max \left[\beta_2 \gamma_t + (1-\beta_2) \left(\frac{z_t - P_t}{z_t} \right), 0 \right] & \text{if } P_t - z_t > w \end{cases} \end{cases} \tag{2.1}$$

We have introduced the auxiliary variable $z_t = P_{t-1}$ in order to have a system of difference equations of the first order, which allows us to manage the system more easily.

2.3 Study of the deterministic model

Map (2.1) regulates the dynamics of the asset price and of the beliefs of optimistic and pessimistic traders. The system admits only one equilibrium:

$$E(F, 0, 0, F) \tag{2.2}$$

At the equilibrium the asset price is coincident with the fundamental value and both kinds of technical traders believe that the price is not going to change. In other words, at the equilibrium the pessimists and optimists become equal, both convinced that the current price is right and will stay constant.

The stability properties of the equilibrium must be studied by taking into account the piecewise definition of the map and in particular the fact the equilibrium

is located at the border between two different dynamic regimes. So calculating the jacobian matrix in a neighborhood of the equilibium means to calculate two matrixes: one when there is a small increase (i.e. when $0 < P - z < w$) and another one with there is a decrease(i.e. when $-w < P - z < 0$).

In the first case dynamics are regulated by the following dynamical system:

$$\begin{cases} P_{t+1} = P_t + \alpha \left[f(F - P_t)^3 + c_1 \sigma_{t+1} P_t - c_2 \gamma_t P_t \right] \\ \sigma_{t+1} = \beta_1 \sigma_t + (1 - \beta_1) \left(\frac{P_t - z_t}{z_t} \right) \\ \gamma_{t+1} = \gamma_t \\ z_{t+1} = P_t \end{cases} \tag{2.3}$$

to which corresponds the following jacobian matrix calculated at the equilibrium:

$$J_1 : \begin{bmatrix} 1 + \alpha c_1 (1 - \beta_1) & \alpha c_1 \beta_1 F & -\alpha c_2 F & -\alpha c_1 (1 - \beta_1) \\ \frac{1-\beta_1}{F} & \beta_1 & 0 & -\frac{1-\beta_1}{F} \\ '0 & 0 & 1 & -0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.4}$$

The eigenvalues of J_1 are:

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, 1, 1, \alpha c_1 (1 - \beta_1) + \beta_1) \tag{2.5}$$

so it is a particular scenario where for any combination of the parameters two eigenvalues are exactly one, another one is zero, implying strong convergence on one dimension, while the last one, always positive, can be both lower and higher than one, in fact:

$$\lambda_4 > 1 \Leftrightarrow \alpha > \frac{1}{c_1} \tag{2.6}$$

By considering the second case, a slightly decreasing price, we have dynamics regulated by the system:

$$\begin{cases} P_{t+1} = P_t + \alpha \left[f(F - P_t)^3 + c_1 \sigma_t P_t - c_2 \gamma_{t+1} P_t \right] \\ \sigma_{t+1} = \sigma_t \\ \gamma_{t+1} = \beta_1 \gamma_t + (1 - \beta_1) \left(\frac{z_t - P_t}{z_t} \right) \\ z_{t+1} = P_t \end{cases} \tag{2.7}$$

and the corresponding jacobian matrix at the equilibrium:

$$J_2 : \begin{bmatrix} 1 + \alpha c_2 (1 - \beta_1) & \alpha c_1 F & -\alpha c_2 \beta_1 F & -\alpha c_2 (1 - \beta_1) \\ 0 & 1 & 0 & -0 \\ -\frac{1-\beta_1}{F} & 0 & \beta_1 & \frac{1-\beta_1}{F} \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.8}$$

and the eigenvalues:

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, 1, 1, \alpha c_2 (1 - \beta_1) + \beta_1) \tag{2.9}$$

so the scenario is not so different. Only the forth eigenvalue changes, with c_2 replacing c_1.

In such cases it is quite complicated to know which kind of dynamics are going to occur, and only numerical simulations may help to understand it.

Figure 2.1 represents a bifurcation diagram obtained by keeping all the parameters but α fixed. We used $c_1 = c_2 = 1$, $\beta_1 = 0.95$. $\beta_2 = 0.98$, $f = 0.2$, $F = 1$ and $w = 0$, that is with no confirmation bias.

We can see that the equilibrium is reached even for values of α higher than 1, that for our setting will cause a fourth eigenvalue higher than one. This may only occur due to the piecewise definition of the map in addition to the special case with two eigenvalues equal to one.

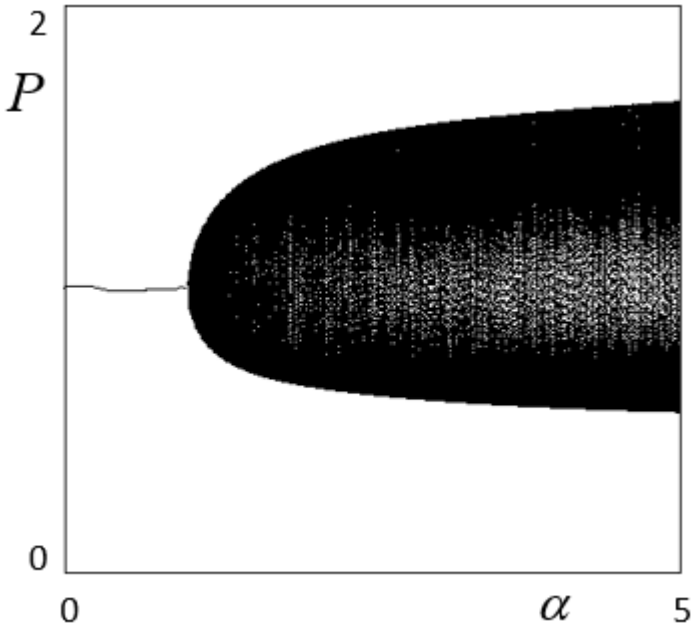

Figure 2.1: Bifurcation diagram

2.3.1 The role of confirmation bias

Let us check if confirmation bias plays the role we expect it to play: that is a failure in the bayesian updating. In order to do that, we numerically investigate the dynamic evolution of the beliefs of the two kinds of technical traders with and without confirmation bias, that is with two different values of w (Figure 2.2). We have also inserted a noise both in the fundamental value and in the asset price. In particular we add a stochastic term uniformly distributed which varies between -0.1 and +0.1 to both. In such a way it is possible to observe the reaction of optimists and pessimists to some exogenous shocks. In all the panels we used the same combination of parameter used in Figure 2.1, we fix $\alpha = 2$ and we vary the confirmation bias parameter w.

Panel (a) is obtained without confirmation bias ($w = 0$) . We used initial beliefs quite distant: ,$\sigma_0 = 0.045$ and $\gamma_0 = 0.047$. In other words, optimistic

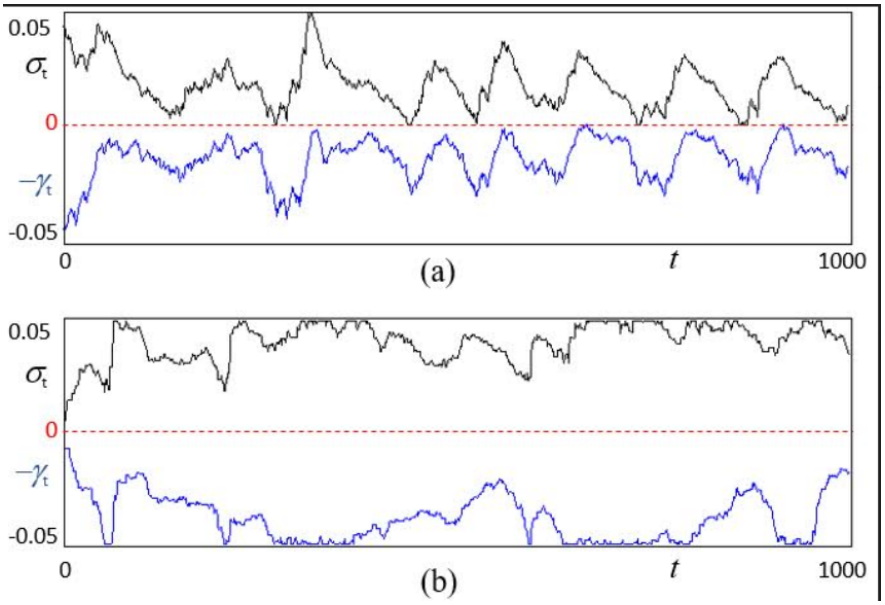

Figure 2.2: Confirmation bias and the dynamics of beliefs. Panel (a) is obtained without confirmation bias ($w = 0$), while panel (b) with a strong confirmation bias ($w = 0.1$).

traders expect an increase of the price of 4.5%, while pessimistic one a decrease of 4.7%. Nevertheless, the two beliefs become quite soon very close, meaning that optimistic traders reduce their bullish expectations and the same is true for pessimistic traders and their bearish beliefs. At the opposite, in panel (b) we assume a strong confirmation bias ($w = 0.1$). It means that optimists do not change their bullish expectations unless the price drops by more than 10% (and the same happens for pessimistic traders in the case of an increasing price). We start with quite close initial beliefs: we assume $\sigma_0 = 0.005$ and $\gamma_0 = 0.007$, that is optimistic traders think the price is going to rise by 0.5% while pessimistic traders believe in a decreasing of the 0.7%.

The dynamic evolution of the beliefs show that they tend to become more and more extreme, getting close to the most extreme value we admit, that is an expectation of a 5% increase or decrease for the two groups of traders. So they look at the same price patterns but confirmation bias prevents the convergence

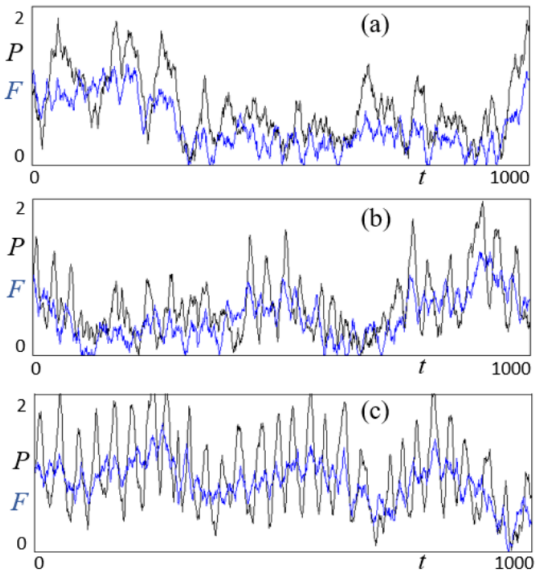

Figure 2.3: Anchoring and fluctuations. Panel (a) is obtained with $\beta_1 = 0.95$ and $\beta_2 = 0.98$. For panel (b) we used $\beta_1 = 0.85$ and $\beta_2 = 0.9$. Finally, for panel (c) we have $\beta_1 = 0.83$ and $\beta_2 = 0.87$.

of their beliefs and make them more and more distant.

2.3.2 The role of anchoring

It is also interesting to note the role played by parameters β_1 and β_2, anchoring the beliefs to their current value even when traders decide to adjust them.

We expect that the less traders are anchored to their previous beliefs, the more prices will vary as a consequence of their fastly changing strategies. In fact, this is what we get when we reduce the two parameters (Figure 2.3). Fluctuations are more and more accentuated with lower anchoring.

Figure 3.

Now that we get a qualitative view of the kind of dynamics occurring for price and beliefs by varying some relevant behavioral parameters, we can try to tune them in order to replicate some important stylized facts of financial markets,

trying to understand if confirmation bias may play a role in them.

2.4 Stochastic numerical simulations

In previous sections, we have discussed how much confirmation bias affects the formation and the fluctuations of price pattern. We have also shown that the more the strength of confirmation bias increases, the more the different expectations become far. Using a mixture of theoretical and numerical tools, we have also highlighted that when the confirmation bias is strongly present, the price of the asset is far from its fundamental value for longer periods, leading these fluctuations to have an impact on financial markets. However, in financial markets when bubbles occur other stylized facts, as high kurtosis, fat tails, uncorrelated returns and high volatility, are present. In this section, we focus our attention on the ability of our model to mimic these dynamics which characterize financial markets. However, our deterministic model only provides a stylized explanation of the dynamics mentioned above. To overcome this limit, we have introduced a stochastic component to the fundamental value of the risky asset, $r \sim N(0, 0.06)$.

As in Tramontana et al. (2015), we have determined the value of this parameter via a "trial and error" calibration approach, by varying the variable until we obtain satisfactory dynamics from the model.However, this choice has also an immediate economic interpretation: assets in real financial markets have an aleatory component, which can be caused for example by exogenous and unpredictable shocks that can affect the fundamental value of the stock. The stochastic component is thus able to replicate these random dynamics.

Figure 2.4 shows a representative simulation run of our stochastic model in presence of confirmation bias ($\omega = 0.1$). The underlying time series is 1000 periods long. By looking at the figure, it is easy to notice the statistical properties

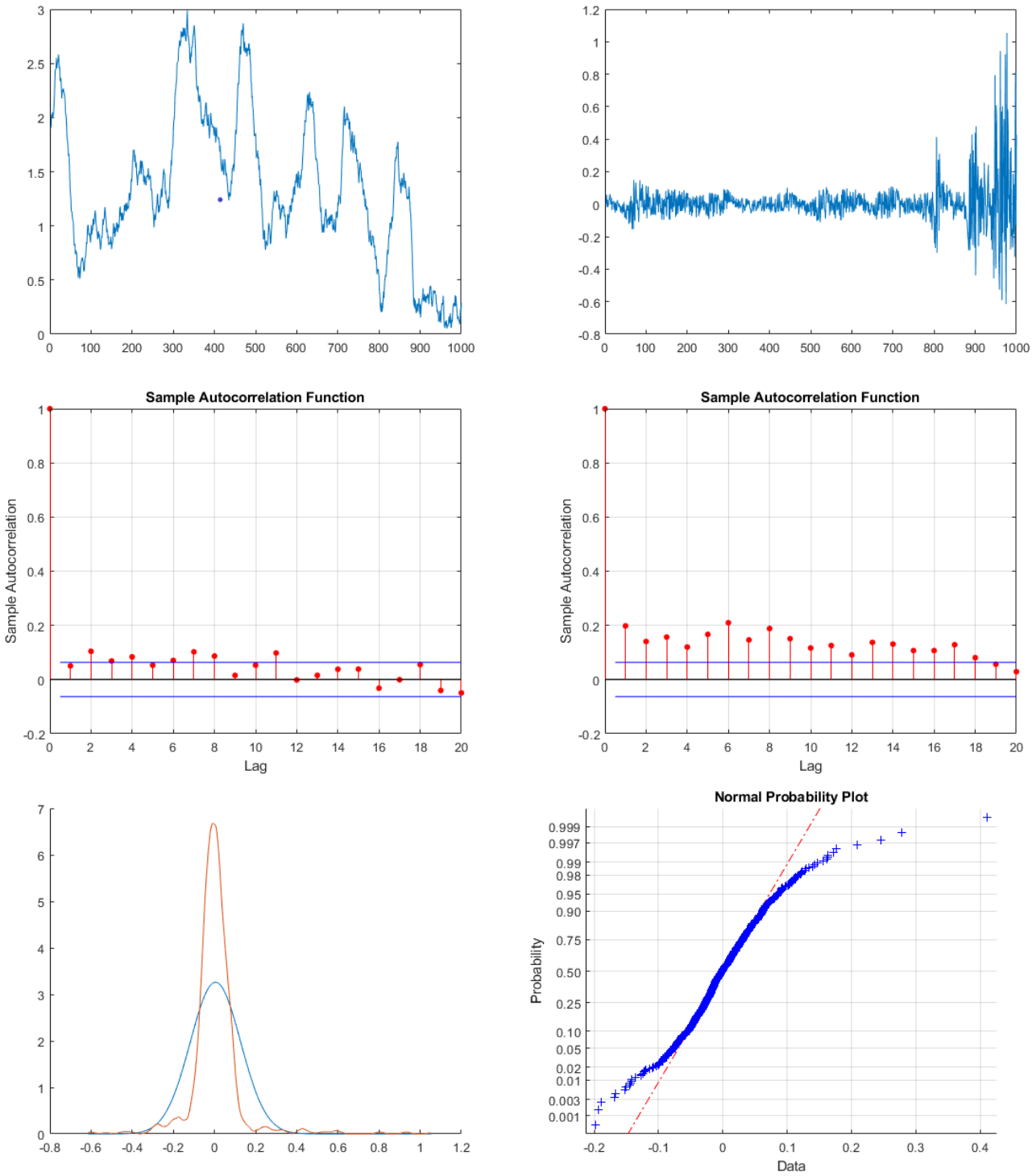

Figure 2.4: A simulation run of our stochastic model. The length of the time series is 1000 simulations. Parameter setting as in section 3, except for $F_0 = 2$. $\omega = 0.1$

characterizing financial markets. Looking at the first two panels, in which the price and the returns of the asset respectively are shown, it is clear that bubbles and crashes occur. An excess of volatility is also present. The price of the asset experiences high amplitudes, with strong price changes, which cause fat tails and a crash at the end of the series. The distribution of returns is asymmetric and left skewed. In the left central panels, the auto-correlation of returns is also shown. Almost all the coefficients of raw returns are insignificant, meaning that there is no auto-correlation between them: despite its boom and bust behaviour, our asset follows a random walk process. On the other hand, the coefficients of the absolute returns, depicted on the panel on the right, are highly significant for the first eighteen periods. We can interpret this result as indicating the succession between periods of low volatility and periods of high volatility. The two bottom panels show the distribution of returns. In the figure on the left, we can observe the theoretical normal distribution of returns (the blue bell) compared with the distribution of returns generated by our model (the red bell). It is evident that the distribution is unimodal, almost symmetric and bell shaped, but it also presents a high concentration around the mean. Finally, the bottom right panel contains the normal probability plot of returns, which evidences significant departures from normality and an S-shaped distribution.

To check if our model really does mimic the dynamics which characterizes real financial markets, we have compared the dynamics explained above with the behaviour of two important indexes: the FTSEMIB index (figure 2.5) and the Standard and Poor index (figure 2.6). The length of the series is different from our simulated series, but we have similar dynamics. In the panel on the left on the top of figure 2.5, we can see that the returns are affected by high volatility, especially in the first part of the series. The length considered is 3890 observations, from 2004 to 2018. The figure on the right contains the normal

probability plot of returns: it is evident that the distribution is far from normality, with several outliers in the tails. As in our simulated model, the returns are not auto-correlated, exceeding only twice the five percent level, while in absolute value the coefficients are always highly significant. Similarly in this context, we can deduce that the returns follow a random walk process. Finally, the bottom panel shows the comparison between the theoretical normal distribution of returns (in blue) and the effective distribution of returns of the FTSEMIB index (in red). As we can see from the red bell, in this case again the distribution is unimodal, almost symmetric and bell shaped with an high concentration around the mean.

A definitely similar behaviour belongs to the Standard and Poor index: for this series again, we considered 3890 observations, from 2004 to 2018. The returns are highly volatile for the first part of the series, and their S-shaped distribution is far from normality. The returns here as well are auto-correlated only in absolute values and their distribution is again bell-shaped, concentrated around the mean and slightly left skewed.

To sum up, our model is able to reproduce some important statistical features of financial markets. We are satisfied with this outcome because, despite it is a deterministic with several economic assumptions behind it, it is nonetheless able to match the data, also thanks to its stochastic component. Our goal in the future is to find other combinations of parameters that will make the analysis even more representative.

2.5 Conclusions

In this paper we have studied a simple financial market model in which three different groups of traders coexist. These groups differ in the way they form their expectations about the price of the asset. The map originated from the different

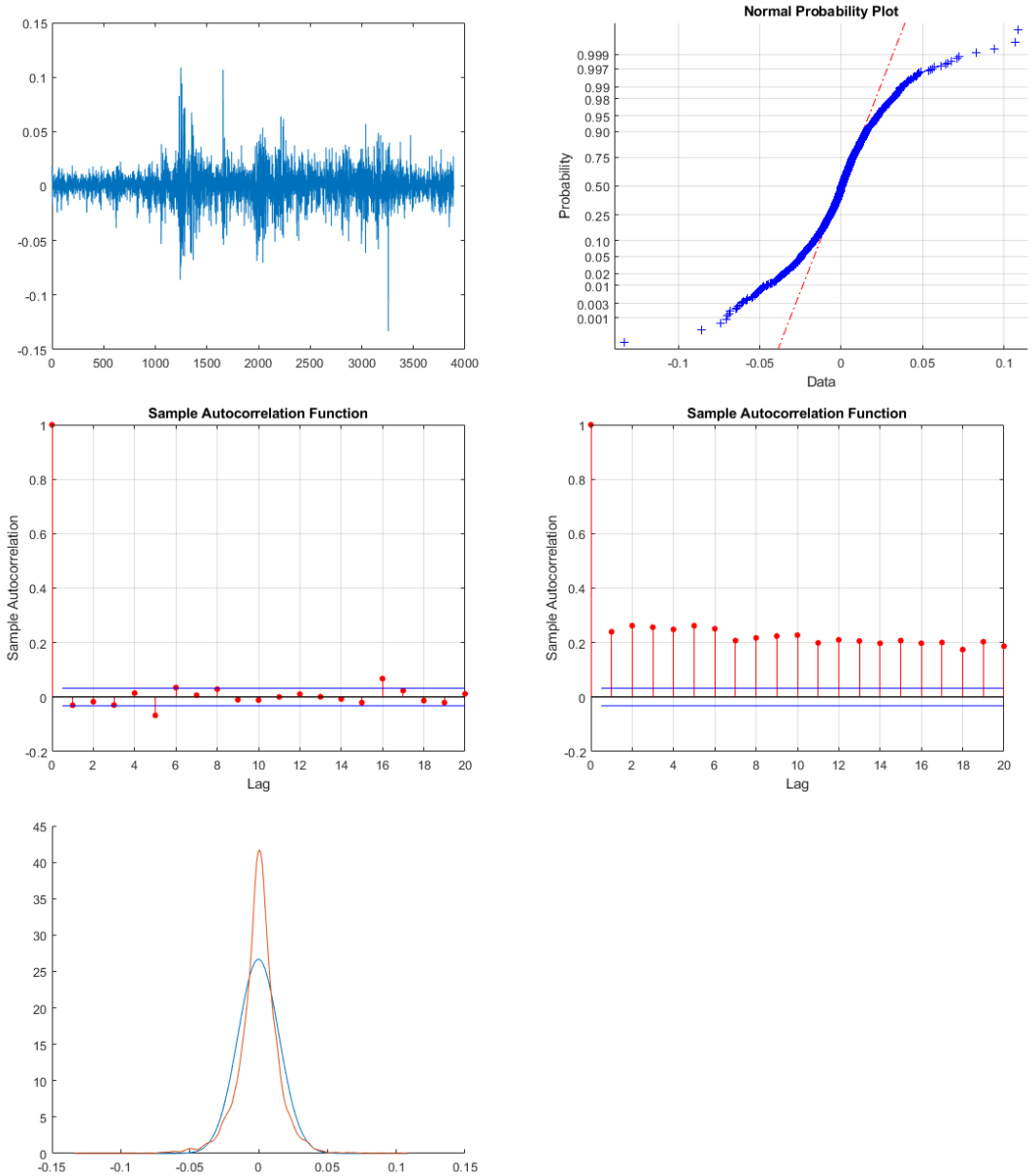

Figure 2.5: The dynamic of the FTSEMIB index. The underlying time series runs from 2004 to 2018 and contains 3890 observations.

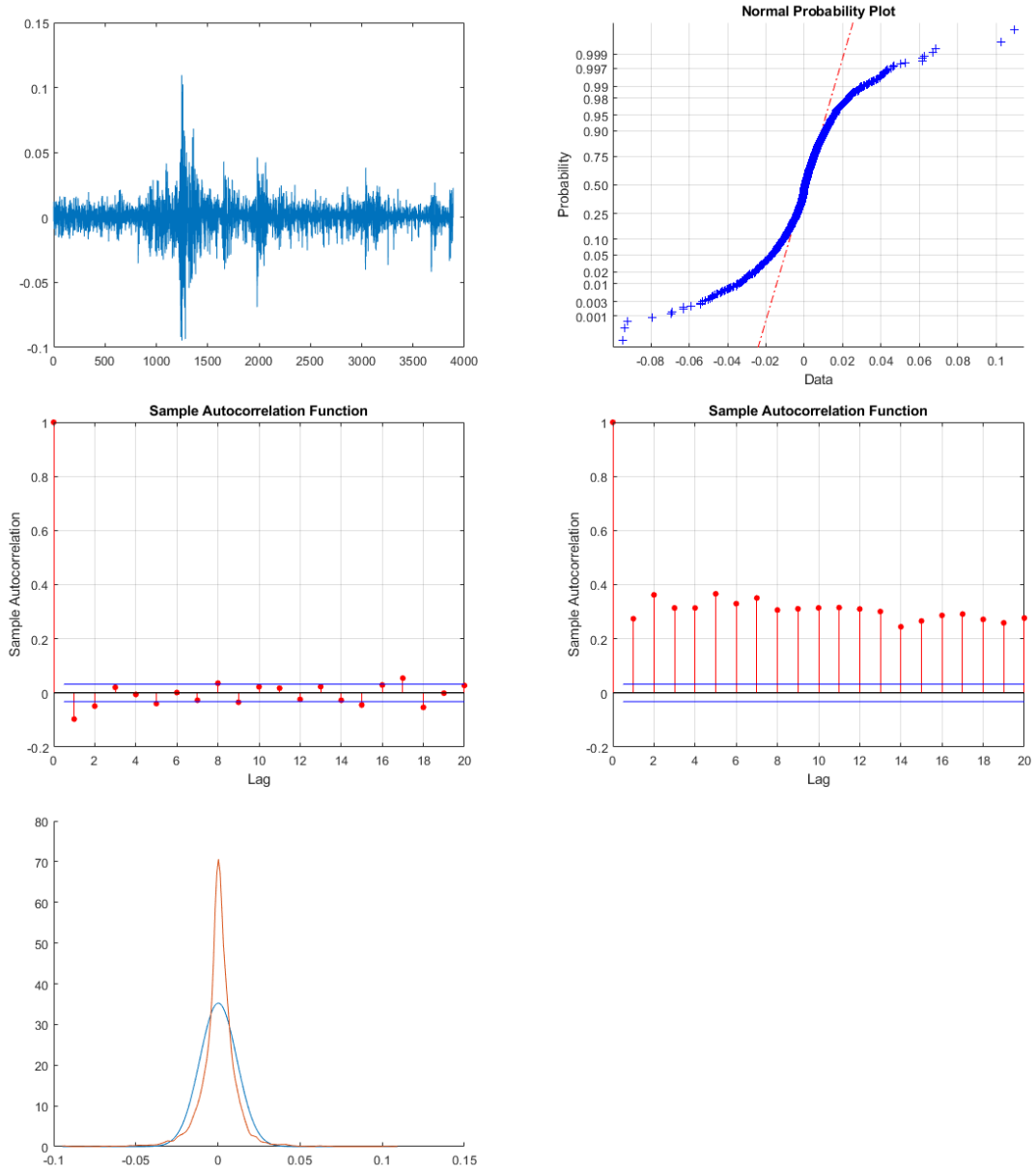

Figure 2.6: The dynamic of the SP index. The underlying time series runs from 2004 to 2018 and contains 3890 observations.

expectations is piecewise- defined and, at the equilibrium, the price of the asset coincides with its fundamental value. The different kinds of expectations differ according to the initial guess of traders about the future price of the asset: they may in fact believe that the price will increase, decrease or follow the fundamental value. Their expectations are affected by confirmation bias. From an economic perspective, we expected that this psychological phenomenon was caused by a failure in the bayesian updating. Our expectations have been confirmed by the simulation of our model, since our results were that confirmation bias makes the different beliefs farther from each others. As for the role of anchoring, we observed that the less traders are anchored to their previous beliefs, the more the prices vary, because of their fast changing strategies. Despite the fact that our model is simple, we found out that it is able to explain the dynamics present in financial markets, such as high volatility, fat tails in distribution of returns, unpredictable price patterns. We have also checked if the dynamics simulated by our model are representative of the behaviour of data- in particular of the FTSEMIB index and of the Standard and Poor index- and we noticed that it provides a good representation of reality. To bring our model to reality, we have introduced a stochastic component which describes some random events that take place in financial markets. Our model may be extended and improved in various ways: we can introduce switching strategies between optimists and pessimists, we can add a stochastic component also to the two reaction parameters of fundamentalists and chartists or we can assume to use a more complex demand function. Another possible extension is to think about possible policy implications to the dynamics we get from the model. About this, it is fundamental to design more precisely how financial markets work.

Chapter 3

The Effects of Negative Nominal Rates on the Pricing of American Calls: Some Theoretical and Numerical Insights

This last chapter investigates the effects caused on options pricing by negative risk-free rates when the underlying is an equity with null dividends. In such anomalous conditions, in fact, the fair value of the American Call would not match the value of the European Call with the same financial features. It has often been discussed about the reasons that invalidate Black pricing model, so in this study we aim to investigate the indirect consequences of these market conditions. In particular, we analyze why in this financial context the property according to which, given an American call written on an equity underlying that does not pay dividend, its early exercise is not optimum, often is not verified. Thus, from a practical point of view, the fair value of an American option is not equal to the one of a European call option having the same financial features, but

it can be considered just an approximation. We originally motivate this assumption with theoretical arguments. We then move to an empirical investigation where we put at work some quasi-closed formulas for pricing an American option and the stochastic trinomial trees, which are one of the most efficient numerical techniques for pricing an American option, written on an equity underlying. Finally, we implement and validate these mathematical models. We then draw the conclusion that from a numerical viewpoint, the bias between the fair value of the American Call and the value of the corresponding European Call is mainly due to approximation errors, which can be mitigated when Trinomial Stochastic Trees are used. Once we have tested the pricing routines, we have estimated a potential error that can be made because of the approximation of the price of an American call with the equivalent European, in the scenario we are considering.

3.1 Introduction

As outlined in a recent note from the Actuarial Association of Europe (2016) , nowadays negative nominal interest rates for long term maturities are observable in European financial markets, because of the 2007 financial crisis, together with the 2011 Euro area sovereign debt crisis. Several central banks, in fact, as the ECB, the Swiss National Bank, the Swedish Riksbank, the Bank of Japan and the Central Bank of Hungary have employed negative interest rate policies to give a further monetary policy stimulus to try to boost the economy. (Arteta et al., 2017). In addition to the economic effects, this led to several technical problems, as the existing pricing models do not give proper valuations, so that either the financial position cannot be correctly priced or the results can be questioned (Kooiman, 2015). The risk management, the regulation system and the pricing models are not able to handle this situation because they have not been thought

to incorporate negative interest rates, since nobody believed that it could have happened (Hull, 2012; Haugh, 2007; Giribone et al., 2017). This scenario has also caused the invalidation of the stochastic models of Black and Sholes, 1976, Cox, Ingersoll and Ross (1985), Black-Derman-Toy (1990), because they assume the risk-free interest rate to be positive. Generally speaking, it is not easy to forecast and to price derivative products, since they are widely used in hedging risk. (Recchioni, Sun and Tedeschi, 2016). The recent empirical evidence of negative interest rates in both the U.S. and EURO short-term government bond yields, have highlighted the need to partially reconsider these models. More precisely, the stochastic interest rate process should allow for negative values. From a financial standpoint, it is therefore necessary to check to what extent the existing pricing models can be adapted to incorporate negative nominal rates. This aspect has been already investigated in some research papers: Giribone, Ligato and Mulas (2017)and Giribone and Ligato (2016) discuss the issue for options written on interest rates, both from the practical and the theoretical viewpoint; Recchioni, Sun and Tedeschi (2017), focusing on foreign exchange and index options investigate whether the use of models allowing for negative interest rates can improve option pricing and implied volatility forecasting; Abudy and Izhakian (2013) , discusses a new closed form for option pricing that leads to sensitively lower level of error in European options pricing. Besides, Inui (2015) adapts the Nelson-Siegel model (1987) to include the negative interest rates. Finally, the Hull and White model (1994) , has been recently adapted to calibrate in a more proper way in situation when the underlying is a negative interest rate (Hull and White, 2015), Nardon and Pianca (2009), analyze some models of option pricing issued on underlying with dividends, focusing their attention on Haugh et al., (2003) and on other binomial methods, Chen (2010) has calculated American option pricing formula for uncertain financial markets. However, to the best of our knowledge, much

less efforts have been devoted to model the effects of negative nominal interest rates in option pricing for other types of underlying. In particular, some issues might arise in the case of equity that does not pay dividends: finding the fair value at early exercise of an American Call might be tricky, as it could not match the value of a European Call option with the same parameters. The problem is relevant, because of its corporate implications, as the option evaluation could make the difference when valuing a firm. This paper aims to fill this gap. Our research question is discussing whether the known approximation formulas can effectively bypass the above highlighted problems. We illustrate an empirical application, where we compare the estimation of a number of quasi-closed formulas, with that provided by the stochastic trinomial trees algorithm. This algorithm has also been used by Leippold (2004) to improve the Hull-Whites procedure to calibrate the tree to bond prices by circumventing the forward rate induction and numerical root search algorithms, proposed with a computationally efficient implementation of general one factor short rate models with a trinomial tree, and he found out it offers a more efficient and flexible method of constructing a trinomial term structure model. This methodology, which is an improvement of the binomial tree model, thanks to its flexibility, also has the advantage of being easily implemented and calibrated to both prices and to adjust the volatility structure (Leippold and Wiener, 2003) and to allow a stock price to move up, down or stay the same with certain probabilities. Trinomial trees provide an effective method of numerical calculation of option prices within Black-Scholes share pricing model (Clifford et al., 2010) and this methodology may also be applied to price barrier options or to calculate the greeks. We finally highlight how the bias between the fair value of the American Call at early exercise and the value of the corresponding European Call can be strongly mitigated using this latter methodology.

3.2 Theoretical Issues and Methodology

3.2.1 On the Violation of the Equivalence between American and European Call Value

Here we focus on the well-known property according to which there is no convenience of an early exercise with an American Call with null dividend. Let us denote by $C_A = f_A(S, K, T, r, q, \sigma)$ the value of an American Call with spot price S , strike price K , time to expiration T , the risk-free interest rate r , continuous dividend yield q , and volatility σ . We also denote by $C_E = f_E(S, K, T, r, q, \sigma)$ the corresponding value for a European contingent claim. We focus on the well-known property (Hull, 2014) according to which in case of an underlying with null dividends we have:

$$f_A(S, K, T, r, q, \sigma) = f_E(S, K, T, r, q, \sigma) \tag{3.1}$$

To demonstrate this, we suppose to have two different portfolios, A and B. A is constituted by a long position of a European Call option and by a zero-coupon-bond which pays K at the maturity date $T = 0$. The second portfolio is composed by the same underlying of the option S. By varying time t, we can check the following scenarios:

- *Portfolio A*: its value at time T is

$$Port_A = C_E(S, T, K) + K \exp(-rT) \tag{3.2}$$

  At the expiration time ($T = 0$) we can have two different situations, which vary if $S \leq K$ or if $S > K$. When $S \leq K$, at the expiration time, we have:

$$C_E(S,T,K)+K\exp(-rT) = C_E(S_{T=0},0,K)+K\exp(0) = max[S_{T=0}-K,0]+K \tag{3.3}$$

Where $C_E(S_{T=0},0,K)$ is the payoff and $K\exp(0) = 1$. So, under this hypothesis, the value of portfolio A equals the maximum value between $max[S_{T=0}-K,0]+K$. However, since the option expires *out-of-the-money*, its maximum value is K. On the other hand, under the hypothesis that $S_{T=0} > K$, the value of portfolio A is the same of eq. (3.3) but, in this context, the European option expires *in the money*. Its value at the expiration time is then $S_{T=0} - K + K = S_{T=0}$.

- *Portfolio B* is equal to $S_{T=0}$ both when $S_{T=0} \leq K$ and when $S_{T=0} > K$.

Since at time $T = 0$, the value of portfolio A can not be lower than the value of Portfolio B, we can extend the same demonstration to a generic time $t = T$. As a result, we get

$$C_E(S,T,K) + Kexp(-rT) \geq S_T \tag{3.4}$$

This relation holds true for every value of T, independently from the sign of the interest rate. By considering $K\exp(-rT)$, we can check the discounted factor, in ordinary market conditions where $r > 0$, we get that $\exp(-rT) < 1$. This condition has been widely proved in literature. On the other hand, when $r < 0$, $\exp(-rT) > 1$. This result further strengthen the positivity of the first member of the inequality. We thus obtain that:

$$C_E(S,T,K) \geq S_T - K\exp(-rT) \tag{3.5}$$

By considering that the value of an American option can not be lower than

its European equivalent, we get:

$$C_A(S,T,K) \geq C_E(S,T,K) \tag{3.6}$$

It is possible to replicate every right of an European option to an American option. Let us assume a long position of an American option. The owner has the opportunity to exert its right not only at the expiration time, but in every moment of its life. This additional opportunity has a positive (or, at least non-zero) value. To conclude, putting together eq. (3.5) and eq. (3.6), we obtain:

$$C_A(S,T,K) \geq C_E(S,T,K) \geq S_T - Kexp(-rT) \tag{3.7}$$

The second part of the demonstration consists in showing the following inequality:

$$C_A(S,T,K) \geq S_T - K \tag{3.8}$$

The demonstration is divided into two steps:

- In the first step, we show that according to the definition of the option, C_A can not be negative, since it is a limited liability. The direct consequence of its payoff $\max(S-K,0)$ is a $C_A(S,T,K) \geq 0$.

- In the second step, we consider the scenario of a *in-the-money* call -i.e. $S_T > K$. If the corresponding value of a call option would be lower than $S_T - K$, it would be possible to buy the option, to exert it and to to sell it, leading to the possibility of an arbitrage. In this case, the net profit would be $S_T - [C_A(S,T,K) + K] = S_T - K - C_A(S,T,K) > 0$. To avoid the risk of arbitrage, we must have

$$C_A(S,T,K) \geq S_T - K \tag{3.9}$$

Looking at eq. (3.7) and eq. (3.9), we can notice that, under the standard market hypothesis in which $r > 0$, the condition of $S_t - K < S_t - K\exp(-rT)$ implies that an American call written on an underlying with a null dividend is more valuable before being exercised. So, in presence of standard market conditions, the early exercise of a call option written on an underlying with a null dividend is not an optimal choice.

However, in anomalous market conditions characterized by negative interest rates $r < 0$, the condition expressed by $S_T - K < S_T - K\exp(-rT)$ does not hold anymore, because $\exp(-rT) > 1$. In this context is not possible to *a priori* conclude if it is convenient to exercise the option before its expiration time. Its value is then potentially *non null*.

3.2.2 Methodology

Pricing American contingent claims has traditionally represented a stimulating field of analysis as, in contradiction to European options, they can be exercised at any time before or at maturity. In this case, the Black-Scholes methodology cannot be applied, and it is necessary to use approximated schemes. Reviewing the related literature requires huge efforts, besides it is out of the scope of this work: the interested reader can refer to (Hull, 2014). Nevertheless, we are mainly concerned with two sub-groups of the above methods. In the first group, we consider three quasi-closed formulas that conveys in different ways the original idea discussed in (Roll, 1977) . In particular, the Barone-Adesi and Whaley-BAW-model (1987) is a quadratic approximation method for pricing exchange-traded American call and put options on commodities and commodity futures. Using the same notational conventions as in Sec.3.1, we consider an American Call option whose underlying has a cost of carry equal to $b = r - q$. When $b \geq r$, ceteris paribus the value of the American Call is equal to that of the European Call so

that the Generalized Black-Scholes-GBS-formula for European contingent claims applies the *Barone-Adesi-Whaley model*

The method based on the quadratic approximation of Barone-Adesi-Whaley (1987) can be used for American Call and option pricing with an equity underlying with a cost-of-carry equals to b. When $b \geq r$, the value of an American Call equals its European equivalent and its price is determined according to the Black-Scholes equation.

$$C_A(S,K,T) = C_E(S,K,T) + A_2 \left(\frac{S}{S^*}\right)^{q_2}, S < S^* \tag{3.10}$$

$$C_A = S - K, S \geq S^*$$

Where $C_E(S,K,T)$ is the fair-value of a European Call plain-vanilla option.

$$A_2 = \frac{S^*}{q_2}\left[1 - \exp\left[(b-r)\,T\right] N\left[d_1\left(S^*\right)\right]\right], d_1(S) = \frac{\ln\left(S/K\right) + \left(b + \sigma^2/2\right)T}{\sigma\sqrt{T}}$$

$$q_2 = \frac{-(N-1) + \sqrt{(N-1)^2 + 4M/X}}{2}, M = \frac{2r}{\sigma^2}, N = \frac{2b}{\sigma^2}, X = 1 - \exp\left(-rT\right)$$

$$P_A(S,K,T) = P_E(S,K,T) + A_1 \left(\frac{S}{S^{**}}\right)^{q_1}, S > S^{**} \tag{3.11}$$

$$C_A = K - S, S \leq S^{**}$$

Where $P_E(S,K,T)$ is the fair-value of a European Put plain-vanilla option.

$$A_1 = -\frac{S^{**}}{q_1}\left[1 - \exp\left[(b-r)\,T\right]N\left[-d_1\left(S^{**}\right)\right]\right]$$

$$q_1 = \frac{-(N-1) - \sqrt{(N-1)^2 + 4M/X}}{2}$$

S^* is the critical value for the Call option which satisfies the following equality:

$$S^* - K = C\left(S^*, K, T\right) + \left[1 - \exp\left[(b-r)\,T\right]N\left[d_1\left(S^*\right)\right]\right]S^*\frac{1}{q_2} \tag{3.12}$$

$$LHS\left(S_i\right) := S_i - K, RHS\left(S_i\right) = C\left(S_i, K, T\right) + \left[1 - \exp\left[(b-r)\,T\right]N\left[d_1\left(S_i\right)\right]\right]S_i\frac{1}{q_2}$$

Where i is the indicator of the Newton-Raphson iteration. This equation can be solved numerically with the monodimensional root-finding algorithm of Newton-Raphson. The partial derivative of RHS with respect to S_i is:

$$\frac{\partial RHS}{\partial S_i} = \exp\left[(b-r)\,T\right]N\left[d_1\left(S_i\right)\right]\left(1 - \frac{1}{q_2}\right) + \frac{\left[1 - \frac{\exp\left[(b-r)T\right]n\left[d_1(S_i)\right]}{\sigma\sqrt{T}}\right]}{q_2} \tag{3.13}$$

Given a starting value of S_i, the Newton-Raphson method guarantees that the succeeding and better estimation of S_{i+1} is:

$$S_{i+1} = \frac{K + RHS\left(S_i\right) - b_i S_i}{\left(1 - b_i\right)} \tag{3.14}$$

The iterative process continues until it reaches a stop criterion of the algortihm,

which is usually computed in the following way:

$$\frac{|LHS\left(S_i\right) - RHS\left(S_i\right)|}{K} < Tollerance \tag{3.15}$$

S^{**} is the critical value for the Put option that satisfies the following equality:

$$K - S^{**} = P\left(S^{**}, K, T\right) - \left[1 - \exp\left[\left(b - r\right)T\right]N\left[-d_1\left(S^{**}\right)\right]\right]S^{**}\frac{1}{q_1} \tag{3.16}$$

$$VS\left(S_j\right) = K - S_j, HS\left(S_j\right) = P\left(S_j, K, T\right) - \left[1 - \exp\left[\left(b - r\right)T\right]N\left[-d_1\left(S_j\right)\right]\right]S_j\frac{1}{q_1}$$

$$\frac{\partial HS}{\partial S_j} = -\exp\left[\left(b - r\right)T\right]N\left[-d_1\left(S_j\right)\right]\left(1 - \frac{1}{q_1}\right) - \frac{\left[1 + \frac{\exp\left[(b-r)T\right]n\left[-d_1\left(S_j\right)\right]}{\sigma\sqrt{T}}\right]}{q_1} \tag{3.17}$$

$$S_{j+1} = \frac{\left[K - HS\left(S_j\right) + b_j S_j\right]}{\left(1 + b_j\right)} \tag{3.18}$$

As in every iterative research procedures, the Newton-Raphson method needs an initialization seed as well. For this purpose, scholars suggest to follow this dynamic:

$$S_1^* = K + \left[S^*(\infty) - K\right]\left[1 - \exp\left(h_2\right)\right], h_2 = -\left(bT + 2\sigma\sqrt{T}\right)\left[\frac{K}{S^*(\infty) - K}\right]$$

$$S_1^{**} = S^{**}(\infty) + [K - S^{**}(\infty)] \exp(h_1), h_1 = + \left(bT - 2\sigma\sqrt{T}\right) \left[\frac{K}{K - S^{**}(\infty)}\right]$$

Where $S(\infty)$ is the critical price with an infinite expiration time, $S(T \to \infty)$:

$$S^*(\infty) = \frac{K}{1 - 2\left[-(N-1) + \sqrt{(N-1)^2 + 4M}\right]^{-1}}$$

$$S^{**}(\infty) = \frac{K}{1 - 2\left[-(N-1) - \sqrt{(N-1)^2 + 4M}\right]^{-1}}$$

- *The Bjerksund-Stensland model*

The approximation of Bjerksund and Stensland can be used to price an option with different underlyings: equity, futures and currencies. The algorithm is more generally applicable than the previous ones; in addition to that, it has also the characteristic of being computationally more efficient and precise, especially on the estimation of the fair-value of derivatives with longer expirations.

$$C = \alpha S^{\beta} - \alpha\phi(S, T, \beta, I, I) + \phi(S, T, 1, I, I) - \tag{3.19}$$

$$\phi(S, T, 1, K, I) - K\phi(S, T, 0, I, I) + K\phi(S, T, 0, K, I)$$

Where:

$$\alpha = (I - K)I^{-\beta}, \beta = \left(\frac{1}{2} - \frac{b}{\sigma^2}\right) + \sqrt{\left(\frac{b}{\sigma^2} - \frac{1}{2}\right)^2 + 2\frac{r}{\sigma^2}}$$

The function $\phi(S, T, \gamma, H, I)$ is defined as:

$$\phi(S,T,\gamma,H,I) = \exp(\lambda)S^{\gamma}\left[N(d) - \left(\frac{I}{S}\right)^{\kappa} N\left(d - \frac{2\ln(I/S)}{\sigma\sqrt{T}}\right)\right]$$

$$\lambda = \left[-r + \gamma b + \frac{1}{2}\gamma(\gamma-1)\sigma^2\right]T, d = \frac{\ln(S/H) + \left[b + \left(\gamma - \frac{1}{2}\right)\sigma^2\right]T}{\sigma\sqrt{T}}$$

$$\kappa = \frac{2b}{\sigma^2} + (2\gamma - 1)$$

The trigger price, I, is:

$$I = B_0 + (B_\infty - B_0)(1 - \exp[h(T)])$$

$$h(T) = -\left(bT + 2\sigma\sqrt{T}\right)\left(\frac{B_0}{B_\infty - B_0}\right), B_\infty = \frac{\beta}{\beta - 1}K, B_0 = \max\left[K, \left(\frac{r}{r-b}\right)K\right]$$

If $S \geq I$, is optimal to immediately exercise the right of buying and the value of the derivative must equal its intrinsic value $S - K$. At the same time, if $b \geq r$, the exercise is not optimal and its fair value is calculated by the Black-Scholes formula for european options.

Another reason for which this approach is very used is the presence of the following Call-Put Parity for American Options:

$$P_A(S,K,T,r,b,\sigma) = C_A(K,S,T,r-b,-b,\sigma) \tag{3.20}$$

By using this transformation, it is no longer necessary anymore to develop a separated formula for the pricing of American Option.

An alternative to the above-mentioned approximation methods is represented by stochastic binomial and trinomial trees. Assuming the stock price to follow a discrete time process, in the binomial tree scheme the life of the option until the maturity T is decomposed into N time steps of equal length. At each time step, the underlying will move either up or down by a specific factor $u = \exp(\sigma\sqrt{\Delta t})$ or $d = \exp(-\sigma\sqrt{\Delta t})$, with probability p and $(1-p)$, respectively. The value of the American Call exercised at expiration is:

$$C_E^{BIN} = \exp(-rT) \sum_{i=0}^{N} \frac{N!}{i!(N-i)!} p_u^i (1-p_u)^{N-i} \max\left(Su^i d^{N-i} - K, 0\right) \tag{3.21}$$

where: $p_u = \frac{\exp[(r-q)\Delta t]-d}{u-d}$

As previously defined, q is the dividend yield.
To properly assess C_E in case of early exercise, at each node of the tree the following pay-off must be applied:

$$\max\left(S_{i,j} - K, \exp(-r\Delta t)\left[p_u f_{i,j+1} + (1-p_u) f_{i,j}\right]\right)$$

where $f_{i,j}$ is the value of the Call for the node of position (i,j) in the tree. More in details, i represents the number of times the asset price has gone up to reach a node, while j is the node assigned to i. (Haugh, E.G., 2007). The initial

value of the option can be then derived using the standard backward induction technique. A straightforward extension of this procedure is given by the trinomial scheme algorithm (Boyle, P.P.1986), with the underlying that can now assume three different states: up, down or unchanged. The increase in the number of possible states allows to lower the number of necessary steps for the convergence of the procedure, without any loss in the estimation accuracy. The size of the jumps is usually set to: $u = \exp\left(\sigma\sqrt{(2\Delta t)}\right)$ and $d = \exp\left(-\sigma\sqrt{(2\Delta t)}\right)$, so that the probability of reaching upward/downward branches is given by:

$$p_u = \left[\frac{\exp\left(\frac{b\Delta t}{2}\right) - \exp\left(-\sigma\sqrt{\frac{\Delta t}{2}}\right)}{\exp\left(\sigma\sqrt{\frac{\Delta t}{2}}\right) - \exp\left(-\sigma\sqrt{\frac{\Delta t}{2}}\right)}\right]^2$$

$$p_d = \left[\frac{\exp\left(\sigma\sqrt{\frac{\Delta t}{2}}\right) - \exp\left(\frac{b\Delta t}{2}\right)}{\exp\left(\sigma\sqrt{\frac{\Delta t}{2}}\right) - \exp\left(-\sigma\sqrt{\frac{\Delta t}{2}}\right)}\right]^2$$

the probability p_m of reaching the intermediate node is: $p_m = 1 - p_u - p_d$

3.3 Examples and Discussion

We consider three scenarios (A, B, and C), and for each of them we compute the value of the American Call with the approximation schemes illustrated in Sec. 3.2.2. In details:

- A represents a typical market situation, with a positive risk-free rate, and with a dividend-paying stock as underlying;
- B simulates a market situation with a positive risk-free rate, and with a null dividend stock as underlying: in this case, as $r > 0$, Equation (3.1) holds;

Parameters	Scenario A	Scenario B	Scenario C
S	100	100	100
K	100	100	100
r	10%	10%	-0.301%
q	10%	0	0
b	0	10%	-0.301%
σ	25%	25%	25%
T	0.25	0.25	0.25

Table 3.1: Parameters employed in the three scenarios simulation.

Scheme	Scenario A	Scenario B	Scenario C
BAW	4.8908	6.2545	4.9479
BS1993	4.8765	6.2545	4.9479
BS2002	4.8802	6.2545	4.9479
T-tree	4.8801	6.2544	5.0461

Table 3.2: Simulation results for the three scenarios under different estimation models.

- C considers an atypical situation, with a negative risk-free rate. The underlying stock, likewise in the B case, does not pay any dividend.

The parameters employed in the simulation are reported in Table 3.1: we used the annualized value of the volatility, while T is expressed in year fractions; finally, the value of r in the third scenario corresponds to the 3-months tenor of the Euribor at 9 September 2016, as provided by Bloomberg.

The simulation results are shown in Table 3.2, where we employed the following abbreviations: BAW to indicate the Barone-Adesi and Whaley model, BS1993 and BS2002 referring to 1993 and 2002 Bjerksund and Stensland approximation formulas, respectively, and T-TREE for the trinomial tree. In this latter case, the discretization steps were set to $N = 9000$.

Looking at Table 3.2, several remarks arise. First, in scenario A, by construction, the early exercise of the American Call is sometimes optimal, and this is duly taken into consideration by every approximation scheme. In scenario B, according to the option property, as it replicates a situation where the early exercise is never optimal and (3.1) holds, all the examined schemes have properly applied the Black-Scholes formula for the European Call. In the third case, the methods relying on quasi-closed approximation formulas (BAW, BS1993 and BS2002) have still exploited (3.1) which is no more verified, so that they all incorrectly estimated the American Call value. On the other hand, the T-TREE scheme generated a more robust estimation, because the convenience for the early exercise was checked on each node of the tree. As further step, we want to test the numeric error at $N = 9000$. It is possible to test the trinomial tree with the exact formula of an European Call Option only under scenario B. In scenario A, in fact, we have a discrete methodology vs an approximated methodology. Under scenario C, we can not validate the model because we have a negative risk-free rate, that is the object of what we want to demonstrate.

In figure 3.1, we show the convergence of the numeric error with the exact formula of BS, with a positive value of r ($r = 0.10$). We validated the model by quantifying the discretization error of the trinomial tree with respect to the closed formula of BS at the variation by putting the discretization steps as a parameter. As we can see, the error we have introduced using 9000 steps is very low: it is about 10^{-4} As preliminary conclusion, we can therefore state that using the trinomial trees rather than other approximation schemes might be preferable, as this methodology seems being more robust to anomalous parameters values.

We then moved one step further, giving additional instruments to evaluate such a robustness. To such aim, we focused on the scenario C (i.e. the one where critical issues arose) and we studied the behaviour of the discrepancies which arise

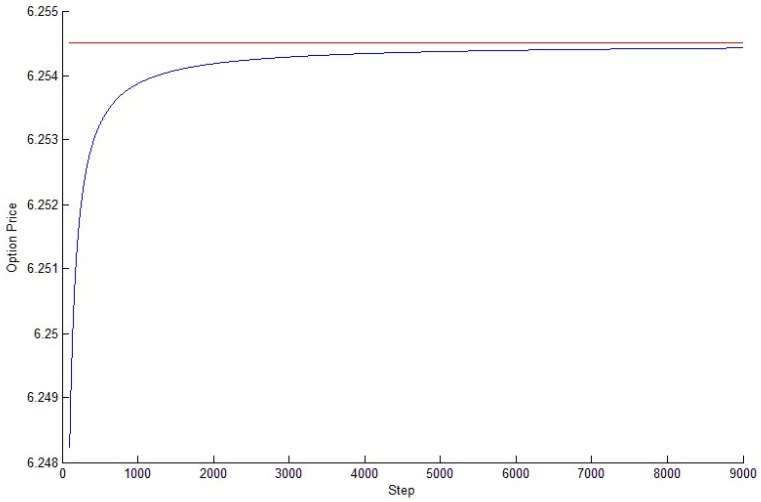

Figure 3.1: Convergence test between the trinomial tree and BS2002, varying steps parameter from 0 to $N = 9000$ and with $r > 0$

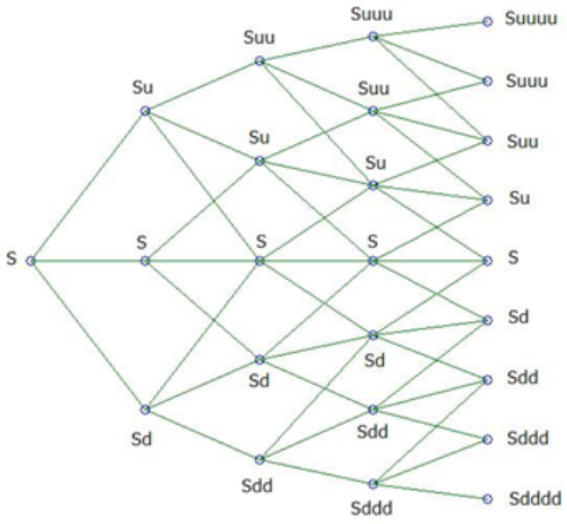

Figure 3.2: Construction of a trinomial tree of Cox-Ross-Rubinstein

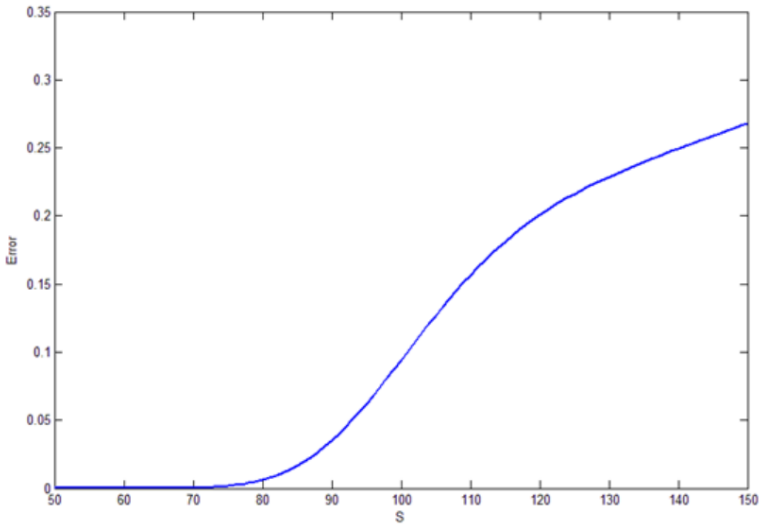

Figure 3.3: Behaviour of the discrepancy varying S

from using different estimation methods, that is the T-TREE ($ValueTTree$) and the BS2002 ($ValueBS2002$) schemes, varying one at a time S, K, T, σ and r. Then, we also considered the behaviour of the relative error of S, K, T, σ and r. The relative error is defines as $(ValueBS - ValueTTree)/ValueTTree$. The choice of BS2002 is motivated as it is generally acknowledged to be the more accurate among the examined quasi-closed formulas. Figure 3.4 shows the behaviour of the discrepancy between $Error = ValueTTree - ValueBS2002$. This is compared with the associated relative error for every parameter.

Examing the results, from Figure 3.3 we observe that, as the spot value S, varies on the x-assis, the gap between the different estimation methods lies within the interval $[0.02, 0.35]$, and tends to increase, originally in a more than proportional fashion. This suggests the existence of model risk, raising as the options moneyness increases. In the second panel, we present the relative error of S. The behaviour of the relative error is slightly different: it decreases as S increases.

Similar considerations also apply to the behaviour of the discrepancy and of

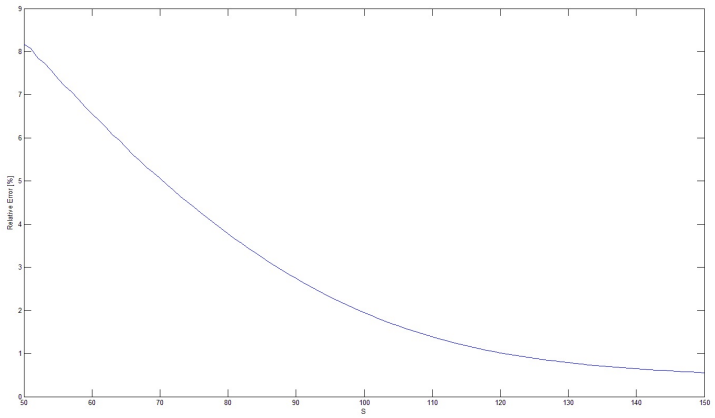

Figure 3.4: Behaviour of the relative error varying S

the relative error with respect to the strike price K, shown in Figure 3.5. In this second case, in fact, the lower K (high moneyness) the higher Error is, i.e. the higher the gap between the T-TREE and the best quasi-closed approximation method. Also in this case, the behaviour of the relative error is different from the one of the discrepancy: it increases when K increases.

In the case of the time to maturity T, observable in Figure 3.7, the divergence between $TTREEandBS2002$ is extremely evident, with the gap varying in the range $[0.05, 0.5]$: the longer the hedging period, the worst the performance of conventional methods is. The relative error here follows the same pattern of the discrepancy between $TTREEandBS2002$.

As regard concerning the behaviour of gap varying r, Figure 3.8 examines only the case of negative risk-free rates. In this case, we can observe an elbow-like curve, with higher discrepancy values (higher than 0.1) concentrated around lowest (and quite unrealistic) negative nominal rates. In every case, as $r < 0$, Error never falls under the 0.09 threshold. Also in r, the relative error follows the same pattern of the behaviour of the discrepancy: we observe an elbow-like

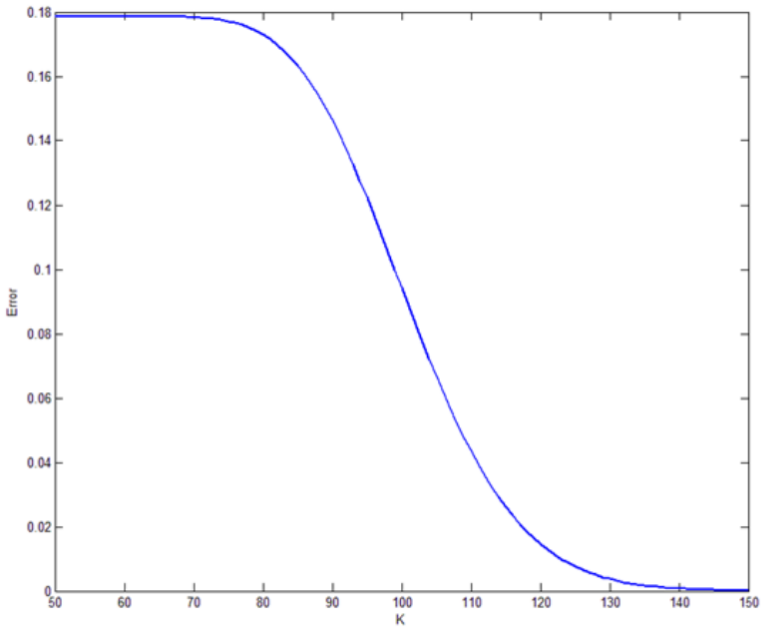

Figure 3.5: Behaviour of the discrepancy varying K

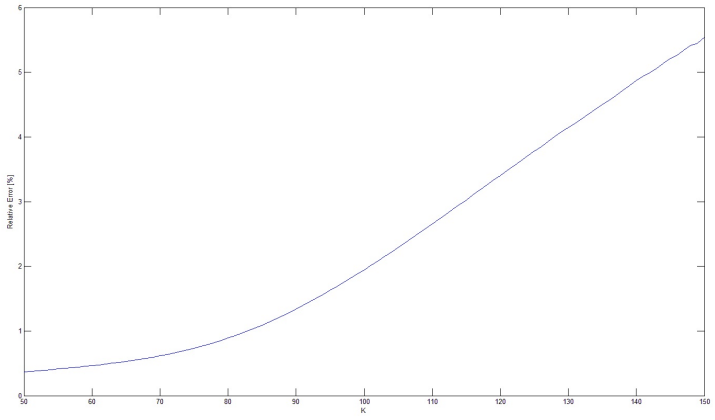

Figure 3.6: Behaviour of the relative error varying K

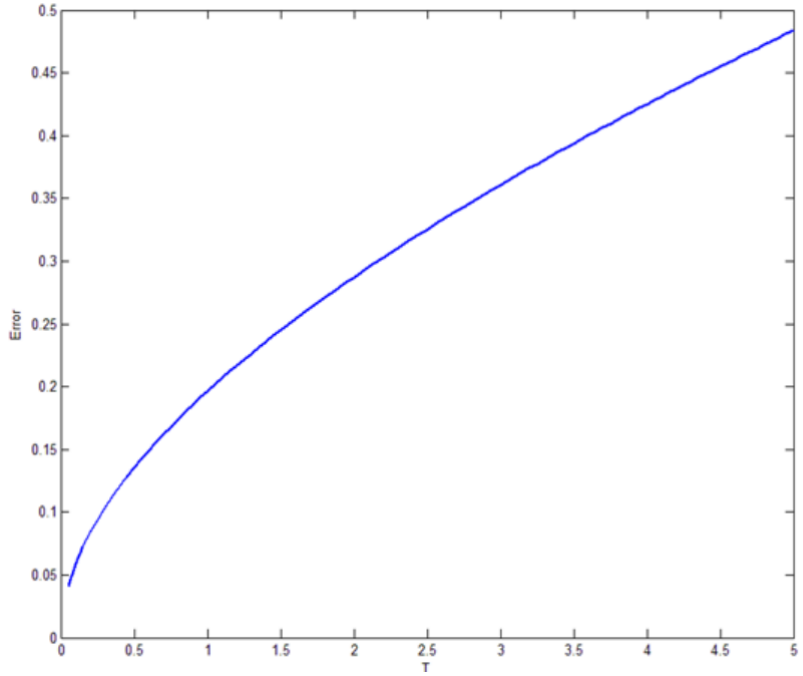

Figure 3.7: Behaviour of the discrepancy varying T

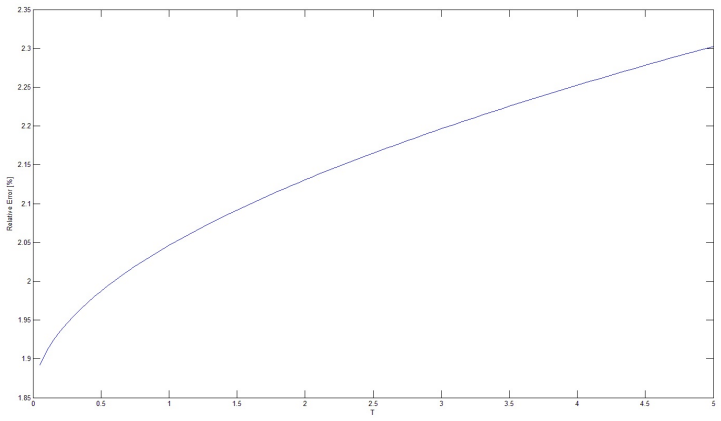

Figure 3.8: Behaviour of the relative error varying T

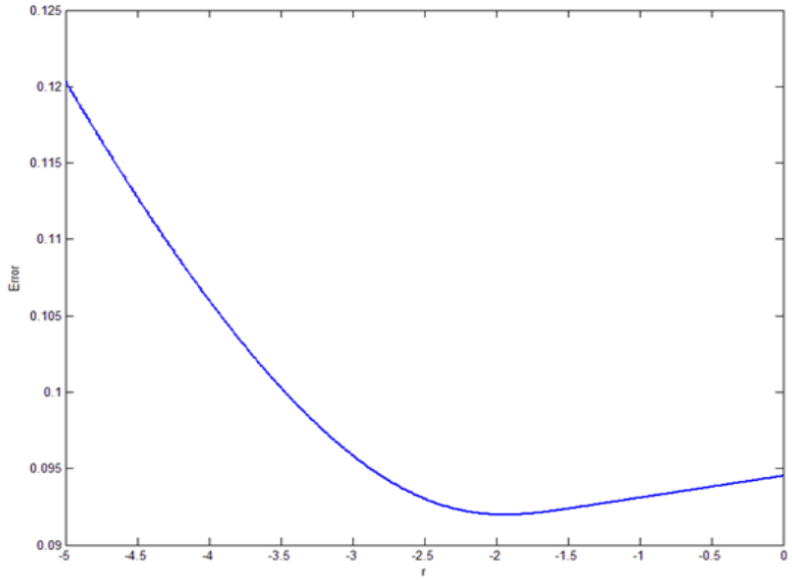

Figure 3.9: Behaviour of the discrepancy varying r

curve, that is flatter at the end of the considered values of r.

Finally, from Figure 3.11 we can state that there is a positive correlation between the behaviour of the discrepancy and σ, with the former monotonically growing as the annualized volatility increases. The same positive correlation is present between the relative error and σ.

Figure 3.13 and 3.14 represents the behaviour of the discrepancy with $r > 0$ and the behaviour of the relative error with a positive r respectively. r varies between 0 and 10: we can notice that as r increases, both the discrepancy and the relative error linearly decrease.

We then examined the impact of different approximation schemes on the value of the most used Greeks (Hull, 2014) , because of the paramount role that they play in the hedging activity. We therefore evaluated Δ, ϑ and Θ, being:

$$\Delta = \frac{\partial C_A}{\partial S} \tag{3.22}$$

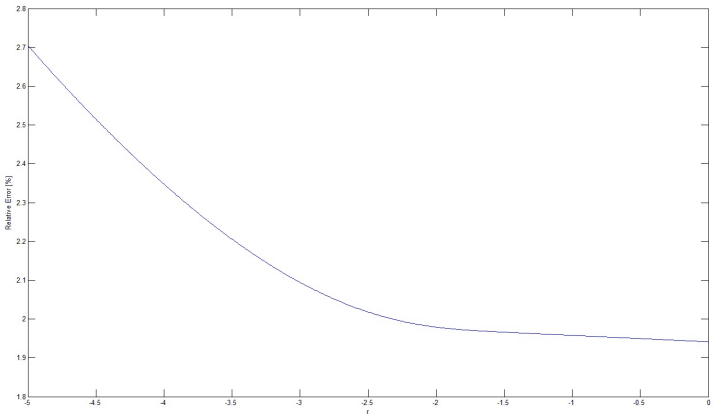

Figure 3.10: Behaviour of the relative error varying r

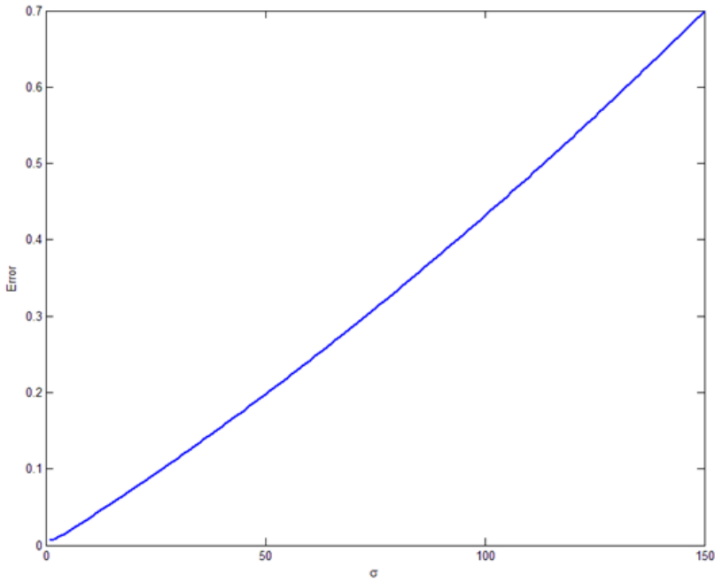

Figure 3.11: Behaviour of the discrepancy varying σ

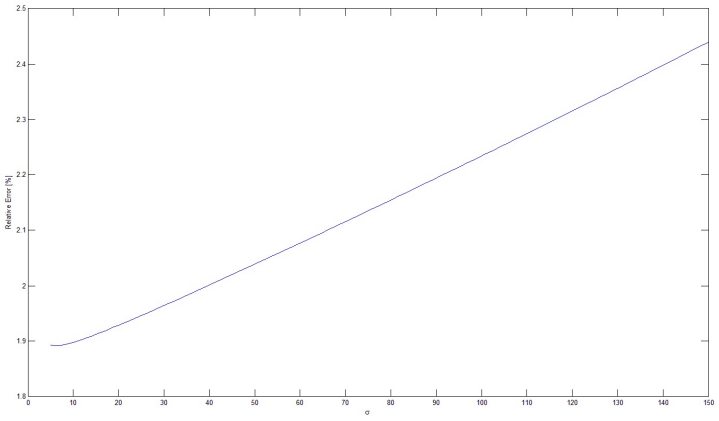

Figure 3.12: Behaviour of the relative error varying σ

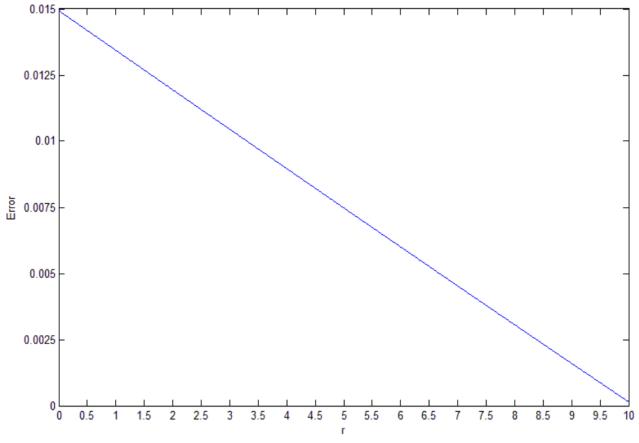

Figure 3.13: Behaviour of the discrepancy with $r > 0$

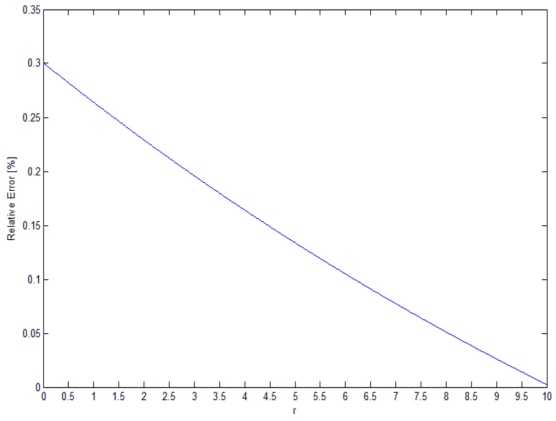

Figure 3.14: Behaviour of the relative error with $r > 0$

$$\vartheta = \frac{\partial C_A}{\partial \sigma} \tag{3.23}$$

$$\Theta = \frac{\partial C_A}{\partial T} \tag{3.24}$$

Where C_A is the option value, and S , σ and T are as usual.

Starting from the results in Table 3.3, we aim at replicating the situations already discussed in the first sensitivity analysis, with all the methods generating the same values for each Greek in the Scenarios A and B, and with the T-TREE scheme providing different results in case C.

3.4 Conclusions

In this paper, we examined how the existing numerical schemes react in the pricing of an American Call option, in presence of anomalous conditions. We focused on

Greeks	Scenario A	Scenario B	Scenario C
ΔBAW	0.5156	0.6035	0.5225
ΔBS1993	0.5151	0.6035	0.5225
ΔBS2002	0.5154	0.6035	0.5225
ΔT-tree	0.5152	0.6033	0.5294
ϑBAW	19.5379	19.2716	19.9153
ϑBS1993	19.4847	19.2716	19.9153
ϑBS2002	19.4951	19.2716	19.9153
ϑT-tree	19.5035	19.2714	20.3289
ΘBAW	-9.4356	-15.0457	-9.8153
ΘBS1993	-9.3404	-15.0457	-9.8153
ΘBS2002	-9.3612	-15.0457	-9.8153
ΘT-tree	-9.3662	-15.0424	-10.0203

Table 3.3: The impact of different approximation schemes on the value of Delta, Vega and Theta Greeks.

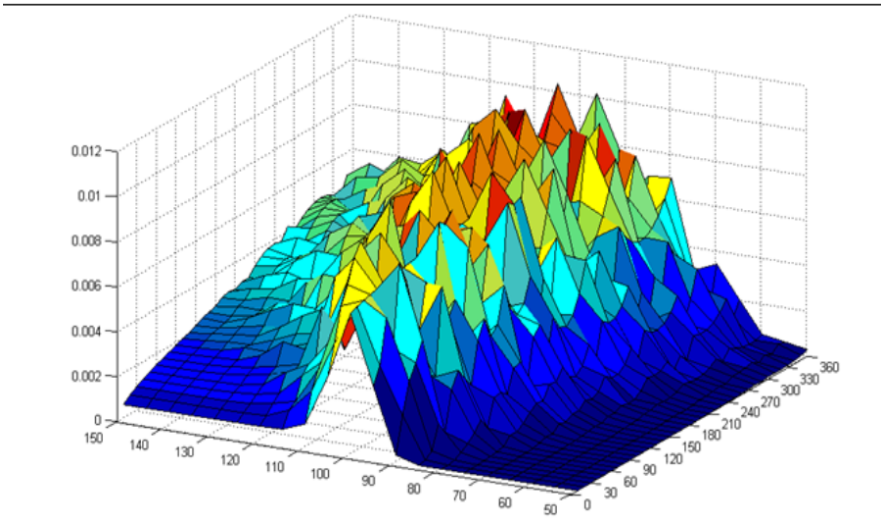

Figure 3.15: Surface of the error made on the evaluation of the Δ of the American option, with S and T varying.

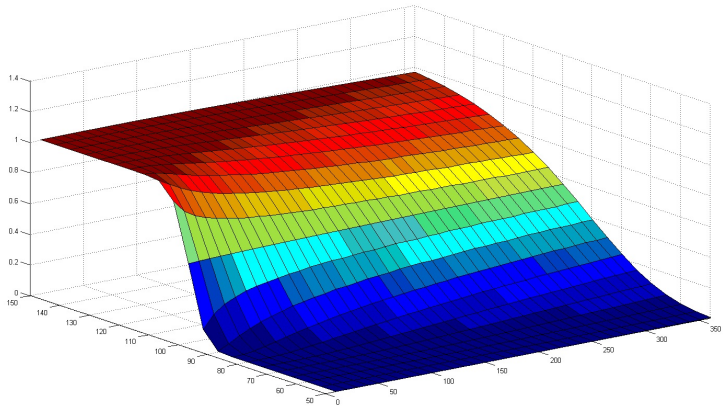

Figure 3.16: Surface of the Δ greek of the American option, with S and T varying.

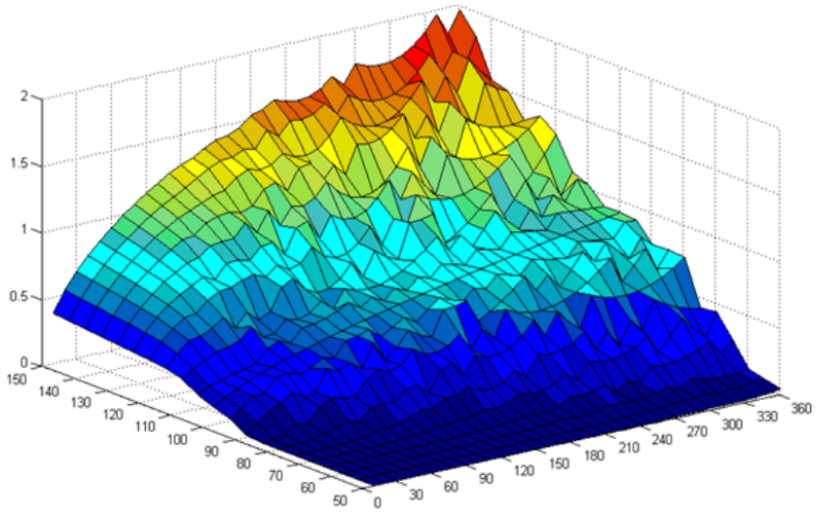

Figure 3.17: Surface of the error made on the evaluation of the Vega of the American option, with S and T varying.

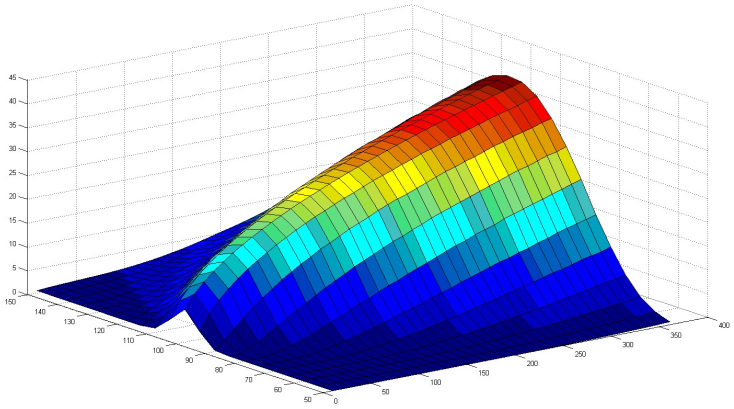

Figure 3.18: Surface of the Vega greek of the American option, with S and T varying.

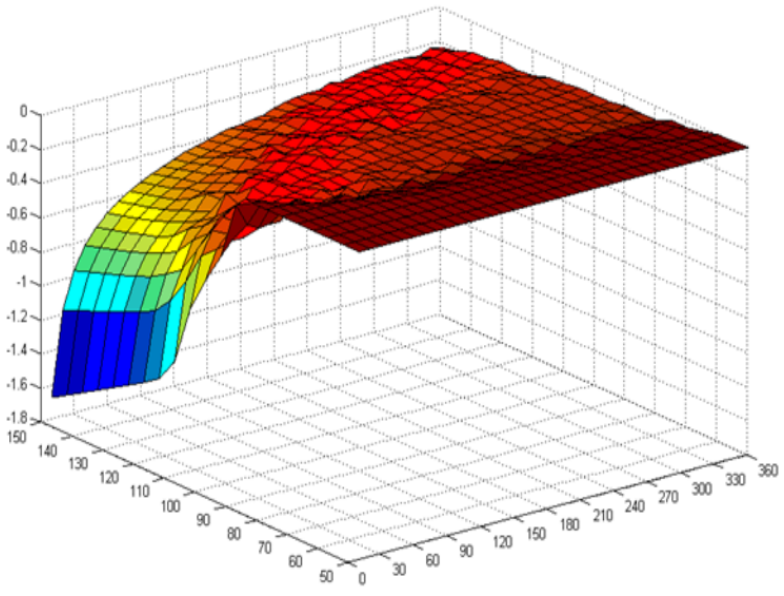

Figure 3.19: Surface of the error made on the evaluation of the Θ of the American option, with S and T varying.

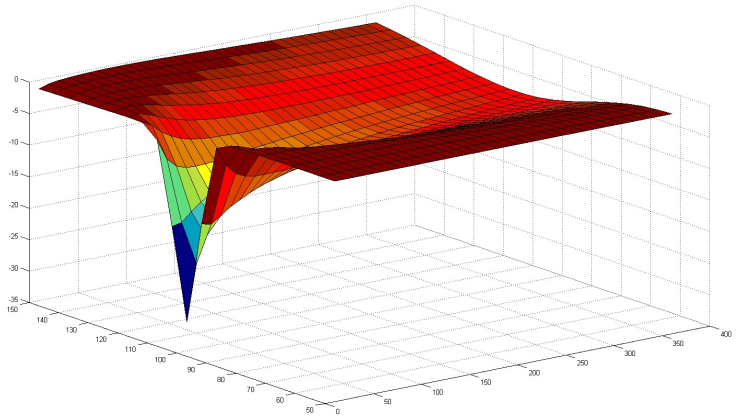

Figure 3.20: Surface of the Θ greek of the American option, with S and T varying.

the case of negative risk-free rate and zero dividends stock as underlying and we put at work three quasi-closed approximation formulas and the Trinomial Trees technique. We then analyzed three simulated scenarios, replicating different market conditions, to conclude that in the case of negative risk-free rate it is preferable to price the American Calls by way of the Trinomial tree (T-TREE) scheme. This is because unlike the other techniques, T-TREE does not price the American Call using the equivalence between its fair value at early exercise and the corresponding value of the European Call with the same financial features, but rather the convenience for the early exercise is checked on each node of the tree. The quadratic approximation method by Barone-Adesi and Whaley (1987), in fact, is fast and good for practical input values: when the cost-of-carry rate is greater than the interest rate, it allows to find the American call value by using the generalized Black-Scholes-Merton (BSM) formula. On the other hand, the Bjerksund and Stensland model is sometimes more accurate for long-term options than the BAW, especially in its 2002 formulation: it may be used to price american options on stocks, futures and currencies. In this way, the T-

TREE is protected from the risk that this property is no longer valid, as in the case with negative nominal rates. Moreover, in such anomalous conditions, the accuracy of the T-TREE with respect to the other methods is very robust to both hard negative values of the risk-free rate, and to increases with respect to the moneyness of the underlying, as well as of the volatility and the maturity of the option. The T-Tree is a good methodology because it tracks in a path-dependent way all the possible conditions of the underlying: it may be used to both time-dependent interest rates and volatilities. It is an alternative of the binomial tree but it works thanks to the same calculations, and it works as other well-known methodologies, as explicit finite difference method. Future research plans include increasing the robustness of our survey by extending the number of approximation methods under comparison, and by considering additional tests to address the error significance.

.

References

[1] M. Abudy and Y. Izhakian, "Pricing stock options with stochastic interest rate," pp. 250–277, International Journal of portfolio Analysis and Management, 1, 2013.

[2] Y. Ait-Sahalia, "Nonparametric pricing of interest rate derivative securities.," pp. 527–560, Econometrica, 1996.

[3] G. Aldashev, T. Carletti, and S. Righi, "Follies subdued: Informational efficiency under adaptive expectations and confirmatory bias," pp. 110–121, Journal of Economic Behavior & Organization 80(1), 2011.

[4] M. J. A G.P. Annaert, J.and Claes, D. Ceuster, and H. Zhang., "Estimating the spot rate curve using the nelsonsiegel model. a ridge regression approach," pp. 482–496, International Review of Economics Finance, 2013.

[5] G. Barone-Adesi and R. Whaley, "Efficient analytic approximation of american option values.," pp. 301–320, Journal of Finance, 42, 1987.

[6] J. Barunik and B. Malinska., "Forecasting the term structure of crude oil futures prices with neural networks," pp. 366379, Applied Energy, 2016.

[7] M. H. Beale, H. M. T., and H. B. Demuth, "Neural network toolbox, users guide," 2014.

[8] C. Bisiere, D. J. P., and S. Lovo, "Risk attitude, beliefs updating, and the information content of trades: an experiment.," pp. 1378–1397, Management Science 61(6), 2014.

[9] P. Bjerksund and G. Stensland, "Closed-form approximation of american options," pp. 87–99, Scandinavian Journal of Management, 9, 1993.

[10] P. Bjerksund and G. Stensland, "Closed-form valuation of american options, norwegian school of economics and business administration," pp. Department of Finance and Management Science, Discussion Paper, 9, 2002.

[11] J. J. F. Bonnans, Gilbert, C. Lemarchal, and C. Sagastizbal, "Numerical optimization, theoretical and numerical aspects," pp. 366379, Springer, second edition, 2006.

[12] S. K. Bose, J. Sethuraman, and S. Raipet., "Artificial neural networks," pp. chapter 8, pages 124138, Finance and Manufacturing, editors, IGI Global, 2006.

[13] M. P. Bowden, "A model of information flows and confirmatory bias in financial markets.," pp. 197–215, Decisions in Economics and Finance 38(2), 2015.

[14] P. Boyle, "Option valuation using three-jump process," pp. 7–12, International Options Journal, 3, 1986.

[15] W. A. Brock and C. H. Hommes, "Heterogeneous beliefs and routes to chaos in a simple asset pricing model," pp. 1235–1274, Journal of Economic dynamics and Control 22(8–9), 1998.

[16] D. Broomhead and D. Lowe, "Multivariate function interpolation and adaptive networks.," pp. 321355, Complex Systems, 2, 1988.

[17] G. Charness and C. Dave, "Confirmation bias with motivated beliefs," pp. 1–23, Games and Economic Behavior 104, 2017.

[18] X. Chen, "American option pricing formula for uncertain financial market," pp. 58–62, Proceedings of the First International Conference on Uncertainty Theory, August 11-19, 2010.

[19] J. Cox, S. Ross, and M. Rubinstein, "Option pricinga simplified approach.," pp. 229–263, 1979, Journal of Financial Economics, 7.

[20] C. Chiarella, R. Dieci, and L. Gardini, "Speculative behaviour and complex asset price dynamics: a global analysis.," pp. 173–197, Journal of Economic Behavior & Organization 49(2), 2002.

[21] C. Chiarella, R. Dieci, and X. Z. He, "Heterogeneity, market mechanisms, and asset price dynamics.," pp. 277–344, In Handbook of financial markets: Dynamics and evolution. North–Holland., 2009.

[22] C. Cortes and V. Vapnik, "Support-vector networks," pp. 321355, Machine Learning, 20(3), 1995.

[23] M. Cottrell, E. de Bodt, and P. Gregoire, "Interest rates structure dynamics: A non parametric approach," 1995.

[24] M. Cottrell, E. de Bodt, and P. Gregoire, "Examining the nelsonsiegel class of term structure models," pp. Technical Report 043/4, Tinbergen Institute, 2007.

[25] V. D'Amato, D. L. E., D. A., N. E., and S. M., "The impact of the discrepancies in the yield curve on actuarial forecasting.," pp. 49–55, Proceedings of the XVI Iberian Italian Conference on Financial and Actuarial Mathematics–

XVI Iberian Italian Conference on Financial and Actuarial Mathematics, 26-27 Maggio 2016.

[26] R. H. Day and W. Huang, "Bulls, bears and market sheep," pp. 299–329, Journal of Economic Behavior & Organization 14(3), 1990.

[27] P. De Grauwe and M. Dewachter, H.and Embrechts, "Exchange rate theory: Chaotic models of foreign exchange markets.," p. Oxford: Blackwell, 1993.

[28] R. de Rezende and M. Ferreira, "Modeling and forecasting the yield curve by an extended nelson-siegel class of models: A quantile autoregression approach," pp. 111123, Journal of Forecasting, 32, 2013.

[29] F. Diebold and C. Li, "Forecasting the term structure of government bond yields," pp. 337364, Journal of Econometrics, 130(2), 2006.

[30] F. Diebold and G. D. Rudebusch, "The dynamic nelsonsiegel approach to yield curve modeling and forecasting," pp. 337364, Journal of Econometrics, 130(2), Princeton University Press, 2013.

[31] E. A. Association, "Negative interest rates and their technical consequence," 2016.

[32] E. E. Securities and M. Authority), "Joint committee discussion paper on the use of big data by financial institutions," From 19 December 2016 to 17 March 2017.

[33] E. Fama and R. Bliss, "The information in long maturity forward rates," pp. 680692, The American Economic Review, 77(4), 1987.

[34] F. stability Board, "Financial stability implications from fintech supervisory and regulatory issues that merit authorities attention," June 2017.

[35] A. Gaunersdorfer and C. Hommes, "A nonlinear structural model for volatility clustering.," pp. 265–288, Long memory in economics Springer, Berlin, Heidelberg., 2007.

[36] M. Gilli, S. Grosse, and E. Schumann, "Calibrating the nelsonsiegel svensson model," pp. 680692, The American Economic Review, 77(4), 2010, Technical report, COMISEF.

[37] P. Giribone, S. Ligato, and O. Caligaris, "Applicazioni delle reti neurali feed-forward per la ricostruzione di superfici di volatilit," 2015, AIFIRM Magazine, Associazione Italiana Financial Industry Risk Managers.

[38] P. Giribone and O. Caligaris, "Modellizzare la curva dei rendimenti mediante metodologie di apprendimento artificiale: analisi e confronto prestazionale tra le tecniche regressive tradizionali e le reti neurali," 2015, AIFIRM Magazine, Associazione Italiana Financial Industry Risk Managers.

[39] G. P. G., O. Caligaris, and S. Fioribello, "Lalgoritmo della fuzzy c-means clustering come tecnica automatica per lindividuazione di anomalie di mercato," 2016, AIFIRM Magazine, Associazione Italiana Financial Industry Risk Managers, 2016.

[40] P. Giribone, S. Ligato, and F. Penone, "Combining robust dynamic neural networks with traditional technical indicators for generating mechanic trading signals," p. International Journal of Financial Engineering (IJFE), 2018.

[41] P. Giribone, S. Ligato, and M. Mulas, "The effects of negative interest rates on the estimation of option sensitivities: The impact of switching from a log-normal to a normal model.," pp. International Journal of Financial Engineering, 4, 1750015, 2017.

[42] P. Giribone and S. Ligato, "Considerazioni sullo stato attuale della valorizzazione delle opzioni cap e floor aventi come parametro di riferimento il tasso euribor," pp. 45–53, AIAF Newsletter, 99, 2016.

[43] P. Gogas, T. Papadimitriou, M. Matthaiou, and E. Chrysanthidou, "Yield curve and recession forecasting in a machine learning framework," pp. 635645, Computational Economics, 45, 2015.

[44] P. Gogas, T. Papadimitriou, M. Matthaiou, and E. Chrysanthidou, "Yield curve and recession forecasting in a machine learning framework," pp. 635645, Computational Economics, 45, 2015.

[45] Z. Grbac and W. Runggaldier, "Interest rate modeling: Post-crisis challenges and approaches," pp. 13541362, Mathematical and Computer Modelling, 55, Springer, 2014.

[46] W. Huang and R. Day, "Chaotically switching bear and bull markets: The derivation of stock price distributions from behavioral rules.," pp. 169–182, R. Day, P. Chen (Eds.), Nonlinear dynamics and evolutionary economics, Oxford University Press., 1993.

[47] W. Huang, H. Zheng, and W. M. Chia, "Financial crises and interacting heterogeneous agents.," pp. 1105–1122, Journal of Economic Dynamics and Control 34(6), 2010.

[48] X. He and Y. Li, "Power-law behaviour, heterogeneity, and trend chasing.," pp. 3396–3426, Journal of Economic Dynamics and Control 31(10), 2007.

[49] M. Henrard, "Interest rate modelling in the multicurve framework," p. Palgrave McMillan, 2014.

[50] C. Hommes, "Behavioural rationality and heterogeneous expectations in complex economic systems," pp. 1–38, Cambridge University Press, 2013.

[51] C. Hommes and F. Wagener, "Complex evolutionary systems in behavioral finance.," pp. 217–276, Handbook of financial markets: Dynamics and evolution, 2009.

[52] J. Hull, "Options, futures and other derivatives. 9th edition, prentice hall, upper saddle river," 2014.

[53] J. Hull and A. White, "Numerical procedures for implementing term structure models i," pp. 7–16, Journal of Derivatives, 2, 1994.

[54] J. Hull, "Options, futures and other derivatives," 2012, 8th Edition, Prentice-Hall, Upper Saddle River.

[55] J. Hull and A. White, "A generalized procedure for building trees for the short rate and its application to determining market implied volatility functions," pp. 443–454, Quantitative Finance, 15, 2015.

[56] K. Inui, "Improving nelson-siegel term structure model under zero/super-low interest rate policy," 2015.

[57] A. Joseph, M. Larrain, and E. Singh, "Predictive ability of the interest rate spread using neural networks," p. Palgrave McMillan, 2011, Procedia Computer Science, 6.

[58] T. Kohonen, "Self-organized formation of topologically correct feature maps," pp. 59–69, Journal of Biological Cybernetics, 1982.

[59] A. Kirman, "Epidemics of opinion and speculative bubbles in financial markets," pp. 354–368, Taylor, M., Money and financial markets. Oxford: Blackwell, 1991.

[60] T. Kohonen, "The self-organizing map," p. Journal of Neurocomputing, 1998.

[61] K. D., "A brief introduction to neural network," 2005.

[62] T. Kooiman, "Negative rates in financial derivatives," pp. Master thesis, Amsterdam University, Amsterdam., 2015.

[63] F. E. Kydland and E. C. Prescott, "Time to build and aggregate fluctutations," pp. 1345–1370, Econometrica, Vol. 50, No. 6, 1982.

[64] E. Larsson, S. M. Gomes, A. Heryudono, and A. Safdari-Vaighani, "Radial basis function methods in computational finance," p. In Proceedings of the 13th International Conference on Computational and Mathematical Methods in Science and Engineering, 2013.

[65] M. Leippold and Z. Wiener, "Efficient trinomial trees for short rate models," pp. 213–239, Review of Derivative Research, 7, 2004.

[66] T. Lux, "Herd behaviour, bubbles and crashes.," pp. 881–896, The Economic Journal, 1995.

[67] T. Lux, "The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions.," pp. 143–16 Journal of Economic Behavior & Organization 33(2), 1998.

[68] T. Lux, "Stochastic behavioral asset-pricing models and the stylized facts," pp. 161–215, Handbook of financial markets: Dynamics and evolution. North–Holland, 2009.

[69] T. Lux and M. Marchesi, "Scaling and criticality in a stochastic multi-agent model of a financial market," p. Nature, 1999.

[70] T. Lux and M. Marchesi, "Volatility clustering in financial markets: a microsimulation of interacting agents," pp. 675–702, International Journal of Theoretical and Applied Finance, 2000.

[71] L. Menkhoff and M. P. Taylor, "The obstinate passion of foreign exchange professionals: technical analysis," pp. 936–972, 2007, Journal of Economic Literature 45(4).

[72] C. Nelson and A. Siegel, "Parsimonious modeling of yield curves.," pp. 473–489, P Journal of Business. 60, 1987.

[73] M. Nardon and P. Pianca 2009, pages = 1-21, WORKING PAPER SERIES, Venezia, Department of Applied Mathematics, University Ca' Foscari of Venice, title = Implied volatilities of American options with cash dividends: an application to Italian Derivatives Market (IDEM), booktitle = WORKING PAPER SERIES, Venezia, Department of Applied Mathematics, University Ca' Foscari of Venice.

[74] C. Nelson and A. F. Siegel, "Parsimonious modeling of yield curves.," pp. 473489, In Journal of Business, 60, 1987.

[75] R. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," pp. 175–220, Review of General Psychology 2, 1998.

[76] J. Park, P. Konana, B. Gu, A. Kumar, and R. Raghunathan, "Information valuation and confirmation bias in virtual communities: Evidence from stock message boards.," pp. 1050–1067, Information Systems Research 24(4), 2013.

[77] "Negative interest rate policies: Sources and implications," p. Policy Research Working Papers, August 2016.

[78] S. Pouget, J. Sauvagnat, and S. Villeneuve, "A mind is a terrible thing to change: confirmatory bias in financial markets.," pp. 2066–2109, The Review of Financial Studies 30(6), 2017.

[79] J. Principe, D. Xu, and J. Fisher, "Information theoretic learning.," pp. Chapter 7, Unsupervised Adaptive Filtering, volume I, John Wiley Sons, New York, 2000.

[80] M. Rabin and J. L. Schrag, "First impressions matter: A model of confirmatory bias.," pp. 37–82, The Quarterly Journal of Economics 114(1), 1999.

[81] M. Recchioni, Y. Sun, and G. Tedeschi, "Can negative interest rates really affect option pricing? empirical evidence from an explicitly solvable stochastic volatility model," pp. 1–19, Quantitative Finance, 2017.

[82] R. Roll, "An analytic valuation formula for unprotected american call options on stocks with known dividends.," pp. 251–258, 1977, Journal of Financial Economics, 5.

[83] D. Rosadi, Y. Nugraha, and R. Dewi, "Forecasting the indonesian government securities yield curve using neural networks and vector autoregressive model," pp. 473489, Technical report, Department of Mathematics, Gadjah Mada University, Indonesia, 2011.

[84] R. Sambasivan and S. Das, "A statistical machine learning approach to yield curve forecasting," pp. Technical report, Chennai Mathematical Institute, 2017.

[85] C. Smith, "Option pricing: A review," pp. 3–51, 1976, Journal of Financial Economics 3(1-2).

[86] L. E. O. Svensson, "Estimating the term structure of interest rates for monetary policy analysis," pp. 163183, Scandinavian Journal of Economics, 98, 1996.

[87] J. Tappinen, "Interest rate forecasting with neural networks," pp. Technical Report 170, Government Institute for Economic Research, 1998.

[88] F. Tramontana and L. Westerhoff, F.and Gardini, "On the complicated price dynamics of a simple one-dimensional discontinuous financial market model with heterogeneous interacting traders," pp. 187–205, Journal of Economic Behavior & Organization 74(3), 2010.

[89] F. Tramontana and L. Westerhoff, F.and Gardini, "Heterogeneous speculators and asset price dynamics: further results from a one-dimensional discontinuous piecewise-linear map," p. Computational Economics 38(3), 2011.

[90] F. Tramontana and L. Westerhoff, F.and Gardini, "A simple financial market model with chartists and funadamentalists: Market entry levels and discontinuities.," pp. 16–40, Mathematics and Computers in Simulation108, 2015.

[91] O. Vasicek, "An equilibrium characterization of the term structure," pp. 177188, Journal of Financial Economics, 5, 1977.

[92] F. Westerhoff, "Multiasset market dynamics.," pp. 596–616, Macroeconomic Dynamics 8(5), 2004.

[93] F. Westerhoff, "Exchange rate dynamics: A nonlinear survey," pp. 287–325, Handbook of research on complexity, 2009.

[94] F. Westerhoff and R. Dieci, "The effectiveness of keynes–tobin transaction taxes when heterogeneous agents can trade in different markets: a behavioral

finance approach," pp. 293–322, Journal of Economic Dynamics and Control 30(2), 2006.