Dynamics of Commodity Prices
A Potential Function Approach
with Numerical Implementation

Giuseppe Ciociola

Contents

Chapter 1

Introduction

The main concern of the present analysis is an attempt to take into account the presence of several forces acting in commodity markets and the difficulty to disentangle their relative price impacts. The analysis starts from one specific characteristic feature, that is the tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others. Such characteristic feature refers to the so-called *price clustering phenomenon*. Price clustering is the phenomenon that some prices are more frequently observed than other prices.

Global commodity markets have experienced significant price swings in recent years. Analysts offer two general explanations. An emphasis on market forces postulates that market fundamentals have changed importantly, whereas an alternative explanation attributes the large changes in price to speculative expectations. Of course, these two explanations are not mutually exclusive, and both market forces and speculative expectations may be responsible. It is well known that the behavior of commodity prices is different from that of traditional financial assets (such as stocks and bonds). Hence, analytical and modelling tools that take into account specific features of commodity prices are needed. Recent developments on the significant and sharp rises and declines in commodity prices seem to indicate that various factors are acting in a very complex way, including geopolitical concerns. In particular, one specific characteristic feature is the tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others, leading to price clustering phenomenon. Explanations of the clustering phenomenon is a subject of extensive research, ranging from fundamental factors to mathematical nonlinear models of price dynamics. It has been noted that the study of commodity prices has long been something of an academic stepchild, and this sort of relative obscurity arguably reflects the niche role of commodities in the broader financial markets. But commodities are in the process of becoming mainstream. In the literature, there are two main approaches which are used to explain the dynamics of commodity price process: structural models and reduced-form models. In the current literature and practice, the commodity price behavior essentially relates to the well known property of mean-reversion. Indeed, mean-reverting class of diffusion models have been widely used to model commodity prices. However, these techniques of analysis are not able to model the phenomenon of multiple attraction regions.

In order to overcame such limitations, our analysis continue to discuss the idea concerning the *potential function model* approach. Assuming that the different variables (forces) acting on the markets are expressed by a vector-valued function, we show how it is possible to construct a corresponding potential function, and how it is able to completely describe all the properties of the original vector field. Starting from a potential function, we are now interested in constructing a dynamic system, which will be defined in terms of the minus gradient of the potential function. Using the concept of potential function, in the framework of this deterministic gradient system we shall attempt to correlate potential function with the different nature of equilibrium position. In order to appreciate the role of potential

function about equilibrium, we need to visualize equilibrium in the light of some external disturbances. In this setting, the analysis of dynamics can be transferred from a vector field into the corresponding potential function by means of the gradient operator. In order to build a more realistic description of the dynamics, we need to consider an external disturbance and we can think of a complementary force in terms of a noise component. The conversion of a model from deterministic form to stochastic form can be formalized adding a diffusion term to our gradient system. The potential function model arises in the context of such a stochastic gradient system, so that what distinguishes the traditional stochastic differential equation from the present approach is that the drift term in the potential model has the special form expressed in terms of the minus gradient of potential function.

When we are interested in describing the dynamics of a certain phenomenon by means of a diffusion model with a potential, a basic issue is how to describe mathematically such an elusive potential function, assuming that such a function exists. Several functional forms of the potential may be considered, ranging from a simple monomial form to more sophisticated functional forms. In our analysis we adopt a data-driven procedure, learning a potential function from a price trajectory given by a collection of data.

In order to make the potential function approach stronger, it is possible to view randomness in an equilibrium perspective. Starting from the origin of Brownian motion, the interplay between Probability, Mathematical Statistics and Statistical Physics has recognized an extremely large and important work (Metropolis algorithm and Simulated Annealing as well refer to such an interplay of these three areas). In particular, the analogy is based on Boltzmann's deduction of equilibrium distribution of ideal gas placed in an external potential field which provides a way of viewing probability density from a perspective of forces/potentials, hidden behind it. By means of a simple heuristic model we provide some insight in order to explain how the potential model works, showing that this is in agreement with economic arguments. Finally, we consider to adopt the Boltzmann-Gibbs distribution, so as to identify (and subsequently estimate) the two main elements of the present model: the potential function and the diffusion parameter.

The present approach of potential model has a fundamental step on fitting the multimodal density of the invariant distribution. Starting from the observed price series, $\{p(t_i)\}_{i=1}^{N}$, we need to estimate the density function by means of fitting the resulting histogram of the historical data. In particular, the estimation method has to provide an analytical expression for the density, since the derivative of the potential needs to be quickly and accurately evaluated. In other words, we will see that for the use of the present model is needed the derivative of the potential function, so the estimation method has to have the advantage that such a derivative does not have to be estimated separately, but can be computed directly from the expression provided by the method. The multimodal density can be estimated in numerous ways, ranging from nonparametric methods to semi-parametric and parametric methods.

Our main contribution in the present analysis is to extend the original approach. We postulate a parametric form of the invariant price distribution in the framework of finite mixture models and fit the potential by means of the maximum likelihood method with a numerical implementation of Expectation-Maximization algorithm for a finite mixture of Gaussians. This way of fitting the potential model extends the original approach, and is a first step, so as it is particularly useful in the case of the multivariate extension of the model.

Our main concern is to consider the framework of finite mixture models in the context of cluster analysis, and the Expectation-Maximization algorithm and its applications to parameter estimation for mixture models from the perspective of latent variables. Mixture models have experienced increased interest and popularity over last decades. The importance of mixture distributions, their enormous developments and their frequent applications over recent years is due to the fact that mixture models offer natural models for *unobserved population heterogeneity*. Suppose that a parametric density $f(x;\theta)$

is capable to describe the phenomenon of interest, where $\theta \in \Theta$ denotes the parameter of the population, whereas x is in the sample space $\mathcal{X} \subset \mathbb{R}$. We call this the *homogeneous case*. However, often this model is too strict to capture the variation of the parameter over a diversity of subpopulations. In this case, we have that population consists of various subpopulations. We call this situation the *heterogeneous case*. In contrast to the homogeneous case, we can consider the same type of density in each subpopulation, but a potentially different parameter. If we consider a sample dataset, here it is not observed which subpopulation the observations are coming from. Therefore, we speak of *unobserved heterogeneity*. We develop such a heterogeneity in the framework of cluster analysis numerical methods. In most applications of cluster analysis a partition of data is sought, in which each individual or object belongs to a single cluster, and the complete set of clusters contains all individuals. The purpose of cluster analysis is to determine the inner structure of clustered data when no information other than the observed values is available. Most clustering done in practice is based largely on heuristic or distance-based procedures, such as hierarchical agglomerative clustering or iterative relocation procedures. Clustering methods based on probability models offer a principal alternative to heuristic-based algorithms. In this context the data are viewed as coming from a mixture of probability distributions, each representing a different cluster. Interest in clustering has increased due to the emergence of new domains of application, such as astronomy, biology, physics and social sciences. In addition to clustering purposes, finite mixtures of distributions have been applied to a wide variety of statistical problems such as discriminant analysis, image analysis and survival analysis. To this extent finite mixture models have continued to receive increasing attention from both theoretical and practical points of view. In order to estimate the parameters of a mixture model we implement the numerical technique Expectation-Maximization, which is one of the most frequently used algorithms for finding maximum likelihood estimators in mixture models. In order to take into account the unobserved heterogeneity, the implemented algorithm refers to the latent variables perspective of mixture distributions in which the discrete latent variables can be interpreted as defining assignments of data points to specific components of the mixture.

The procedure for identifying and estimating the two main elements of the present model - the potential function and the diffusion parameter - is provided. In order to estimate and fit the potential function, we consider a scaled potential and postulate a parametric form of the invariant distribution based on the framework of finite mixture model. The estimate of the diffusion parameter refers to a preliminary discretization of the equation model, together with a combination of the effects of time discretization and random disturbances. In such a way, the new unknown parameter can be estimated by means of a regression procedure. The estimated model is tested in various ways. The analysis of the model residuals is important in order to capture main dependence characteristics of the observed data, and it is related to the estimate of the diffusion parameter and the corresponding changes in volatility within the variance of the model residuals. The performance of the fitted model and its prediction accuracy is tested using the mean square prediction error in a context of cross-validation procedure. More important, the model is tested in terms of predicting the direction of the next price move, by means of the correct up-down moves. Concerning the volatility measures, the potential function model is able to provide an estimate of such an unobserved parameter, so we briefly address to this estimate which can be used as a valid alternative to more traditional techniques. The model is able to generate copies of the observed price series with the same distributional properties, which is useful for applications such as Monte Carlo analysis, scenario testing, and other studies that require a large number of independent price trajectories. We provide numerical schemes in order to simulate price process. Finally, by using an approach based on Monte Carlo simulations, we implement a goodness-of-fit procedure in order to access the validity of the model obtained by testing if there is a lack-of-fit between the stochastic differential equation model and the collection of data used to estimate the drift and the diffusion.

An underlying assumption of the present model is that the potential function and the long-term volatility do not change with time. As we have already said, global commodity markets have experienced significant price swings in recent years. In such a context of new market conditions, new attraction regions can form, changing the shape of the potential and the magnitude of the long-term volatility. In order to investigate further the behavior of the potential model, we have considered a price dataset to ensure the availability of as long a span of high quality data as possible. At first we employ daily spot price data concerning crude oil. Additionally, we have also tested the behaviour of the present model with an agricultural commodity, using soybean data.

A numerical implementation of the present analysis is provided. The implemented code concerns the potential function model, the Expectation-Maximization algorithm, the testing performance of the estimated model, the numerical schemes for simulating price trajectories, the goodness-of-fit test for the stochastic differential equation model.

This work is organized as follows. In Chapter 2 we address to the main hypothesis for explanations of recent commodity price developments, the phenomenon of price clustering and the modelling of commodity price dynamics. In Chapter 3 we introduce the potential function in the context of stochastic gradient system, discuss its different kind of equilibrium levels and how the potential model works in agreement with economic arguments. In Chapter 4 we outline the framework of finite mixture models in the context of cluster analysis, and the Expectation-Maximization algorithm and its applications to parameter estimation for mixture models from the perspective of latent variables. In Chapter 5 we provide the procedure for identifying and estimating the parameters of the model, and apply the procedure to crude oil and soybean prices throughout different periods. Finally, Chapter 6 is devoted to the code developed in order to provide a numerical implementation of the present analysis.

Chapter 2

Commodity Markets and Price Clustering

Global commodity markets have experienced significant price swings in recent years. Analysts offer two general explanations. An emphasis on market forces postulates that market fundamentals have changed importantly, whereas an alternative explanation attributes the large changes in price to speculative expectations. Of course, these two explanations are not mutually exclusive, and both market forces and speculative expenditures may be responsible. It is well known that the behavior of commodity prices is different from that of traditional financial assets (such as stocks and bonds). Hence, analytical and modelling tools that take into account specific features of commodity prices are needed. Recent developments on the significant and sharp rises and declines in commodity prices seem to indicate that various factors are acting in a very complex way, including geopolitical concerns. In particular, one specific characteristic feature is the tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others. Such characteristic feature refers to the so-called *price clustering phenomenon*. Price clustering is the phenomenon that some prices are more frequently observed than other prices. Explanations of the clustering phenomenon is a subject of extensive research, ranging from fundamental factors to mathematical nonlinear models of price dynamics. It has been noted that the study of commodity prices has long been something of an academic stepchild, and this sort of relative obscurity arguably reflects the niche role of commodities in the broader financial markets. But commodities are in the process of becoming mainstream. In the literature, there are two main approaches which are used to explain the dynamics of commodity price process: structural models and reduced-form models. In the current literature and practice, the commodity price behavior essentially relates to the well known property of mean-reversion. Indeed, mean-reverting class of diffusion models have been widely used to model commodity prices. However, these techniques of analysis are not able to model the phenomenon of multiple attraction regions.

2.1 Commodity markets, prices and recent developments

Global commodity markets have experienced significant price swings in recent years. Commodity prices are determined by fundamental supply and demand conditions in physical commodity markets. In the last decade these market fundamentals have changed importantly related to increasing demand for commodities from highly growing emerging countries, alternative uses of commodities for energy production (biofuels), and a reduction in supply due to supply constraints and stagnation in production and productivity related to low investments in the last two decades. Simultaneously to these fundamental supply and demand related changes, trading activities on commodity markets have undergone major changes with the increasing presence of financial investors, including banks, institutional in-

vestors and hedge funds. Trading volumes on commodity derivative markets and the share accounted for by financial investors have increased sharply, particularly since 2005.

In order to give some insights concerning the presence of several forces acting in commodity markets and the difficulty to define a clearly and efficient price formation, in this section we are interested to briefly address to the main hypothesis for explanations of recent commodity price developments.

2.1.1 Recent developments in commodity prices

Over the last decade global commodity markets have experienced significant price swings. In the post-war period, primary commodity prices experienced several cycles (Radetzki 2006 [72]). Prices were generally high in the 1950s in the context of the Korea war while they were low in the 1960s. In light of the two oil price shocks in the 1970s commodity prices increased again. More recently, after two decades of low commodity prices in the 1980s and 1990s, the prices of a wide range of commodities have registered historic price increases. The mid-2000s marked the start of a trend of steeply rising commodity prices. In the late 1990s and particularly since 2002/03, many commodities have registered steep price increases culminating in a peak in mid-2008. The price boom between 2002 and mid-2008 was the most pronounced in several decades. However, in mid-2008 prices fell sharply. This price decline following the eruption of the current global crisis in the second half of 2008 stands out both for its sharpness and for the number of commodities affected. The World Bank noted that commodity prices had lost in a matter of two months in the last quarter of 2008, most of the increase of the preceding 24 months (World Bank 2009 [89]). Since the first half of 2009, global commodity prices have been rising again, with the speed of the rise accelerating since the fall of 2010. Especially food prices reached an all time high in February 2011. While the timing varied for different types of commodities the surge in prices, the sharp correction and the subsequent rebound affected all major commodity categories, including agricultural, metals and energy commodities.

The recent developments in commodity prices between 2008 and 2011 have been extraordinary with respect to its short duration, amplitude and coverage of commodities. These developments have also been exceptional in many ways (UNCTAD 2011 [87]).

The current evolution of commodity prices reflects significant changes in fundamental demand and supply relationships. In contrast to earlier price cycles that were primarily triggered by supply shocks of specific commodities, the recent changes are largely related to demand factors affecting a broad range of commodities (Kaplinsky 2010 [54], Nissanke 2011 [65]). The rapid growth in major emerging economies, particularly China and India among other emerging countries, has led to a rapidly increase in their demand for commodities, particularly since the turn of the century. This rising demand has been driven by heavy investments in infrastructure, increasing urbanization and industrialization, the materials utilized in manufactures and the growing consumption of energy. In recent years, crude oil prices have climbed to unprecedented levels, reaching an all time high of nearly 150 per barrel in July 2008. In the wake of the financial crisis of 2008-2009, oil prices fell below 40 per barrel at the end of 2008. It is often argued that the fast-growing Asian emerging economies are a major source of rising demand for crude oil. The higher energy intensity of their production compared to that of developed economies has contributed decisively to the growing demand (ECB 2010 [26]). This demand slowed down only temporarily as a result of the recent crisis. Moreover, the strong surge in oil prices in recent years cannot be explained without taking into account the role of the supply side. A shift in the supply relations between non-OPEC and OPEC countries can thus be assumed to have a significant impact on the evolution of oil prices. The sudden slowdown in the growth rate of non-OPEC crude oil supply after 2004 is therefore seen as a major factor driving oil price developments (Kaufmann 2011 [55], ECB 2010 [26]). It caused an unexpected increase in OPEC's capacity utilization, lowering OPEC's excess capacity and thus strengthening the role of the cartel as a marginal supplier. Recent

oil price increases are likely to have been accelerated by political tensions and armed conflicts in oil-producing countries, among other factors, although the effect may have been dampened to some extent by declining inventories.

Recent developments in commodity prices and emerging economies are also associated by means of a growing middle class with changing dietary habits, including an increasing food consumption of meat and dairy products, as incomes rise (Farooki and Kaplinsky 2011 [30]). Grain prices have been very volatile in the most recent years. Having peaked in 2008, they declined sharply, but started rising again in 2010. In February 2011 maize prices exceeded the level of June 2008. A number of supply and demand factors contribute to rising food commodity prices. Supply growth is slowing, because agricultural land is limited and productivity growth has slowed. Supply constraints are exacerbated by the effects of climate change (such as extreme weather events), which are already felt in many regions of the world, but are expected to grow dramatically over the next decades. On the demand side, the rising world population and changes in emerging economies towards more protein-rich diets are major long-term factors. As incomes in emerging economies have risen sharply with accelerated economic growth, consumption patterns of the population have also changed. Between 1995 and 2005, world meat consumption rose by 15 percent, East and Southeast Asia being the region with the highest increase at almost 50 percent . Taking into account that the production of 1 kg of meat requires about 7 kg of grains, the impact on grain demand is substantial.

Biofuel production is another decisive demand factor. The decision by some governments to introduce blending requirements and subsidies for biofuel production is considered to play a significant role in the recent price hikes of grains. Biofuel production also affects price movements of agricultural products which are not used in the production of biofuels, because agricultural land is diverted to producing crops needed for biofuel production. As biofuels partly replace petroleum products, they strengthen the link between the oil market and markets of agricultural products used in the production of biofuels (i.e. maize, sugar, oilseeds and palm oil). High oil prices also affect agricultural commodity prices via higher production costs, especially for energy and fertilizers. This may also explain the co-movement of oil prices and some agricultural commodity prices.

In the short run, weather effects have a strong impact on price developments. Often, these are exacerbated by policy measures such as export bans or taxes. Thus, wheat prices were driven up last August by the drought in the Russian Federation and an export ban.

Commodity prices have also been extremely volatile, in many instances with no obvious link to changes on the supply side. Price volatility has long been recognized as a major feature of commodity markets. Indeed, high price volatility has for long been a feature of commodity prices related to specific characteristics and specific shocks of commodities, especially on the supply side of food commodities, and in this way they have generally played a key role in this respect. Although the particular reasons for commodity price volatility differ by commodity, one important common factor is low short-run elasticities of supply and demand which means that any shock in production or consumption (that are frequent for many physical commodities) translates into significant price fluctuations as demand and supply cannot adjust quickly. Commodity price volatility tends to have significant adverse effects. At the macroeconomic level, it can lead to a deterioration in the balance of payments and in public finances, and the associated uncertainty is likely to curtail investment and to significantly depress long-term growth. At the microeconomic level, high and volatile commodity prices have severe impacts on the most vulnerable, especially food- and energy-insecure households. Rapidly growing demand for commodities, especially in emerging economies, as well as the debate about the future use of fossil fuels in the light of global climate change, and about the link between agricultural production and climate change more generally, have clearly had an impact on recent commodity price developments beyond simple commodity-specific shocks. However, since commodity prices have moved largely in tandem across all major categories over the past decade or so, the question arises as to whether the

very functioning of commodity markets has changed.

Unlike in earlier periods, the recent price hikes have occurred in an environment of general price increases across a wide range of commodities, from energy to agricultural commodities. Most of the factors which are often cited as price drivers, such as population growth or changing consumption patterns have been at work for an extended period often coinciding with low commodity prices. Their role in explaining recent price hikes is therefore doubtful. Experiences with the weak forecasting performance of econometric models for oil prices based on fundamentals also suggest that physical supply and demand are not the only factors that drive oil prices (Kaufmann 2011 [55]). The European Commission has also expressed doubts that market fundamentals are the main drivers of commodity prices (EC 2008 [25]). As the following sections show, there is strong evidence that the increasing presence of financial investors in commodity markets plays an important role in price dynamics.

2.1.2 The financialization of commodity markets and price formation

The term of "financialization of commodity trading" is understood to mean the increasing and expanded role of financial motives, financial markets, financial actors and financial institutions in the operation of commodity markets.

Traditional actors on spot and commodity derivative markets are commercial traders, which are actual producers and consumers of commodities that buy or sell on spot markets and try to reduce the related price risks through hedging on future markets. Non-commercial traders are also traditional in these markets, where they do not have an underlying physical commodity position to hedge but take over the price exposure from hedgers in exchange for a risk premium and are hoping to profit from changes in futures prices. Buying future contracts without having an underlying physical position to hedge is considered speculation and not investment. These speculators provide an essential function as they accept price risks in exchange for providing liquidity by actively trading in futures. Until recently, speculators on commodity future markets were dominated by experts of physical markets whose activities were closely linked to the fundamental supply and demand dynamics in the underlying physical markets (Masters and White 2008 [59]).

Over the last two decades and in particular since the early 2000s a major new element in commodity markets is the greater presence of financial investors, a third category of actors that treat commodities as an asset class. Financial investors can be divided into two main groups. The first group consists of index investors with a longer-term horizons. They are institutional investors such as pension funds, sovereign wealth funds, university endowments, public and private foundations and life insurance companies that follow passive trading strategies based on the assumption that commodities have a unique risk premium and form a relatively homogenous class. They generally invest in commodity indexes that are composites of future contracts of a broad range of commodities. The two largest ones are the Standard & Poor's Goldman Sachs Commodity Index and the Dow Jones-Union Bank of Switzerland Commodity Index. Index investors invest in a broad basket of commodities without taking into account the supply and demand fundamentals of individual commodities. Their trading strategy is based on holding long forward positions and taking advantage of the long-term increase in commodity prices. The second group of financial investors consists of financial intermediaries with much shorter time horizons. They are money managers including a range of investors, most importantly hedge funds, floor traders and institutional investors (Farooki and Kaplinsky 2011 [30]). They follow more active trading strategies and take positions on both sides of the market (long and short) which enables them to earn positive returns in rising and declining markets (Mayer 2009 [61]). Their investments are generally smaller in size compared to index investors and characterized by the frequency of their transactions seeking to take advantage of arbitrage and speculation opportunities. As index investors,

the trading activities of money managers are not based on the supply and demand fundamentals of individual commodities.

The fact that these market participants do not trade on the basis of fundamental supply and demand relationships, and that they may hold, on average, very large positions in commodity markets, implies that they can exert considerable influence on the functioning of those markets. Indeed, the greater participation of financial investors may have caused commodity markets to follow the logic of financial markets more closely, than that of a purely goods market.

Financial investors have long been active on commodity markets. This is evidenced by the frequently quoted examples of commodity price bubbles created by financial investors, including the tulip mania in Holland in the 1630s, the Mississippi Bubble in France and the South Sea Bubble in England in the early 1700s (Garber 1990 [32]). Over the last decade some factors are particularly important for the increasing involvement of financial investors. Broader developments in financial markets following the bursting of the equity market bubble in 2000 (dot-com crisis) and more pronounced in the global financial crisis of 2008 led financial investors to search for new investment opportunities given the losses and low returns in traditional investments such as stocks and bonds. This context spurred financial investment in commodities. Financial investors have been engaging in commodities trading for purposes of portfolio diversification ever since it became evident that commodity futures contracts exhibited the same average returns as investments in equities, while over the business cycle their returns are perceived negatively correlated with those on equities and bonds. Moreover, the returns on commodities were less volatile than those on equities or bonds, because the pair-wise correlations between returns on futures contracts for various commodities were relatively low. In such a way, there was growing acceptance of the notion that commodities as an asset class are a quasi-natural hedge against positions in equity markets (Buyuksahin et al. 2010 [21], Gorton and Rouwenhorst 2006 [37]). Such portfolio diversification considerations gained further impetus in the early 2000s with the increasing recognition in both academic circles (Radetzki 2006 [72]) and among potential investors (Heap 2005 [44]) that commodities were entering a new super cycle. It was believed that rapidly growing demand associated with emerging economies had triggered a new, prolonged increase in real commodity prices. Commodity futures contracts were also found to have good hedging properties against inflation, since their return was positively correlated with inflation (Bodie 1983 [12], Edwards and Park 1996 [27]). This is because they represented a bet on commodity prices, such as prices of energy and food products, which have a strong weight in the goods baskets used for measuring current price levels. Also, futures prices reflect information about expected changes in commodity prices, so that they rise and fall in line with deviations from expected inflation. Furthermore, investing in commodity futures contracts may provide a hedge against changes in the exchange rate of the dollar. Most commodities are traded in dollars and commodity prices in dollar terms tend to increase as the dollar depreciates. Measured in a currency basket, commodity prices are generally less correlated with the dollar (the sign of the correlation is reversed). This suggests that changes in the value of the dollar against other currencies may partly explain the negative correlation between the prices of dollar-denominated commodities and the dollar. Financial innovation has also played a facilitating role, as tracking commodity indexes is a relatively new phenomenon. Commodity market deregulation, such as enacted by the Commodity Futures Modernization Act of 2000, was a further facilitating factor.

In spite of the extent and the important implications of current commodity price dynamics, there is no consensus about the causes of these developments and how the increasing presence of financial investors has impacted on commodity prices (Irwin and Sanders 2011 [51]). The discussion involves a theoretical debate about how futures markets work and if speculation can move future prices and trigger speculative bubbles and an empirical debate about the factors behind the recent price developments. The impact of financial traders on commodity prices is difficult to quantify. Part of this difficulty is

due to the fact that the financialization of commodity trading became a major factor roughly at the same time as demand for physical commodities from emerging economies started to increase rapidly. These roughly simultaneous developments make it difficult to disentangle their relative price impacts as commodity prices are determined on the basis of expectation formation by heterogeneous market participants.

Accordingly, empirical studies come to different conclusions stating that financial investors have either no impact on future prices that are solely determined by fundamental factors (Sanders and Irwin 2010 [74]) or that they have moderate up to considerable impact on future as well as spot prices (Gilbert 2010 [36], Tang and Xiong 2010 [81]).

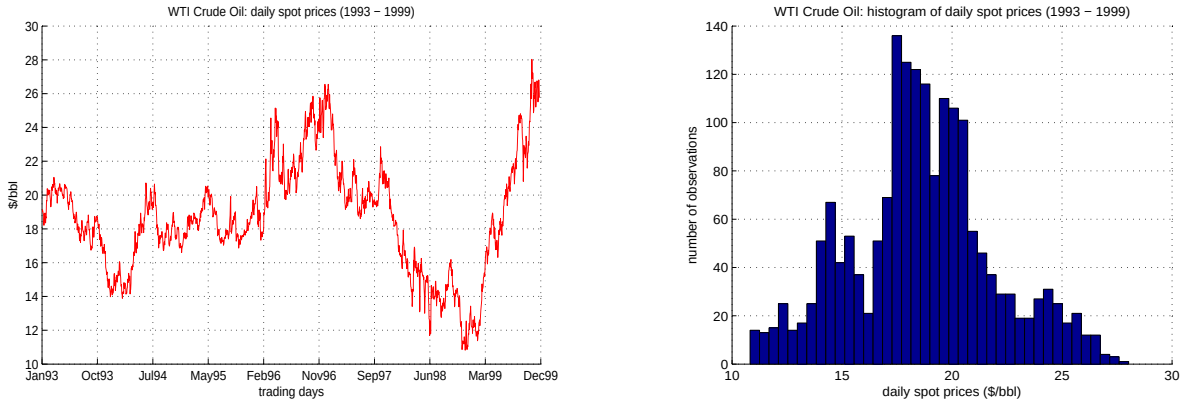
International institutions have also had different views on the role of financial investors in determining commodity prices. A special study on commodity markets by the World Bank (2009 [89]) or periodical analysis on commodity price developments reported in the World Economic Outlook by the International Monetary Fund (2009 [50]) interpret price dynamics basically in terms of fundamental demand and supply developments and do not consider the effects of financial investors. In contrast, the UNCTAD supports the financialization hypothesis stating that financial investors and their trading strategies can have sizable impacts on commodity prices, and that such a strong impact on prices may be considered the new normal of commodity price determination.

Concerning the practitioners, commodity future market traders generally agree that their trading activities have an effect on price developments as can be seen in the quotes from recent research reports that are cited by Henn (2011 [45]).

2.2 The price clustering phenomenon and possible explanations

It is well known that the behavior of commodity prices is different from that of traditional financial assets (such as stocks and bonds). Such factors as seasonal supply and demand, weather conditions, and storage and transportation costs, are specific to commodities. Hence, analytical and modelling tools that take into account specific features of commodity prices are needed. Moreover, we have just seen that recent developments on the significant and sharp rises and declines in commodity prices seem to indicate that various factors are acting in a very complex way, including supply-demand fundamentals, speculative market forces and geopolitical concerns. There is no consensus at this moment on which factor dominates the rise in commodity prices.

The main concern of the present analysis is an attempt to take into account the presence of several forces acting in commodity markets and the difficulty to disentangle their relative price impacts. The analysis starts from one specific characteristic feature, that is the tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others. Such characteristic feature refers to the so-called *price clustering phenomenon*. Price clustering is the phenomenon that some prices are more frequently observed than other prices. A striking example is the series of daily prices of crude oil, as seen from a plot of daily closing prices of West Texas Intermediate (WTI) crude oil daily spot prices over the trading days from January 4, 1993 to December 30, 1999 (Figure 2.1 (a)). The phenomenon of price clustering can be clearly seen on the resulting histogram of daily oil prices (Figure 2.1 (b)). The oil price is clustered around several preferred regions at approximately 12, 14, 17, 20, 24 dollars per barrel in that period, and most trading occurs there. In these regions the price is quite well stable, whereas it lies outside these regions only relatively briefly and its behavior is then rather unstable (16, 18, 23 dollars per barrel). Indeed, when the oil price fluctuates around 16 dollars per barrel, it will likely either drop towards 14 dollars or rise towards 17 dollars, depending on whether the market is falling (bearish) or rising (bullish) as a result of various market forces acting in that



(a) (b)

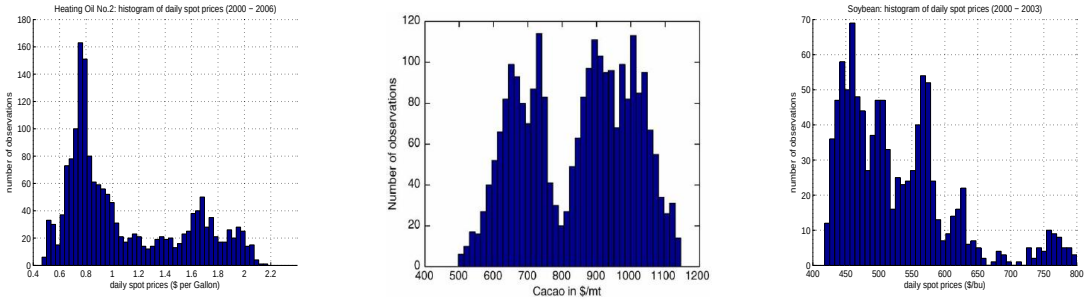
Figure 2.1: WTI crude oil daily spot prices over the trading days from January 4, 1993 to December 30, 1999. The observed time series (a) and the corresponding histogram (b), where can be clearly seen the phenomenon of price clustering.

period. Such price clustering is a well-known phenomenon in commodities markets, and traders know which price levels are more persistent than others. Spot as well as future prices of several agricultural commodities (coffee, cocoa, soybean) and other energy commodities (heating oil, gasoline) exhibit similar behavior (Figure 2.2).

Explanations of the clustering phenomenon is a subject of extensive research. Concerning fundamental factors, the explanation for the peaks in the histogram of prices of agricultural commodities can refer to seasonal effects, but a more important factor is alternating good and bad harvest years, whose occurrence is not periodic. Possible explanations in terms of fundamentals for energy commodities involve such factors as the global balance of oil supply and demand, OPEC quotas and target price bands, economic planning in the petroleum industry, the cyclic development of new exploration technologies, the strategic importance of oil. Economists are searching for theoretical explanation of the cluster formation in terms of macroeconomic factors, and there is a vast literature on the subject (see e.g. Cashin et al. 2002 [22], Pindyk 2001 [70], Adelman 2002 [1], Brook et al. 2004 [20]). Another, more empirical approach is to build mathematical nonlinear models of price dynamics with two or more attracting regions. This approach, more established in econometric literature, is based on the assumption of heterogeneous beliefs, where different expectations about the future price for different trader types can lead to complicated (chaotic) dynamics of prices with strange attractors and having such properties as multiple stable price levels which correspond to multiple attracting regions (see e.g. Hommes 2001 [48], Brock and Hommes 1997 [17] 1998 [18], Brock et al. 2001 [19], Gaunersdorfer et al. 2000 [33]).

2.3 Modeling the dynamics of commodity prices

It has been noted that the study of commodity prices has long been something of an academic stepchild, and especially in the domain of specific fields, notably agricultural commodities (Pirrong 2011 [71]). In finance, this sort of relative obscurity arguably reflects the niche role of commodities in the broader



(a) (b) (c)

Figure 2.2: The tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others: (a) heating Oil, (b) cocoa and (c) soybean exhibit price clustering phenomenon.

financial markets, as compared to more traditional equity and fixed-income markets. But it is easy to convince oneself that commodities are in the process of becoming mainstream. In the previous sections we have just seen that a confluence of forces has dramatically increased the importance of, and interest in, commodities and commodity prices. Moreover, this increase in the presence of investors and large financial intermediaries in commodity markets combined with extraordinary price movements in commodities to make commodity prices an important political issue. Unfortunately, the modeling of commodity prices has not kept pace.

2.3.1 Structural and reduced-form models

The issue of commodity pricing started with the work of Hoteling (1931 [49]) but especially with the seminal paper of Working (1933 [93]). This author was the first to identify spreads between spot and future prices as a measure of the return to storage and to derive a "supply of storage" curve relating these spreads to the amount of commodity in store. Working's research motivated the work of Kaldor (1939 [53]) who advanced the idea of a "convenience yield" for which those holding stocks receive an implicit stream of benefits from holding inventories (analogous to a dividend), and the marginal value of this stream is declining in stocks. Based on these ideas the theory of storage was introduced (Kaldor 1939 [53], Working 1949 [94], Brennan 1958 [14], Telser 1958 [83]). In its original formulation, the theory of storage links the commodity spot price with the contemporaneous futures price through a no-arbitrage relationship known as the "cost-of-carry model". This theory and its various and different versions was intuitively appealing and has informed much research in commodity markets (Pindyck 1994 [69], Williams and Wright 1991 [92]).

Following the seminal work of Samuelson (1965 [73]), it is now widely accepted that commodity prices fluctuate randomly. In the literature, there are two main approaches which are used to explain the dynamics of commodity price process: structural models and reduced-form models.

Structural models of commodity prices explicitly account for the implication of intertemporal optimization through storage. This approach dates from the pathbreaking work of Gustafson (1958 [40]) and has been developed further by many authors (Scheinkman and Schectman 1983 [75], Williams and Wright 1991 [92], among others).

A more recent reference on structural model approach is Pirrong (2011 [71]). The reduced-form

class of models has gained widespread acceptance and dominates the current literature and practice. In the present analysis the modeling of commodity prices refers to the reduced-form class of models. Reduced-form models explicitly specify the dynamics of a set of underlying state variables, such as the commodity spot price, the convenience yield or the instantaneous interest rate. Starting from the Black-Scholes option pricing formula, most practitioners have adopted reduced-form models to analyze commodity prices and price commodity derivatives. Probably the first research to apply the theory of real options to commodities is due to Brennan and Schwartz (1985 [15]), where they consider a mine producing a resource whose price can be modelled as a geometric Brownian motion (GBM) as follows

$$dP(t) = \alpha P(t)dt + \beta P(t)dB(t) \tag{2.1}$$

where $dP(t) = P(t + dt) - P(t)$ are the price increments, α and β are real constant, and $dB(t) = B(t + dt) - B(t)$ are the increments of a standard Brownian Motion. This model allowed the procedures developed for valuing financial options to be easily extended to valuing commodity based contingent claims. Another classic early commodities pricing model is due to Gibson and Schwart (1990 [35]), where the spot price is modelled as a GBM, and the convenience yield as a mean-reverting Ornstein-Uhlenbeck process. The convenience yield approach is also the motivation for one of the most widely used reduced-form commodity derivatives pricing models due to Schwartz (1997 [76]). However, Schwartz (1997 [76]) and Baker et al. (1998 [8]), among other, have emphasized the inadequacy of using GBM to model commodity prices. Under GBM the expected price level grows exponentially without bound.

In this more realistic context it is not unreasonable to expect that the workings of supply and demand will result in commodity prices that exhibit some sort of mean reversion. There is also empirical research that supports this claim. For example Bessembinder et al. (1995 [10]) find support for mean-reversion in commodity prices by comparing the sensitivity of long-maturity futures prices to changes in spot prices. If mean reversion is accepted as a desirable property, there are several possible stochastic models to choose from which incorporate mean reversion. The simplest mean reverting process, the Ornstein-Uhlenbeck process, is given as

$$dP(t) = \alpha(K - P(t))dt + \sigma P(t)dB(t) \tag{2.2}$$

where α is a real constant and referred to as the speed of mean reversion, the mean price level K denotes the long-run equilibrium price level (assumed constant) that P will tend towards, and σ is also a real constant and denotes the instantaneous volatility. Note that the equation model (2.2) is expressed in a common variation of the Ornstein-Uhlenbeck process, because the conditional variance of P depends on the level of P, thereby preventing P from becoming negative.

The mean reverting class of models of equation 2.2, while an improvement over geometric Brownan motion, are not entirely satisfactory. Another approach is the regime switching class of models. Initially proposed by Hamilton (1989 [42] and 1990 [43]), this approach employs a nonlinear model that admits the possibility of changes in regime, that is occasional discrete shifts in the parameters governing the behavior of the time series. In order to better capture the main characteristics, using a regime switching model the observed stochastic behavior of a specific time series is assumed to be comprised of several separate regimes or states. For each regime or state, one can define a separate and independent underlying stochastic process. The switching mechanism between each regime is typically assumed to be governed by an unknown random variable that follows a Markov chain. We have seen that various factors may contribute to the random shift between regimes, spanning from weather conditions to changes in government policies. In a regime switching model, spot prices can jump discontinuously between different states governed by state probabilities and model parameters. The regime switching model can be used to capture the shifts between “abnormal” and “normal” equilibrium states of supply and demand for a commodity.

Another class of models concerns the addition of other stochastic factors which may be important in modeling commodity prices. These models take into consideration the need to introduce the presence of jumps and/or to incorporate stochastic volatility. Jumps in commodity prices are often driven by discrete events such as weather, disease, or economic booms and busts which may persist for months or years.

In devising better models for commodity prices we are faced with a trade-off between increased realism through the addition of more stochastic factors, jumps, etc., and the added complexity and difficulty of solving the resulting equation model. The more factors incorporated into the model, the more complicated is the solution of the model. It is desirable to find an approach to modeling prices which, while adequately rich, still allows for the solution of the related model using more standard approaches.

2.3.2 Weakness of mean-reverting models

In the previous Section 2.2 we have seen that price commodities regularly move between attraction regions, although the time spent at a given region can be long and unpredictable. In particular, the price model has to be able to allow multiple attraction regions. This way, an observed time series is considered as a realization of a stochastic process with an invariant measure, that gives the relative frequency with which the process visits different regions of its allowed values. According to such a behavior the observations concentrated around some regions correspond to the underlying invariant measure. Hence, to model the phenomenon of multiple attraction regions, we need to mimic the invariant distribution of the underlying process.

In the current literature and practice, the commodity price behavior essentially relates to the well known property of mean-reversion. Indeed, we have just seen that mean-reverting class of diffusion models have been widely used to model commodity prices. However, these techniques of analysis are not able to model the phenomenon of multiple attraction regions. In the case of mean-reversion class of models the Gaussian distribution is the underlying invariant measure and has only one attraction point, so the models postulate the existence of just one price equilibrium level, and hance cannot generate processes with multiple attraction regions. Moreover, the rate of mean reversion is constant, so that the influence of the drift term is the same across all price regions. In the case of regime-switching class of models, such an approach allows for a non-constant mean reversion rate by specifying a finite number of regimes (two, three, ...) each with its own mean-reversion rate. But the big restriction is that each mean-reversion rate is still constant within each regime.

By contrast, the invariant measure of a clustering process has to be able to concentrates around a number of attraction points in order to be multimodal. Moreover, it is reasonable to assume a non-constant mean-reversion rate, and more general way of modelling this is to allow the mean-reversion rate to be a continuous function of the distance to the mean price level. This can be incorporated into the model by allowing for a larger class of possible drift forms. In the next chapter we analyze this idea, by specifying the drift term by means of a *potential function*. The resulting model is able to capture price behavior and overall characteristics of the data remarkably well, and it is as simple and tractable for estimation as traditional mean-reverting models. Moreover, the potential model allows for a continuously varying reversion rate providing a flexible way to model price. Then, it is more versatile, as it allows modelling of multiple stable price equilibrium levels (attracting regions) and encompasses a mean-reverting models as special cases. Allowing for a continuum of different reversion rates provides a richer model structure and significantly extends the regime-switching models. In a certain sense, the potential model is an extension of the regime-switching model, with a "continuum" of regimes and "continuum" of reversion rates, rather than a finite number of regimes each with its

own mean-reversion rate.

Rethinking the importance of commodity markets and the need to devise better models for commodity prices, one question arises. Since the idea behind the approach of a pricing model with a potential function approach is to take into account the different variables (forces) acting on the markets, can we think about this approach as a more natural bridge between structural models and reduced-form models?

Chapter 3

Potential Function Model

In the present chapter our main interest is to discuss the idea concerning the potential function model approach. Assuming that the different variables (forces) acting on the markets are expressed by a vector-valued function, we show how it is possible to construct a corresponding potential function, and how it is able to completely describe all the properties of the original function. Starting from a potential function, we are now interested in constructing a dynamic system, which will be defined in terms of the minus gradient of the potential function. Using the concept of potential function, in the framework of this deterministic gradient system we shall attempt to correlate potential function with the different nature of equilibrium position. In order to appreciate the role of potential function about equilibrium, we need to visualize equilibrium in the light of some external disturbance. In this setting, the analysis of dynamics can be transferred from a vector field into the corresponding potential function by means of the gradient operator. In order to build a more realistic description of the dynamics, we need to consider an external disturbance and we can think of a complementary force in terms of a noise component. The conversion of a model from deterministic form to stochastic form can be formalized adding a diffusion term to our gradient system. The potential function model arises in the context of such a stochastic gradient system, so that what distinguishes the traditional stochastic differential equation from the present approach is that the drift term in the potential model has the special form expressed in terms of the minus gradient of potential function. When we are interested in describing the dynamics of a certain phenomenon by means of a diffusion model with a potential, a basic issue is how to describe mathematically such an elusive potential function, assuming that such a function exists. Several functional forms of the potential may be considered, ranging from a simple monomial form to more sophisticated functional forms. In our analysis we adopt a data-driven procedure, learning a potential function from a price trajectory given by a collection of data. In order to make the potential function approach stronger, it is possible to view randomness in an equilibrium perspective. Starting from the origin of Brownian motion, the interplay between Probability, Mathematical Statistics and Statistical Physics has recognized an extremely large and important work (Metropolis algorithm and Simulated Annealing as well refer to such an interplay of these three areas). In particular, the analogy is based on Boltzmann's deduction of equilibrium distribution of ideal gas placed in an external potential field which provides a way of viewing probability density from a perspective of forces/potentials, hidden behind it. By means of a simple heuristic model we provide some insight in order to explain how the potential model works, showing that this is in agreement with economic arguments. Finally, we consider to adopt the Boltzmann-Gibbs distribution, so as to identify (and subsequently estimate - Chapter 5) the two main elements of the present model: the potential function and the diffusion parameter.

3.1 Force field and potential function

Starting from a vector-valued function F, in this section our main concern is to briefly discuss how it is possible to construct a corresponding *potential* function U, and by means of it how we are able to completely describe all the properties of the original function F.

Let us consider an open set $A \subset \mathbb{R}^n$ and suppose to define on A a vector-valued function or a vector field F, so that we write

$$\begin{aligned} F : A \subset \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto y = F(x) \end{aligned} \tag{3.1}$$

a correspondence associating to each element $x \in A$, a vector $x = (x_1, \ldots, x_n)$ of n components, one and only one (at most one) element $y \in \mathbb{R}^m$, a vector $y = (F_1(x), \ldots, F_m(x)) = F(x)$ of m components. In such a way, it is possible to express a vector-valued function F as follows

$$\begin{aligned} y_1 &= F_1(x_1, \ldots, x_n) \\ y_2 &= F_2(x_1, \ldots, x_n) \\ &\vdots \\ y_m &= F_m(x_1, \ldots, x_n) \end{aligned}$$

where the vector-valued F is expressed in terms of m scalar function $F_i : \mathbb{R}^n \to \mathbb{R}$, with $i = 1, \ldots, m$. A simple use of vector field in economics is a market of n goods where the demand function for each good depends on the corresponding n prices and the consumer's income, so that the vector field is $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$.

3.1.1 How a potential function arises

Starting from a vector field F, we are interested in constructing a *dynamic system*. In order to construct such a system, let us consider the coordinate axes in a Cartesian system for a three-dimensional Euclidean space $\mathbb{R}^3$, $O(x_1, x_2, x_3)$, and suppose the existence of a *material particle* (a sort of "ball", which may be an idealisation of an extended body) with coordinates $x = (x_1, x_2, x_3)$ that moves in this space. In such a way, we identify the position $P(x_1, x_2, x_3)$ of the particle in motion with the vector $x = (x_1, x_2, x_3)$, that is the particle moves from the origin, denoted by O which has coordinates $(0, 0, 0)$, to the actual position P. We can say that the particle moves through a *force field*. Therefore, such a force field can be expressed as a vector field, $F(x)$, which describes the force acting on a particle at various positions in space, where the force F depends on the vector position $x = (x_1, x_2, x_3)$ of the particle. Now suppose moving the particle from position P, which coordinates are $x = (x_1, x_2, x_3)$, to another position P' which coordinates are $x + dx = (x_1 + dx_1, x_2 + dx_1, x_3 + dx_1)$, where the vector $dx = (dx_1, dx_2, dx_3)$ describes the infinitesimal space interval (the displacement) along a generic path γ. Our objective is now to evaluate the work W done by force field F along the path. The total work done by force is equal to the sum of infinitesimal works calculated along the path, so that

$$\begin{aligned} dW &= F(x) \bullet dx \\ &= (F_1(x), F_2(x), F_3(x)) \bullet (dx_1, dx_2, dx_3)^T \\ &= F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3 \end{aligned} \tag{3.2}$$

where we have denoted by $\bullet$ the scalar product, the vector $(F_1(x), F_2(x), F_3(x))$ denotes the coordinates of the force, whereas the vector (dx_1, dx_2, dx_3) denotes the coordinates of the infinitesimal space interval

along the path γ. Integrating expression (3.2) we can write the total work of the force as

$$W_{PP'} = \int_P^{P'} F(x) \bullet dx \tag{3.3}$$

evaluated along the path $\gamma = PP'$.

In general, the total work $W_{PP'}$ is dependent on the starting point P and ending position P', and also on the specific path $\gamma = PP'$ taken from the particle. However, when the work is dependent only on the position of the particle and is independent of any path taken from point P to point P', then we say that the force field F is *conservative*. Thus a conservative force is a force with the property that the work done in moving a particle between two points is independent of the path taken.

Now we want to show that, for a given conservative force field F, it is possible to determine a scalar function such that it is able to summarize all its properties. Fix an arbitrary point O in the space and consider the work W done by force acting on the particle and moving it from fixed starting position O to ending position P, which we think as a moving point. Such a quantity is considered as a function of vector x representing the position P. Let us consider an open set $A \subset \mathbb{R}^n$, then any regular function in $C^2(A)$ taken the form

$$U : A \subset \mathbb{R}^n \to \mathbb{R} \tag{3.4}$$

$$x \mapsto U(x) = -\int_O^P F(x) \bullet dx \tag{3.5}$$

is called *potential function* associated with force field F. The potential function summarizes the work W done along the path $\gamma = OP$, where we have changed the sign. In the integration path one can arbitrary choose the starting point O so that the integral can be extended to any path from position O to new position P.

Thus the potential is a measure of the amount of work done on the particle and hence the ability of the particle itself to do work, that is to give back the work done on it. Potentials are immensely useful, because they are so much easier both to understand and to calculate.

A conservative force is a force with the property that the work done in moving a particle between two points is independent of the path taken. Equivalently, if a particle travels in a closed loop, the net work done (i.e. the sum of the force acting along the path multiplied by the distance travelled) by a conservative force is zero. A conservative force is dependent only on the position of the object. If a force is conservative, it is possible to assign a numerical value for the potential at any point. When an object moves from one location to another, the force changes the potential energy of the object by an amount that does not depend on the path taken. If the force is not conservative, then defining a scalar potential is not possible, because taking different paths would lead to conflicting potential differences between the start and end points.

Knowing the potential function U allow us to give a complete description of the force field F. Now, we want to evaluate the work done from forces in moving the particle from starting position P to new position Q. If we consider the partial path from O to P and hence from P to Q, we get a total path from O to Q such that, following the expression (3.5), the total work done is

$$-U(x_Q) = \int_O^Q F(x) \bullet dx \tag{3.6}$$

where the vector x_Q is the coordinate of position Q. It is easy to see that such a work is equivalent to the work done along the path where the particle moves from O to P, so that we can evaluate it as

$$-U(x_Q) = -U(x_P) + W_{PQ} \tag{3.7}$$

where we have added the work W_{PQ} done along the path where the particle moves from P to Q. Finally, we can write

$$\begin{aligned} W_{PQ} &= U(x_P) - U(x_Q) \\ &= \int_P^Q F(x) \bullet dx \end{aligned} \tag{3.8}$$

so that the work done from forces acting on the particle and moving it from starting position P to ending position Q is equivalent to the difference of the potential function evaluated in any two fixed points along the path of integration. In the definition of potential function (3.5), we can arbitrary choose the initial point O for the path of integration because we have just seen that in the above expression only the difference between the value of potential function evaluated in the two points has to be considered. Hence, any constant of integration added to U doesn't affect the evaluation of the work.

On the other hand, if we know the potential function U, then we are able to evaluate the force F in every point by a simple differentiation. According to what has been said so far, in evaluating the infinitesimal work dW done by forces acting on particle in moving from position x to infinitesimal new position $x + dx$ we get as follows

$$\begin{aligned} dW &= F(x) \bullet dx \\ &= F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3 \\ &= -(\frac{\partial U(x)}{\partial x_1}dx_1 + \frac{\partial U(x)}{\partial x_2}dx_2 + \frac{\partial U(x)}{\partial x_3}dx_3) \\ -dU &= -(U(x) - U(x + dx)) \end{aligned} \tag{3.9}$$

where we have considered an arbitrary space interval dx. Consequently, we get the following fundamental relation

$$\begin{aligned} F(x) &= -\frac{\partial U(x)}{\partial x} \\ &= -\nabla U(x) \end{aligned} \tag{3.10}$$

so that the potential function U is related with the conservative force field F. In other word, such an expression has to be interpreted as the vector field $F(x)$ can be expressed in terms of the *minus gradient of potential* function for each point $x \in A$. In such a way the three component functions $F_1(x)$, $F_2(x)$ and $F_3(x)$ of the force are related and can be expressed in terms of a single function: the potential U.

The historical development of the concept of potential function may be described as: Brahe collected data on the paths of planets in the sky, then Kepler analysed that data to produce "laws", next Newton found differential equations consistent with Kepler's laws and produced further laws. Later Lagrange set down the potential function for the gravitational potential case, namely $U(x) = -G/|x_0 - x|$, with G the constant of gravitation where x denotes location in $\mathbb{R}^3$ (see [46]). This function leads to attraction of a particle at the position x towards the position x_0 (see [47] pages 277-289, [82] pages 13-17).

3.1.2 Equilibrium for a potential function

The goal of this section is to look into different kinds of *equilibrium positions* for a particle motion. We shall not go for details of every aspects of equilibrium, but will limit the analysis to the context

of equilibrium in conservative force field. For a particle moving in conservative force field we can define a potential function as a function of position. Using the concept of potential function, we shall attempt to correlate potential function with the nature of equilibrium in three categories: (1) stable, (2) unstable and (3) neutral (static or dynamic).

Since we are interested here to establish the characterizing features of equilibrium, we shall keep our discussion limited to one-dimension and attempt to find the required correlation. Consider a conservative force field F acting on the particle and moving it along a path according to linear motion where the forces are in the same x-direction of such a motion. Then, in one-dimensional case the expression of potential function (3.5) becomes

$$U(x) = -\int_{x_0}^{x} F(x)\, dx \tag{3.11}$$

where the potential function can be evaluated, for simplicity, in the start point $x_0 = 0$. If we know the potential function U, then we are able to find the force field F by a simple differentiation of the potential function

$$\begin{aligned} F(x) &= -\frac{dU(x)}{dx} \\ &= -U'(x) \end{aligned} \tag{3.12}$$

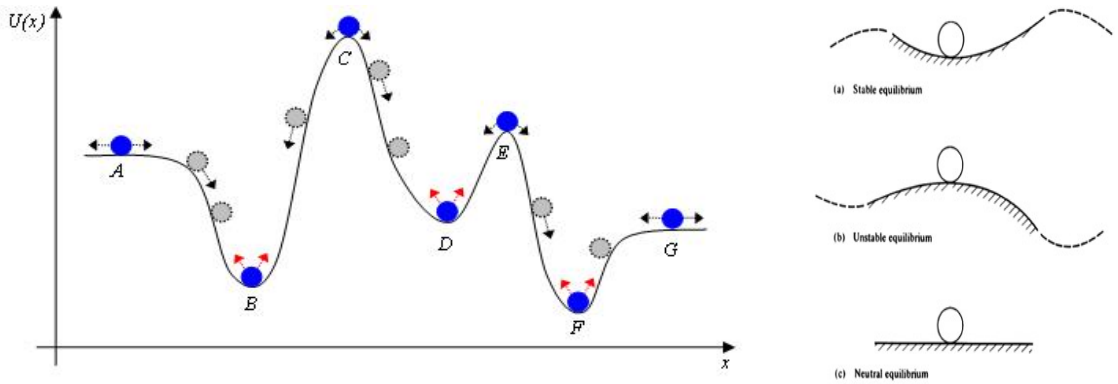
so that the force is equivalent to minus spatial derivative of potential function. It is easy to realize that the dynamics of a particle subject to a conservative force and moving along a straight line can be analysed starting from force field F which describes the total forces acting on the particle, likewise starting form the corresponding potential function U. If we know the expression of F or that of U, then it is always possible to calculate each other. In order to evaluate the different kinds of equilibrium, we have to consider all the positions where the force is equal to zero

$$F(x) = -\frac{dU(x)}{dx} = 0 \tag{3.13}$$

so that this relationship can be used to interpret equilibrium, if we have values of potential function $U(x)$ with respect to displacement x. A plot of potential function will indicate position of equilibrium, when the function's derivative is equal to zero, where tangent to the plot is parallel to x-axis so that slope of the curve is zero at that point, then the net force acting on the system is equal to zero. When an object is located at one of these positions or in one of these regions it is said to be in a *state of equilibrium*. Therefore, all the equilibrium positions are interpreted as minimum or maximum points for the function U. For the equilibrium analysis we refer to the setting illustrated in Figure 3.2.

In order to appreciate the role of potential function about equilibrium, we first need to visualize equilibrium in the light of *external disturbance*. We should keep this in mind that disturbance that we talk about is a relatively small force. In particular, in our analysis it is considered as a noise or random force. A typical set of example to illustrate the nature of equilibrium consist of three settings of a small ball: (1) inside a spherical shell, (2) over the top of a spherical shell and (3) over a horizontal surface. These three settings are shown in Figure 3.1.

In the first case, what do we expect when the ball inside the shell is slightly disturbed to its left. A component of gravity acts to decelerate the motion, then brings the ball to a stop and accelerates the ball back to its original position and beyond. Restoration by gravity continues till the ball is static at the original position, depending upon the friction. The equilibrium of the ball inside the shell is said to be *stable* equilibrium as it is unable to move out of its setting. The identifying nature of this equilibrium is that a *restoring force* comes into picture to restore the position of the object. In order



(a) (b)

Figure 3.1: Potential function $U(x)$ with (a) some possible equilibrium points and (b) the nature of different equilibrium positions.

to correlate potential function with stable equilibrium, we draw an indicative x-displacement plot of the particle motion, potential function plot and force-displacement plot corresponding to potential function curve (Figure 3.2 (a)). On the Figure 3.1 (a), equilibrium point corresponding to C is a point of stable equilibrium (bowls): if the system is slightly displaced to either side the forces on either side will return the object back to these positions. In particular, if we give a small disturbance to the ball, its motion is bounded by the energy imparted during the disturbance. We, therefore, conclude that the equilibrium of an object located in such a positions is a stable equilibrium and that this situation with respect to such a ball is similar to the spherical ball placed inside a spherical shell. More formally, at such an equilibrium point $x = x_0$, we conclude that

$$\frac{dU(x)}{dx}|_{x=x_0} = 0 \qquad \frac{d^2U(x)}{dx^2}|_{x=x_0} > 0 \qquad U(x) = U_{\min} \tag{3.14}$$

where the first derivative of potential function with respect to x-displacement is equal to zero, the second derivative is greater than zero so that the potential of the body is *minimum* for stable equilibrium for a given potential function. Consider the region $(x_0 - \triangle x, x_0 + \triangle x)$ around the minimum in $U(x)$. For $x = x_0 - \triangle x$, the corresponding force is $F = -dU/dx > 0$, which means a force tending to decrease the displacement in a negative x-direction. For $x = x_0 + \triangle x$, the corresponding force is $F = -dU/dx < 0$, which means a force tending to decrease the displacement in a positive x-direction. Taken together, these imply that equilibrium at $x = x_0$ is stable. The position of stable equilibrium is bounded by the maximum allowable energy imparted during the disturbance and force acts to restore the original position of the body in stable equilibrium. It is important to note that for $x = x_0$, $F = -dU/dx = 0$, which means that the influence of force (potential) is zero and the evolution is largely determined by random forces. In Section 3.5 we provide motivations that this setting is in agreement with economic arguments.

Let us now consider the second case in which the ball is placed over the shell. We can easily visualize that it is difficult to achieve this equilibrium in the first place. Secondly, when the ball is disturbed with a smallest touch, it starts falling down. The gravity here plays a different role altogether. It aids in destabilizing the equilibrium by pulling the ball down. The equilibrium of the ball over the top of the sphere is called *unstable* equilibrium. The identifying nature of this equilibrium is that once equilibrium ends, there is no returning back to original position as there is no restoring mechanism

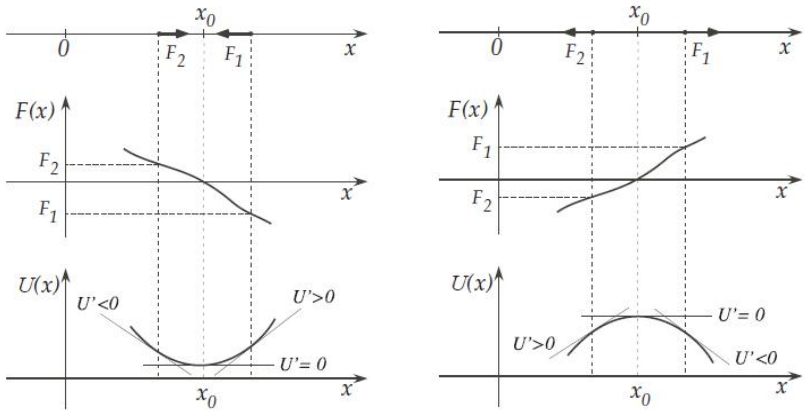


Figure 3.2: The behavior of potential function U and force F in the region $(x_0-\triangle x, x_0+\triangle x)$ around a point of (a) stable equilibrium and (b) unstable equilibrium.

available. The nature of the potential function plot for unstable equilibrium is inverse to that of stable equilibrium curve. A typical plot is shown in the Figure 3.1. The position of equilibrium at point x_0, where slope of the curve is zero corresponds to maximum potential, which in turn is equal to maximum energy allowable for the body. We can refer this plot to the case of a ball placed over a spherical shell. It is easy to realize that the ball has maximum potential at the top as is shown in potential function plot. Further, as force is negative of the slope of the potential curve, it is first negative when slope of potential curve is negative and becomes negative when slope is positive. An indicative displacement plot, potential function plot and force-displacement plot corresponding to potential function curve is shown in Figure (ref). On the Figure 3.1 (a), equilibrium points corresponding to B and E are positions of *unstable* equilibrium (a crest or peak): if the object is displaced ever so slightly from this position, the internal forces on either side will act to encourage further displacement instead of returning it back to B (or to E). As such, it is stopped after some distance due to friction or keeps moving with uniform velocity if surface is smooth. We, therefore, conclude that the equilibrium of an object located in such a positions is an unstable equilibrium and that this situation with respect to such a ball is similar to the spherical ball placed over a spherical shell. More formally, at such an equilibrium point $x = x_0$, we conclude that

$$\frac{dU(x)}{dx}|_{x=x_0} = 0 \qquad \frac{d^2U(x)}{dx^2}|_{x=x_0} < 0 \qquad U(x) = U_{\max} \tag{3.15}$$

where the first derivative of potential function with respect to x-displacement is equal to zero, the second derivative is less than zero so that the potential of the body is *maximum* for unstable equilibrium for a given potential function. Consider the region $(x_0 - \triangle x, x_0 + \triangle x)$ around the maximum in $U(x)$. For $x = x_0 - \triangle x$, the corresponding force is $F = -dU/dx < 0$, which means a force tending to increase the displacement in a negative x-direction. For $x = x_0 + \triangle x$, the corresponding force is $F = -dU/dx > 0$, which means a force tending to increase the displacement in a positive x-direction. Taken together, these imply that equilibrium at $x = x_0$ is unstable. There is no bounded allowable energy imparted during the disturbance like in the case of stable equilibrium, so that there is no restoring force acting on the body: a slightly external force is enough to leave the original and unstable position. As in the stable case, note that for $x = x_0$, $F = -dU/dx = 0$, which means that the influence

of force (potential) is zero and the evolution is largely determined by random forces.

In the last case, when ball is disturbed, it moves on the horizontal surface. If the surface is smooth it maintains the small velocity so imparted. The distinguishing aspect is that component of gravity in horizontal direction (direction of motion) is zero. As such gravity neither plays the role of restoring force nor that of an aid to the disturbance. The equilibrium of the ball on the horizontal surface is called *neutral* (*static* or *dynamic*) equilibrium. The identifying nature of this equilibrium is that once equilibrium ends, there is neither the tendency of returning back nor the tendency of moving away from the original position. The nature of the potential function plot for neutral equilibrium is easy to visualize. Let us consider equilibrium of the ball, which is lying on horizontal surface. If we consider the horizontal surface to be the zero reference potential level, then potential function plot is simply the x-axis itself, if not, it is a straight line parallel to x-axis. On the other hand, force-displacement plot is essentially x-axis as component of gravity in horizontal direction is zero. When ball is disturbed from its position, it merely moves till friction stops it. If the surface is smooth, ball keeps moving with the velocity imparted during disturbance. On the Figure 3.1 (a), equilibrium points corresponding to A and D are positions of *neutral* equilibrium (a plateau): since there is no net force acting on the object it must be at rest (static) or must be moving at a constant velocity (dynamic). More formally, at such an equilibrium point $x = x_0$, we conclude that

$$\frac{dU(x)}{dx}|_{x=x_0} = 0 \qquad \frac{d^2U(x)}{dx^2}|_{x=x_0} = 0 \tag{3.16}$$

where the first derivative of potential function with respect to x-displacement is equal to zero, the second derivative is also equal to zero so that the potential of the body neither has minimum nor maximum for a given potential function. For the position of neutral equilibrium there is no bounded allowable energy imparted during the disturbance like in the case of stable equilibrium, so that there is no restoring force acting on the body: external force neither acts to restore the body nor aid in acquiring energy by the body.

3.2 Deterministic and stochastic gradient systems

As we have seen so far, the analysis of dynamics can be transferred from a vector field F into the corresponding potential function U, by means of the gradient operator ∇. Such a construction allows us to define a *dynamic gradient system* in the following way

$$\begin{aligned} \frac{dx(t)}{dt} &= -\nabla U(x(t)) \\ dx(t) &= -\nabla U(x(t))\ dt \end{aligned} \tag{3.17}$$

where $U : A \subseteq \mathbb{R}^n \to \mathbb{R}$ is a scalar function in $C^2(A)$. It is possible to define a gradient system in $\mathbb{R}^n$ as a system of n differential equations of the form (3.17), so that such an expression is equivalent to the following system

$$\begin{cases} dx_1(t) &= -\frac{\partial U(x)}{\partial x_1} dt \\ &\vdots \\ dx_n(t) &= -\frac{\partial U(x)}{\partial x_n} dt \end{cases} \tag{3.18}$$

where the partial derivative of the function U is evaluated with respect to every n-th spatial component (Hirsch et al. 2004 [47]). The possibility to get a complete description of the properties of a vector-valued function by means of the known corresponding potential function and, therefore, to express the

ensemble in the form of a dynamic gradient system is a preferred condition. It is easy to realize that working with a gradient system has advantages in statistical modeling for the potential function U since it is real-valued, as opposed to matrix and vector-valued forms.

In order to build a more realistic description of the dynamics, we need to consider an external disturbance and we can think of a complementary force in terms of a noise component. The conversion of a model from deterministic form to stochastic form can be formalized adding a *diffusion* term to our deterministic gradient system. For a stochastic process $(X(t))_{t\in\mathbb{T}}$ we are able to get the so-called *stochastic gradient system* as follows

$$dX(t) = -\nabla U(t, X(t))\ dt + V(t, X(t))\ dB(t), \tag{3.19}$$

where $U : \mathbb{R}^n \to \mathbb{R}$ is the potential function and its minus gradient describes the drift term coefficient, whereas the stochastic process $(B(t))_{t\in\mathbb{T}}$ is a Brownian motion and the matrix V whose components are the scalar factors that measure the magnitude of random fluctuations, i.e. the influence of the Brownian Motion on the evolution of the process. It is to be noted that Brownian motion itself is a remarkable example of stochastic gradient system (3.19) corresponding to *constant* potential U and also constant matrix V. The expression (3.19) is a general form for a diffusion model since the drift term coefficient $-\nabla U$ and the diffusion term coefficient V depend not only on state variable vector $X(t)$ but also on time $t \in \mathbb{T}$. If they are functions of just state variables $X(t)$ it is called a time-homogeneous diffusion process.

In one-dimensional case, if we consider a stochastic price process $P = (P(t))_{t\in\mathbb{T}}$, the diffusion model in describing the evolution of the price value $P(t)$ is the traditional stochastic differential equation

$$dP(t) = a(t, P(t))\ dt + b(t, P(t))\ dB(t), \tag{3.20}$$

where the functions $a(t, P(t))$ and $b(t, P(t))$ are real-valued and the stochastic process $(B(t))_{t\in\mathbb{T}}$ is a one-dimensional Brownian motion. If we use a diffusion model with a potential function U, then we have to replace the drift term as follows

$$\begin{aligned} a(t, P(t)) &= -\frac{\partial U(t, p)}{\partial p} \\ &= -U'(t, P(t)) \end{aligned} \tag{3.21}$$

and the general form (3.20) becomes

$$dP(t) = -U'(t, P(t))\ dt + b(t, P(t))\ dB_t, \tag{3.22}$$

so that what distinguishes the traditional stochastic differential equation from the present approach is that the drift term in the potential model has the special form $-U'(t, P(t))$, being expressed in terms of the minus gradient of potential function. It will be seen that the modeling situation is simplified when such a function U is assumed to exist.

3.3 Learning a potential function from a trajectory

A basic issue is how to describe mathematically a potential function $U : \mathbb{R}^n \to \mathbb{R}$, supposing one exists. When we are interested in describing the dynamics of a certain phenomenon by means of a diffusion model with a potential, the main objective is how we can get such an elusive potential function, assuming such a function exists.

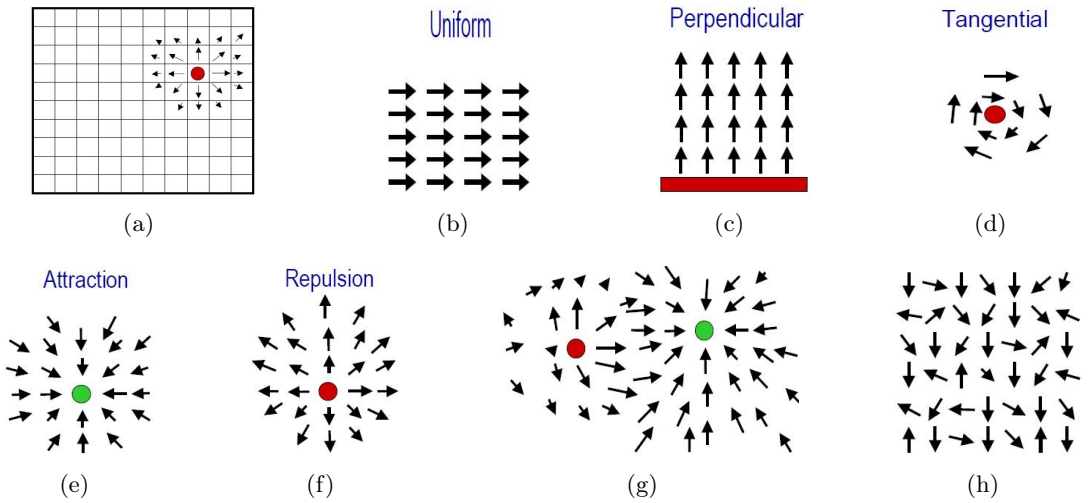


Figure 3.3: A moving particle in some primitive types of potential fields. In a hybrid process, the primitive components (g) can be combined and (h) a noise component (random force) added.

For the aims of our analysis it is important to distinguish the pure mathematical point of view from the empirical data-based application. Starting from a vector-valued function F, from a purely mathematically point of view, the general problem here is as follows. Given a vector field F, defined on an open simply connected region A, one thing we may be asked is to find a potential function for F. That is, we want to determine whether or not there exists a scalar valued function U such that $\nabla U = F$, and, if so, find such a potential function. In general, we cannot guarantee the existence of such a function. To establish the existence of a potential for F, that is to show that F is conservative, we can use standard results from calculus. If it is the case that F is conservative, then we can find the potential U trough a systematic procedure.

In the case of applications we do not know a vector field F, but we are interested to describe the forces expressed by F. In a continuing large-scale experiment, simulations or observations can be collected and we get a dataset $x = \{x_i\}_{i=1}^n$, so that we turn to the main issue: how to mathematically describe a function U, supposing such a function to be the potential function for a conservative vector field F. A basic issue becomes how to learn and mathematically describe a potential function starting from a trajectory given by the dataset $x = \{x_i\}_{i=1}^n$.

Several forms of potential may be considered, ranging from a simple monomial form to more sophisticated functional forms. For example, we might suppose U is linear in a vector-valued parameter β making study and estimation easier. The potential functions might be represented via basis functions. The basis functions are given and might be monomials, polynomial expansions, cosinusoids or Gaussian densities, for example. Another example of potential is a kernel node based, where as a specific example of kernel one has the radial basis thin plate splines. More in general, for the structure of the potential it is important to consider *attraction* and *repulsion* region, in order to mimic empirical evidences for the data generating process governing the dynamics. Consider a region $A \subset \mathbb{R}^n$ and a point x outside A. Potential functions can be set down allowing attraction or repulsion from A. Specifically, if one consider a distance function $\mathrm{d}_A(x)$ which denotes the minimum distance from the point x outside A to the region A, so that we can set the potential $U(x) = \beta(\mathrm{d}_A(x))^\alpha$. If the signum

of parameter is $\alpha > 0$, then one has attraction to A and repulsion if $\alpha < 0$. One can reverse attraction and repulsion by changing the sign of the function d_A. It is easy to realize that one might consider a hybrid process for more general purposes. Then, the above functional forms may be added together to provide other and more realistic structures of potential. Some primitive types of potential are shown in Figure 3.3. Moreover, it can be convenient to use nonlinear function. When appropriate one can set down stochastic differential equations or their approximations with non-Gaussian and autocorrelated stimulation as well.

There is theoretical and applied literature about models concerning stochastic gradient systems (see for example Brillinger et al., 2011 [16] and the mentioned references). In particular, Ao (2004 [3]) introduces the potential in stochastic differential equations by means of a novel construction based on a mathematical structure discovered during a study in gene regulatory network dynamics. Further possible analogy for applications in economics and finance are explored, using Boltzmann-Gibbs distribution (Ao 2007 [4], Kwon et al. 2005 [58], Ao et al. 2007 [5]). In a moving potential approximated by a proper quadratic function Takayasu et al. (2006 [79]) investigate potential force observed in market dynamics, whereas Watanabe et al. (2009 [88]) consider potential forces observed in a financial crisis.

3.4 Randomness as an equilibrium

In order to make the potential function approach stronger, we believe that viewing randomness in an equilibrium perspective is an interesting aspect to study further. In this section we consider randomness as viewed through an analogy between the physical quantity of ideal gas density and the mathematical construct of probability mass or density function (Grendar and Grendar, 2001 [38]). Starting from the origin of Brownian motion, the interplay between Probability, Mathematical Statistics and Statistical Physics has recognized an extremely large and important work (Jaynes and Rosenkrantz, 1983 [52]). For example, *Metropolis algorithm* and *Simulated Annealing* as well refer to such an interplay of these three areas (Metropolis et al. 1953 [64], Kirkpatrick et al. 1983 [57]).

The analogy is based on Boltzmann's deduction of equilibrium distribution of ideal gas placed in an external potential field which provides a way of viewing probability density from a perspective of forces/potentials, hidden behind it. Since the probability mass or density function is a mathematical construct, the fundamental equivalence offers a physical model of the probability distribution, so the probability density function is materialized by the density function of the ideal gas placed in an external potential field.

For our purposes, we make a very short exposition of Boltzmann's deduction of equilibrium distribution in the context of potential field. Molecules of the gas, closed in an infinitesimally small cube with sides dx,dy,dz are subjected to two forces. For simplicity, restricting to one-dimensional case, the two forces are given by

$$dF^{(1)}(x) = -\frac{dU(x)}{dx} n(x) d(x) \qquad dF^{(2)}(x) = -\frac{dn(x)}{dx} d(x) \tag{3.23}$$

where force $F^{(1)}$ is represented by a potential function $U(x, y, z)$, and force $F^{(2)}$ is induced by spatial difference of density $n(x, y, z)$ of the ideal gas. In the equilibrium, the two forces should compensate

each other, so

$$\begin{aligned} dF^1(x) + dF^2(x) &= 0 \\ -\frac{dU(x)}{dx} n(x) d(x) - \frac{dn(x)}{dx} d(x) &= 0 \\ \frac{1}{n(x)} \frac{dn(x)}{dx} &= -\frac{dU(x)}{dx} \end{aligned} \tag{3.24}$$

hence, the equilibrium condition (3.24) is a differential equation solved by

$$n(x) = ke^{-U(x)} \tag{3.25}$$

that is the famous Bolzmann's formula. The formula expresses distribution of the ideal gas placed in a potential field, *in the equilibrium*.

Using the equilibrium condition (3.24), the equilibrium view of Boltzmann's deduction can be transferred into probability context. In order to provide the analogy, a sort of principles of the physicalization of randomness can be formulated as follows: (P1) randomness has its own, *intrinsic* field; (P2) any randomness takes place also in an *external* field. Consider a random variable $X : \Omega \to \mathbb{R}$ with probability distribution or mass function $f_X : S \subseteq \mathbb{R} \to [0, 1]$ defined over a support S. Fundamental terms of vocabulary of the equilibrium view of randomness are built up in analogy with above deduction. The randomness is subjected to two intensities, defined as

$$E^s(x) = -\frac{1}{f_X(x)} \frac{df_X}{dx} \qquad E^c(x) = -\frac{dU(x)}{dx} \tag{3.26}$$

where E^s is the intrinsic intensity of its potential field, whereas E^c is the intensity of external field and U is the potential function of the external field. The intrinsic intensity E^s represents the *stochastic* component of randomness which can be also called stochastic intensity, and the external intensity E^c represents the *causal* component of randomness which can be also called causal intensity. If the two intensities of the internal and external fields compensate each other, so

$$\begin{aligned} E^s(x) + E^c(x) &= 0 \\ -\frac{1}{f_X(x)} \frac{df_X}{dx} - \frac{dU(x)}{dx} &= 0 \\ \frac{1}{f_X(x)} \frac{df_X(x)}{dx} &= -\frac{dU(x)}{dx} \end{aligned} \tag{3.27}$$

hence the equilibrium condition (3.27) is a first order linear differential equation with variable coefficients, solved by

$$f_X(x) = ke^{-U(x)} \tag{3.28}$$

where $k = \frac{1}{Z}$ is the normalizing constant and $Z(x) = \sum e^{-U(x)}$ denotes the statistical sum. In this way, the random variable X attains its equilibrium distribution in potential U.

To get rid of the normalization constant Z, we can take the normalized potential $\widetilde{U}(x)$ of the random variable X as follows

$$\begin{aligned} \ln(f_X(x)) &= \ln(\frac{1}{Z}) + \ln(e^{-U(x)}) \\ &= \ln(\frac{1}{Z}) - U(x) \\ -\ln(f_X(x)) &= U(x) - \ln(\frac{1}{Z}) \\ \widetilde{U}(x) &= U(x) + \ln(Z) \end{aligned} \tag{3.29}$$

where it can be shown that for every U there exist the corresponding normalized potential $\widetilde{U}$ for which the evolution has the same invariant distribution (3.28) with the normalized $\widetilde{Z}(x) = \sum e^{-\widetilde{U}(x)} = 1$. So that, we thus assume $Z = 1$ for the potential we consider ($\ln(1) = 0$). In this way, according to relations (3.28) and (3.29), we can write

$$\widetilde{U}(x) = -\ln(f_X(x)) \tag{3.30}$$

so there is a *unique* relationship between normalized potential and equilibrium distribution.

According to what has been developed so far, equilibrium probability distribution generated by causal intensity E^c is given by

$$\begin{aligned} f_X(x) &= e^{\int E^c(x)\ dx} \\ &= e^{-\int \frac{dU(x)}{dx}\ dx} \end{aligned} \tag{3.31}$$

where we assume $k = 1$. Now, we want to briefly discuss the simplest form of causal intensities and their corresponding equilibrium distributions generated via (normalized) potentials.

In the case of causal intensity equals to zero

$$E^c(x) = -\frac{dU(x)}{dx} = 0 \tag{3.32}$$

then, in the equilibrium, such a causal intensity generates *uniform* probability density function. Equivalently, uniform distribution reveals absence of causal component of randomness, so that uniform distribution becomes intrinsic distribution of randomness, left alone. Assuming N elements, the potential which generates this effect is $\widetilde{U}(x) = \ln(N)$, that is the corresponding (normalized) potential of the uniform discrete distribution.

In a different situation, when causal intensity equals a real constant

$$E^c(x) = -\frac{dU(x)}{dx} = -a \tag{3.33}$$

then, in the equilibrium, such a constant causal intensity generates the *exponential* probability density function, e^a. Equivalently, exponential distribution reveals presence of *constant* causal intensity in a studied effect. The potential which generates this effect is

$$\widetilde{U}(x) = ax - \ln(a) \tag{3.34}$$

that is the corresponding (normalized) potential of the exponential distribution. Another context arises when causal intensity has a linear form

$$E^c(x) = -\frac{dU(x)}{dx} = -bx \tag{3.35}$$

then, in the equilibrium, a linear causal intensity generates a *Normal distribution* whit mean $\mu = 0$ and variance $\sigma^2 = \frac{1}{b}$. Equivalently, Normal distribution reveals presence of linear causal intensity in a studied effect. The potential which generates this effect is

$$\widetilde{U}(x) = \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} + \ln(\sqrt{2\pi\sigma^2}) \tag{3.36}$$

that is the corresponding (normalized) potential of the normal distribution. In other words, it is a linear harmonic oscillator. When we consider superposition of constant and linear intensities

$$E^c(x) = -\frac{dU(x)}{dx} = -a - bx \tag{3.37}$$

then such a causal intensity generates a probability density function $f_X(x) = ke^{-ax-b\frac{x^2}{2}}$, where k is a normalizing constant.

Causal intensity of the form

$$E^c(x) = -\frac{\Gamma'(x+1)}{\Gamma(x+1)} + \ln(\lambda) \tag{3.38}$$

generates Poisson distribution, $Poi(\lambda)$. Equivalently, Poisson distribution reveals presence of the above force. The potential which generates this effect is

$$\widetilde{U}(x) = \lambda - x\ln(\lambda) + \ln(\Gamma(x+1)) \tag{3.39}$$

that is the corresponding (normalized) potential of the Poisson distribution.

When causal intensity has the form

$$E^c(x) = -\frac{a}{x} - b \tag{3.40}$$

then, it generates Gamma distribution, $\Gamma(1-a, \frac{1}{b})$. Equivalently, Gamma distribution reveals presence of a superposition of constant and reciprocal causal intensities. The potential which generates this effect is

$$\widetilde{U}(x) = (1-\alpha)\ln(x) + \frac{x}{\beta} + \ln(\Gamma(\alpha)\beta^\alpha) \tag{3.41}$$

that is the corresponding (normalized) potential of the Gamma distribution, $\Gamma(\alpha, \beta)$.

Finally, we want to consider the family of Pearson distributions. The following causal intensity

$$E^c(x) = \frac{x-a}{b_0 + b_1 x + b_2 x^2} \tag{3.42}$$

is defined in the context of Person's system of distributions, an important tool used for densities approximation ([78]). It can be shown that such a causal intensity generates the family of Pearson distributions.

We believe that such a way of viewing probability density from a perspective of forces/potentials, hidden behind it, is an interesting aspect to study further. For instance, in the next Section 3.5 we will clearly seen interpretation of any probability distribution through its relationship to potential in the case of Langevin equation and the well known Ornstein-Uhlenbeck process as its solution. Indeed, normal distribution is from the equilibrium perspective a distribution of linear harmonic oscillator. Among the possible gains of such a weird and elusive effort one can find, first of all, a clear splitting of any random effect into its stochastic part, that is its own intrinsic component, and deterministic parts, that is the potential component. Furthermore, such a perspective is a valid support in terms of search for causes/forces in study of a random effect. Revealing forces behind a random effect thus can help to understand the effect, and consequently help to predict, and also govern or regulate the future outcomes of the effect.

3.5 The potential function model

In this section our main interest is to present the approach of potential function to model continuous as well as discrete time processes that exhibit the phenomenon of state variable clustering, that is the

multiple attraction regions (Borovkova et al., 2003 [13]). What distinguishes the above models from the present approach is that the purpose of the potential function is to model the distribution of the process underlying the observed time series. A realization of a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is given by an observed time series $x = \{x_i\}_{i=1}^n$. Such an observed time series gives the underlying invariant measure, that is the relative frequency with which the process visits different regions of its allowed values. The phenomenon of state variable clustering can be clearly seen on the histogram showing that the observations prefer to concentrate around some regions (see Figure 2.1). To model such a phenomenon, we need to mimic this multimodal invariant distribution of the underlying process. In the framework of the potential field, the present approach suggest using the potential function as the main factor governing the evolution of the process.

In order to introduce the formal model, let's consider a continuous time process $P = (P(t))_{t\in\mathbb{T}}$ in one-dimensional case, $P(t) \in \mathbb{R}$, where the time index set is $\mathbb{T} = [t_0, T]$. In general, a stochastic differential equation is a typical problem of the following type

$$\begin{cases} dP(t) = a(t, P(t))\ dt + b(t, P(t))\ dB(t) \\ P(t_0) = \xi_{t_0}(\omega) \end{cases} \tag{3.43}$$

where $a, b : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$ are measurable scalar valued functions of deterministic type depending on the time, $t \in \mathbb{T}$, and state variable, $P(t) \in \mathbb{R}$. The function a is referred to as the *drift* coefficient, while the function b is called the *diffusion* coefficient. The stochastic process $(B(t))_{t\in\mathbb{T}}$ denotes a Wiener process or standard Brownian motion on a probability space $(\Omega, P, \mathcal{F})$ with respect to the filtration $\mathbb{F} = (\mathcal{F}(t))_{t\in\mathbb{T}}$. The initial condition may be a random variable $\xi(\omega, t_0)$ or also a scalar $\xi(\omega, t_0) = p_{t_0}$, that is a degenerate random variable which only takes a single value p_{t_0}.

For the purpose of doing a model with a potential function we consider the following choices

$$a(t, P(t)) = -U'(P_t) \qquad \forall t \in \mathbb{T} \tag{3.44}$$

$$b(t, P(t)) = \kappa \qquad \forall t \in \mathbb{T} \tag{3.45}$$

where, therefore, the drift coefficient a becomes the minus first derivative of a function U with respect to the state variable P_t, whereas the diffusion coefficient b is assumed to be a real constant, $\kappa \in \mathbb{R}$. In such a way we postulate that the evolution of the state variable $P(t)$ is described according to the following stochastic differential equation

$$dP_t = -U'(P(t))\ dt + \kappa\ dB(t) \tag{3.46}$$

where $U : A \subseteq \mathbb{R} \to \mathbb{R}$ is a potential function, the stochastic process $(B(t))_{t\in\mathbb{T}}$ denotes the standard Brownian motion, and $\kappa \in \mathbb{R}$ is the scalar factor that measures the magnitude of random fluctuations, that is the influence of the Brownian motion on the evolution of the process.

According to what we have done so far, now take a first look at the underlying assumptions of such a potential model. The first assumption is that, whereas the potential function U is dependent from the current value of the state variable, but the shape of the potential do not change with time. The second assumption is that the scalar factor κ is constant, so that also such a parameter do not change with time. In particular, we interpret κ as the average volatility or long-term volatility. In Section 5.7 are given more details about the share-out of the magnitude of the price increments that in some part are determined by the derivative of the potential and the remaining part by random fluctuations. In particular, such a deterministic component (via potential's derivative) plays a significant role in order to explain part of the price variability.

The equation model (3.46) is an example of a diffusion process with drift. Moreover, starting from general equation (3.43), if we consider the following choices

$$a(t, P(t)) = -\alpha P(t), \qquad b(t, P(t)) = \kappa \qquad \forall t \in \mathbb{T}$$

we get the traditional problem

$$\begin{cases} dP(t) = -\alpha P(t)\ dt + \kappa\ dW(t) \\ P(t_0) = p_0 \end{cases} \tag{3.47}$$

that is a *Langevin equation* for which the stochastic process $P = (P_t)_{t\in\mathbb{T}}$ as a solution of the equation (3.47) is the well known *Ornstein-Uhlenbeck process*. It is easy to realize that if the potential is a quadratic function, we get

$$U(p) = ap^2 + c \tag{3.48}$$

$$\frac{dU(p)}{dp} = 2ap = \alpha p \tag{3.49}$$

where $\alpha = 2a$. As a consequence, the Ornstein-Uhlenbeck process is a particular case of a model with potential. In Section 2.3 we have seen that the Ornstein-Uhlenbeck process stands for modeling mean reversion property and that mean-reverting class of diffusion models have been widely used to model commodity prices. However, in the case of mean-reversion class of models the Gaussian distribution is the underlying invariant measure and has only one attraction point, so the models postulate the existence of just one price equilibrium level, and hence cannot generate processes with multiple attraction regions.

3.5.1 How the potential model works

In Section 3.1.1 we have seen that the minus sign in potential model (3.46) comes from classical mechanics and is traditional. In Section 3.1.2 we have taken a look into different kinds of equilibrium positions of a particle moving in a potential field. One notes then that depressions in U correspond to points of attraction, while ridges lead to repulsion. At this stage, it is important to provide some insight, so that we want to illustrate the approach of a potential model by a simple heuristic model. In a discrete time setting the time variable becomes $t \in \mathbb{N}$. If we denote with $(P_t)_{t\in\mathbb{N}}$ the series of prices, then the evolution of price (or price increments) movement can be modeled as

$$P_{t+1} - P_t = -U(P_t) + \epsilon_{t+1} \tag{3.50}$$

$$P_{t+1} = P_t - U(P_t) + \epsilon_{t+1} \qquad t = 0, 1, 2, \ldots \tag{3.51}$$

where $U : \mathbb{R} \to \mathbb{R}$ denotes a potential function and the process $(\epsilon_t)_{t\in\mathbb{N}}$ denotes some random perturbations. As the evolution illustrated in Figure 3.4, in one dimension the potential function can be thought of as like a "multi-bowl" with dips at local minima, and the observed process is the horizontal coordinate of a ball moving along the bowl walls, subject to gravity and some random impulses. Because of the gradient of the potential is a direction to move from current location to the maximum value, as a consequence of minus sign the next value P_{t+1} of the series tend to move in the direction of the nearest minimum of the potential function with respect to the current value P_t. The random fluctuations ϵ_{t+1} make sure that the series is not trapped in a local minimum and can continue to move from one local minimum to another. The ball has a natural tendency to move downhill into the dips, but random shocks prevent it from settling there and force it up the walls and move it from one dip to

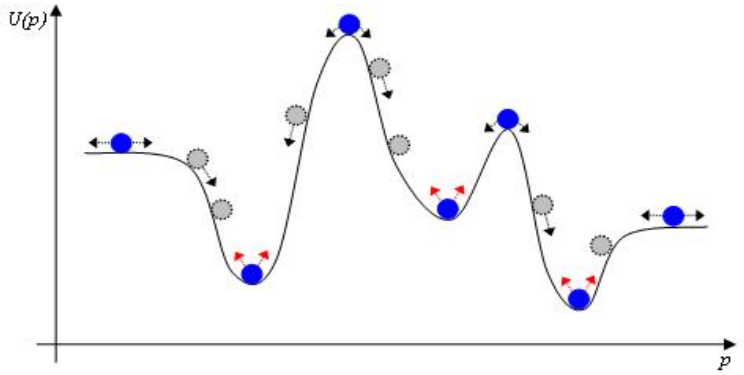


Figure 3.4: The idea behind the diffusion model where the evolution of the process $p = (P(t))_{t\in\mathbb{T}}$ is governed by a potential function U.

another. The role of the random fluctuations $(\epsilon_t)_{t\in\mathbb{N}}$ can be thought of as placing the ball randomly in other parts of the potential, not necessarily in the direction of the downhill gradient. Note that the deeper a local minimum of the potential, the longer (more time) the process spends there. So that in a series of observations on such a process, there are more observations in the neighborhood of the minima of the potential than at other locations.

It is important to note that the main forecasting power of the model lies in its improved ability to predict the direction of the next move, once evolution departs from a local minimum. There the influence of the deterministic potential field prevails over random fluctuations, while at a local minimum the derivative of the potential field is close to zero and the evolution is largely determined by random forces. This is in agreement with economic arguments: if the price is far from an equilibrium price, external forces of the market drive the price towards nearest equilibrium, while at equilibrium price fluctuations are largely due to random shocks.

It is interesting to note that the present potential model has been inspired by the applications of diffusions for global optimization (see Geman and Hwang, 1986 [34]), which in turn are related to an heuristic technique called *simulated annealing*. In the problem of optimization, we can find a local minimum of a function U on $\mathbb{R}^n$ by starting at an arbitrary point $x_0 \in \mathbb{R}^n$ and solving the following deterministic gradient system

$$\begin{aligned} \frac{dx(t)}{dt} &= -\nabla U(x(t)) \\ dx(t) &= -\nabla U(x(t))\ dt \end{aligned} \tag{3.52}$$

as we have seen in Section 3.2. When looking for a global minimum, the path $x(t)$ should not only follow downhill gradients but be forced to a sort of "climb hills" in order to escape from a local minimum. This can be done by introducing the diffusion term into the path $x(t)$, so that the gradient system becomes stochastic, and we write

$$dX(t) = -\nabla U(X(t))\ dt + \sqrt{2T}\ dB(t) \tag{3.53}$$

where $(B(t))_{t\in\mathbb{T}}$ denotes the standard Brownian motion, whereas T is the *temperature* which controls the magnitude of the random fluctuations. As the temperature tends to decrease, $T \to 0$, then the

invariant measure concentrates on the global minima of function U. Simulated annealing is the discrete version of the above minimization problem, and the name "annealing" comes from the physical procedure where a physical substance is melted and then slowly cooled to find a low energy configuration. The simulated annealing is a probabilistic method proposed in Kirkpatrick et al. (1983 [57]) and Cherny (1985 [23]). This method is now applied to a wide variety of problems (Bertsimas and Tsitsiklis, 1993 [9]). The simulated annealing is also applied to reliability data analysis and linked to mixture distribution analysis for applications in reliability testing and assessment (Tan, 2008 [80]).

In the present application of such a diffusion potential to time series modeling, the factor κ (which corresponds to the temperature T in the optimization problem) does not tend to 0 but is intrinsically determined by the process underlying the observed time series. As we have seen just above, in the potential model the factor κ measures the influence of the random fluctuations on the process relative to the influence of the potential field.

3.5.2 A distribution model: Boltzmann-Gibbs measure

Starting from equation (3.46) we only know the structure of the model, so we need to perform a procedure to get an estimate of the potential function U and the volatility parameter κ, so as to model the price evolution by letting the price process be governed according to such a diffusion. In fact, given a dataset of an observed price series, $\{p(t_i)\}_{i=1}^N$, neither the potential function nor the volatility parameter are known. However, it is possible to adopt the following distributional result in order to identify such unknown elements (Matkovsky and Schuss, 1981 [60]). Starting from the historical data $\{p(t_i)\}_{i=1}^N$, our approach is to perform a procedure to learn a potential U from such a trajectory, and subsequently to get an estimate for the parameter κ.

Assuming that the dynamics of the process $(P_t)_{t\in\mathbb{T}}$ follows an evolution described by equation (3.46), then it is possible to show that, under suitable conditions on function U, the distribution of state variable P_t approaches an equilibrium in a weak sense, which is a probabilistic structure called *Boltzmann-Gibbs distribution* with the following density function

$$f(p;\kappa,U) = \frac{1}{Z} e^{-\frac{2\ U(p)}{\kappa^2}} \tag{3.54}$$

where

$$Z(p;\kappa,U) = \int_{\mathbb{R}} e^{-\frac{2\ U(p)}{\kappa^2}}\ dp < \infty \tag{3.55}$$

is the normalizing constant. The conditions for the function U are that should be twice continuously differentiable, $U \in C^2(A)$, and that U should tend to infinity with $\|p\|$ and the integral (3.55) should be finite. These conditions assure that the partition function Z is finite and that the expression (3.54) is indeed a density function. In the present case of potential model, since the approach is to learn the potential U from the trajectory given by the historical data, such a construction of function U is able to hold the above condition automatically.

Concerning the normalizing constant, without loss of generality, it can be taken as $Z = 1$. The density (3.54) can be rewritten as

$$f(p;\kappa,U) = e^{-\frac{2(U(p)+\frac{1}{2}\kappa^2\log(Z))}{\kappa^2}} \tag{3.56}$$

and it can be shown that for every U there is a "new" potential $\widetilde{U}$ that has the same invariant distribution for which

$$\widetilde{Z}(p;\kappa,U) = \int_{\mathbb{R}} e^{-\frac{2\ \widetilde{U}(p)}{\kappa^2}}\ dp = 1 \tag{3.57}$$

is the corresponding normalization constant. Since for the potential $\widetilde{U}$ and U the evolution has the same invariant distribution, then we can take the new potential as

$$\begin{aligned}\widetilde{U} &= U(p) + \frac{1}{2}\kappa^2 \log(Z) \\ &= U(p) + \underbrace{\frac{1}{2}\kappa^2 \log(1)}_{=0}\end{aligned} \tag{3.58}$$

so that we thus assume $Z = 1$ for the potential we consider.

It was Metropolis et al. (1953 [64]) who first introduced the Boltzmann-Gibbs distribution based idea into numerical analysis, in their fundamental paper. The idea arises through constructing a means for simulation of a system at some fixed temperature. In particular, if a system is in some current energy state, E^{cur}, and some system aspects are changed to make the system potentially achieve a new energy state, E^{new}, the the Metropolis simulation always has the system go to the new state if $E^{new} < E^{cur}$. On the other hand, if $E^{new} \geq E^{cur}$, then the probability of the system going to the new state is

$$e^{\frac{E^{new} - E^{cur}}{k_B T}} \tag{3.59}$$

which is known as the *Metropolis criterion*. After a large number of such decisions and outcomes, the system eventually reaches an equilibrium where the system state is governed by the Boltzmann-Gibbs distribution (3.54). This is predicated on the system being at the fixed temperature T (Spall, 2003 [77] p.210).

3.5.3 Parameters estimation via mixture models and EM algorithm

The present approach of potential model has a fundamental step on fitting the multimodal density of the invariant distribution. Starting from the observed price series $\{p(t_i)\}_{i=1}^N$, we need to estimate the density function by means of fitting the resulting histogram of the historical data. In particular, the estimation method has to provide an analytical expression for the density, since the derivative of the potential needs to be quickly and accurately evaluated. In other words, we will see that for the use of the present model is needed the derivative of the potential function, so the estimation method has to have the advantage that such a derivative does not have to be estimated separately, but can be computed directly from the expression provided by the method.

The multimodal density can be estimated in numerous ways. One could use nonparametric methods such as to fit a kernel density estimation, but this method can be slow and, above all, do not provide the analytical expression for the derivative of the potential.

Other method is to fit a polynomial of high degree. In Borovkova et al. (2003 [13]) the authors found that fitting a polynomial is the fastest and most accurate way to apply the model, especially in univariate case, because such a method allows for fast calculation of the potential's derivative. It is important to note that they investigate further the influence of the polynomial degree, because of this element has to be taken into account concerning the robustness of the estimated model.

Parametric method can be used to estimate the multimodal density. In the present analysis, we postulate a parametric form of the invariant price distribution in the framework of finite mixture models and fit the potential by means of the maximum likelihood method with a numerical implementation of Expectation-Maximization algorithm for a finite mixture of Gaussians. This way of fitting the potential model extends the original approach, and this is particularly useful in the multivariate extension of the model. Mixture models have experienced increased interest and popularity over last decades. In

Chapter 4 we address on finite mixture models and Expectation-Maximization algorithm we are using in the present analysis.

Chapter 4

Finite Mixture Models and EM Algorithm

In the present chapter our main concern is to consider the framework of finite mixture models in the context of cluster analysis, and the Expectation-Maximization (EM) algorithm and its applications to parameter estimation for mixture models from the perspective of latent variables. Mixture models have experienced increased interest and popularity over last decades. The importance of mixture distributions, their enormous developments and their frequent applications over recent years is due to the fact that mixture models offer natural models for *unobserved population heterogeneity*. Assuming that a parametric density $f(x;\theta)$ is capable to describe the phenomenon of interest, where $\theta \in \Theta$ denotes the parameter of the population, whereas x is in the sample space $\mathcal{X} \subset \mathbb{R}$. We call this the *homogeneous case*. However, often this model is too strict to capture the variation of the parameter over a diversity of subpopulations. In this case, we have that population consists of various subpopulations. We call this situation the *heterogeneous case*. In contrast to the homogeneous case, we can consider the same type of density in each subpopulation, but a potentially different parameter. If we consider a sample dataset, here it is not observed which subpopulation the observations are coming from. Therefore, we speak of *unobserved heterogeneity*. We develop such a heterogeneity in the framework of cluster analysis numerical methods. In most applications of cluster analysis a partition of data is sought, in which each individual or object belongs to a single cluster, and the complete set of clusters contains all individuals. The purpose of cluster analysis is to determine the inner structure of clustered data when no information other than the observed values is available. Most clustering done in practice is based largely on heuristic or distance-based procedures, such as hierarchical agglomerative clustering or iterative relocation procedures. Clustering methods based on probability models offer a principal alternative to heuristic-based algorithms. In this context the data are viewed as coming from a mixture of probability distributions, each representing a different cluster. Interest in clustering has increased due to the emergence of new domains of application, such as astronomy, biology, physics and social sciences. In addition to clustering purposes, finite mixtures of distributions have been applied to a wide variety of statistical problems such as discriminant analysis, image analysis and survival analysis. To this extent finite mixture models have continued to receive increasing attention from both theoretical and practical points of view. In order to estimate the parameters of a mixture model we implement the numerical technique Expectation-Maximization (EM), which is one of the most frequently used algorithms for finding maximum likelihood estimators in mixture models. In order to take into account the unobserved heterogeneity, the implemented algorithm refers to the latent variables perspective of mixture distributions in which the discrete latent variables can be interpreted as defining assignments of data points to specific components of the mixture.

4.1 Finite mixture models

Consider a D-dimensional random variable, $X : \Omega \to \mathbb{R}^D$, taking values in a sample space $\mathcal{X} \subset \mathbb{R}^D$. The probability distribution of X is characterized by its probability density function, $f(x)$ on $\mathbb{R}^D$, defined with respect to an appropriate measure on $\mathbb{R}^D$ which is either the Lebesgue measure, or a counting measure, or a combination of the two, depending on the context. The probability density function describes the occurrence of the observed realization, $X(\omega) = x$.

Additionally, consider a population Ω made up of $k = 1, \ldots, K$ subgroups, mixed at random in proportion to the relative group sizes $\alpha_1, \ldots, \alpha_K$. Assume that interest lies in some random feature X which is heterogeneous across and homogeneous within the subgroups. Due to heterogeneity, we can think that X has a different probability distribution in each group, so that we suppose that the random feature arises from a finite mixture model, and the probability density function of this distribution, $f(x)$, takes the form of a mixture density, written in terms of its K conditional densities $f_{X|\Omega_k}$, as follows

$$\begin{aligned} f_X(x) &= \alpha_1 \ f_{X|\Omega_1}(x) + \ldots + \alpha_K \ f_{X|\Omega_K}(x) \\ &= \sum_{k=1}^{K} \alpha_k \ f_{X|\Omega_k}(x) \qquad \forall x \in \mathcal{X} \end{aligned} \tag{4.1}$$

where each $f_{X|\Omega_k}$ is a conditional probability density function for all $k = 1, \ldots, K$. A single density $f_{X|\Omega_k}$ is referred to as the *component* density or *marginal* density of the mixture. The number of components is denoted by K. The parameters $\alpha_1, \ldots, \alpha_K$ are the weights of the population's subgroups, and are called *mixing weights*. The vector $\alpha = (\alpha_1, \ldots, \alpha_K)$ is called the weight distribution, and is defined by the following constraints

$$0 \leq \alpha_k \leq 1 \tag{4.2}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \tag{4.3}$$

so that the mixture distribution is a convex linear combination of the K conditional densities $f_{X|\Omega_k}$. Moreover, it is clear the analogy of the mixing coefficient α_k with the probability that an element of population Ω belongs to the subgroups Ω_k.

4.1.1 Mixture models in the parametric context

In many applications the component densities are assumed to belong to some parametric family, and we are interested in the formulation of mixture models in the parametric context. There is no requirement that the component densities should all belong to the same parametric family, but as in most applications this is the case, we will restrict our attention to the case where all component densities have a common functional form. In this case, one assumes that all component densities arise from the same parametric distribution family $\mathbb{F} = \{f_X(x;\theta), \theta \in \Theta\}$, where the density is specified as $f_X(x;\theta)$ and indexed by $\theta \in \Theta$, which is the unknown vector of parameters of the postulated form for the k-th component of the mixture. We can then write $f_{X|\Omega_k}(x) = f_{X_k}(x;\theta_k)$, where θ_k denotes the parameters occurring in the k-th component density f_{X_k}. The finite mixture density function (4.1) will then have

the form

$$\begin{aligned} f_X(x;\theta) &= \alpha_1 \ f_{X_1}(x;\beta_1) + \ldots + \alpha_K \ f_{X_K}(x;\beta_K) \\ &= \sum_{k=1}^{K} \alpha_k \ f_{X_k}(x;\beta_k) \qquad \forall x \in \mathcal{X} \end{aligned} \tag{4.4}$$

where $\theta = (\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_K)$ denotes the vector containing the complete collection of all the unknown (and distinct) parameters occurring in the mixture model. The mixture density function f_X is indexed by the parameter $\theta = (\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_K)$ taking values in the parameter space $\Theta_{(K)} = \Theta^K \times \Psi_{(K)}$. The vector of mixing distribution $\alpha = (\alpha_1, \ldots, \alpha_K)$ describes the structure of the mixture, whereas the parameters $\beta = (\beta_1, \ldots, \beta_K)$ describe the structure of each k-th component density, f_{X_k}. In other words, the whole functional form of the mixture is governed by the unknown vector θ, which has to be estimated. We assume that the finite mixture distribution is unconstrained in the sense that no constraints are imposed on the component parameters $\beta = (\beta_1, \ldots, \beta_K)$ and that the weight distribution $\alpha = (\alpha_1, \ldots, \alpha_K)$ is unconstrained apart from the natural constraints (4.2) and (4.3).

It is important to show that the mixing coefficients $\alpha = (\alpha_1, \ldots, \alpha_K)$ constitute a valid probability distribution. First of all, we know that for each k-th marginal component f_{X_k} is

$$\int_{\mathbb{R}^D} f_{X_k}(x;\beta_k) \ dx = 1$$

so that its integral with respect to the variable x is normalized (i.e. it sums up to 1), as it should be for any density function. The same normalizing condition is valid for the mixture density f_X, so that we can write

$$\begin{aligned} 1 &= \int_{\mathbb{R}^D} f_X(x;\theta) \ dx \\ &= \int_{\mathbb{R}^D} \sum_{k=1}^{K} \alpha_k \ f_{X_k}(x;\beta_k) \ dx = \sum_{k=1}^{K} \alpha_k \int_{\mathbb{R}^D} f_{X_k}(x;\beta_k) \ dx \\ &= \sum_{k=1}^{K} \alpha_k \end{aligned} \tag{4.5}$$

where we have used the above condition on the component integrals. Now, the requirement that for each component density is $f_{X_k} \geq 0$, together with the mixture density $f_X \geq 0$, this implies that also the mixing coefficients are needed to take only non-negative values, $\alpha_k \geq 0$. Combining this with the condition (), we obtain

$$0 \leq \alpha_k \leq 1 \tag{4.6}$$

so that the above constraints (4.2) and (4.3) arise as a natural conditions. This is the motivation for that the mixture coefficients are also called *mixing probabilities*, because they satisfy the requirements in order to be valid probabilities.

4.1.2 Mixture of Gaussians

For our purposes, we are mainly concerned with finite mixture model from well known parametric distribution families. The most widely used finite mixture distributions are those involving Gaussian

components. Using multivariate normal distribution, $N(\mu, V)$, with density function $f_N(x; \mu, V)$, the mixture density (4.4) is given by

$$\begin{aligned} f_X(x;\theta) &= \alpha_1 \; N_1(x;\mu_1,V_1) + \ldots + \alpha_K \; N_K(x;\mu_K,V_K) \\ &= \sum_{k=1}^{K} \alpha_k \; N_k(x;\mu_k,V_k) \end{aligned} \tag{4.7}$$

where each k-th marginal component X_k has a multivariate Gaussian distribution, $X_k \sim N_k(x; \mu_k, V_k)$, with density

$$f_{X_k}(x;\mu_k,V_k) = \frac{1}{\sqrt{\det(V_k)}\sqrt{(2\pi)^D}} e^{-\frac{1}{2}((x-\mu_k)'V_k^{-1}(x-\mu_k))} \qquad k = 1,\ldots,K \tag{4.8}$$

each k-th component being governed by its vector of mean μ_k and its covariance matrix V_k. The reason for the importance and widespread use of Gaussian mixture models include the fact that the normal density is symmetric, unimodal and has a simple and concise representation requiring only two parameters: the mean $\mu = (\mu_1, \ldots, \mu_K)$ and the covariance $V = (V_1, \ldots, V_K)$. These distributional characteristics along with its well studied status give Gaussian mixtures the power and effectiveness that other mixture densities can hardly surpass. By using a sufficient number of Gaussians, and by adjusting their means and covariances as well as the coefficients in the linear combination, almost any continuous density can be approximated to arbitrary accuracy.

In the univariate case, for $D = 1$, the Gaussian mixture model (4.7) becomes

$$\begin{aligned} f_X(x;\theta) &= \alpha_1 \; N_1(x;\mu_1,\sigma_1^2) + \ldots + \alpha_K \; N_K(x;\mu_K,\sigma_K^2) \\ &= \sum_{k=1}^{K} \alpha_k \; N_k(x;\mu_k,\sigma_k^2) \end{aligned} \tag{4.9}$$

where each k-th mixture component X_k has a univariate Gaussian distribution, $X_k \sim N_k(x; \mu_k, \sigma_k^2)$, with density

$$f_{X_k}(x;\mu_k,\sigma_k^2) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_k}{\sigma_k})^2} \qquad k = 1,\ldots,K \tag{4.10}$$

each k-th component being governed by its mean μ_k and its variance σ_k^2.

The univariate normal mixture distributions have a long history, and the problem of estimating their parameters is one of the oldest estimation problem in the statistical literature. Indeed, the first major analysis involving the use of finite mixture models was undertaken by Pearson (1894 [67]) who fitted a mixture of two univariate normal probability density functions, with different means and different variances, to some data provided by Weldon (1892 [90], 1893 [91]).

4.2 Estimation of mixture parameters

The estimation of the parameters of a mixture distribution can be handled by a variety of techniques. They include graphical methods, method of moments, minimum-distance methods, maximum likelihood, and Bayesian methods. An exhaustive review of those methods is reported in Titterington (1985 [85]). The main reason for the huge literature on estimation methodology for mixtures is the fact that explicit formulas for parameter estimates are typically not available. In this section we are

interested in the maximum likelihood method that has focused many attentions, mainly due to the existence of an associated statistical theory. Since the advent of the Expectation-Maximization (EM) algorithm, maximum likelihood has been by far the most commonly used approach to the fitting of mixture distributions. Before proceeding to consider the EM algorithm, we will first briefly define maximum likelihood estimation in general and introduce some associated notation.

4.2.1 Maximum likelihood estimation

For estimating the unknown parameters, we apply the standard maximum likelihood estimation, which possesses desirable properties such as, under very general conditions, the estimates obtained by the method are consistent, i.e. they converge with probability 1 to the true parameter values. We first establish a general likelihood function and give the likelihood equation.

The goal of maximum likelihood estimation is to find parameters that maximize the probability of having received certain observations (or measurements) of a random variable distributed by some probability density function. Consider a D-dimensional random variable X, which describes the random feature we are interested in, with its probability density function $f_X(x;\theta)$ that is governed by the parameter $\theta \in \Theta$. Given a measurement vector $\mathbf{x} = \{x_1, \ldots, x_N\} = \{x_i\}_{i=1}^N$, that is a sample dataset of size N, supposedly drawn from this distribution, such that we assume that these data vectors are independent and identically distributed according to distribution f_X. The probability of receiving some measurement, x_i, is given by the density function $f_X(x_i;\theta)$. Therefore, the resulting density for the samples is

$$f_X(\mathbf{x};\theta) = \prod_{i=1}^{N} f_X(x_i;\theta) \tag{4.11}$$

that is the probability of having received the whole series of measurements. The likelihood function is defined as follows

$$\mathcal{L}(\theta;\mathbf{x}) = f_X(\mathbf{x};\theta) \tag{4.12}$$

so that it is thought of as a function of the parameters θ, where the data $\mathbf{x}$ are considered fixed. We also use the log-likelihood function

$$\begin{aligned} \ell(\theta;\mathbf{x}) &= \log \mathcal{L}(\theta;\mathbf{x}) = \log f_X(\mathbf{x};\theta) \\ &= \log \prod_{i=1}^{N} f_X(x_i;\theta) = \sum_{i=1}^{N} \log f_X(x_i;\theta) \end{aligned} \tag{4.13}$$

because it is analytically easier to maximize. In the maximum likelihood problem, we attempt to find the particular vector θ that maximizes the log-likelihood function such that

$$\theta^\star = \underset{\theta}{\operatorname{argmax}}\ \ell(\theta;\mathbf{x}) \tag{4.14}$$

and we wish to find the solution $\theta^\star$. This maximization can be dealt with the traditional way by differentiating the log-likelihood function with respect to the components of θ, and by just setting the derivative to zero

$$\frac{\partial \ell(\theta;\mathbf{x})}{\partial \theta} = 0 \tag{4.15}$$

in order to get the likelihood equation (4.15). Depending on the functional form of density function, $f_X(x;\theta)$, this problem can be easy or hard. Often the log-likelihood function cannot be maximized analytically, and the corresponding log-likelihood equation has no explicit solutions. In the following sections we address the problem for the important case of mixture distributions.

4.2.2 Maximum likelihood for mixture models

According to the maximum likelihood estimation method, if the probability density function $f_X(x;\theta)$ is simply a single Gaussian distribution where the vector parameter is $\theta = (\mu, \sigma^2)$, then we can set the derivative of the log-likelihood function $\ell(\theta;\mathbf{x})$ to zero, and solve directly for μ and σ^2, and this is the result in the standard formulas for the mean and variance of a given dataset. However, for many problems such a closed form is not always as easy to be derived, because it is not possible to find an analytical expression. In other cases, solutions can show the problem of circularity (i.e. the chicken-and-egg dilemma) and we must resort to more elaborate techniques. This is the case of the mixture distributions.

In the framework of finite mixture models, we assume that the occurrences of the random feature X are described according to the following probabilistic model

$$f_X(\mathbf{x};\theta) = \sum_{k=1}^{K} \alpha_k \; f_{X_k}(\mathbf{x};\beta_k) \tag{4.16}$$

where the vector parameter is composed of $\theta = (\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_K)$. Given a sample dataset of N independent and identically observations, $\mathbf{x} = \{x_i\}_{i=1}^N$, supposedly drawn from this mixture distribution, we assume that each measurement x_i is generated according to the mixture density (4.16), as follows

$$f_X(x_i;\theta) = \sum_{k=1}^{K} \alpha_k \; f_{X_k}(x_i|C=k;\beta_k) \tag{4.17}$$

where the conditional $x_i|C=k$ denotes that the i-th measurement is generated from the k-th component density. For more insight, we assume that each data point in the given sample, $x_i \in \mathbf{x}$, is produced as follows. As a first step, the k-th alternative component that produces the measurement is chosen, $C=k$. Then, according to its parameter vector β_k, the actual measurement x_i is produced. Since there are K components, the resulting mixture density value $f_X(x_i;\theta)$ is composed according to each associated component weight α_k.

Based on some dataset of observations, $\mathbf{x} = \{x_i\}_{i=1}^N$, concerning the underlying random variable X, we wish to model this data using the mixture model (4.16). If we assume that the data points are drawn independently from this distribution model, then the log-likelihood function is given by

$$\begin{aligned}
\ell(\theta;\mathbf{x}) &= \log f_X(\mathbf{x};\theta) \\
&= \log \prod_{i=1}^{N} f_X(x_i;\theta) \\
&= \sum_{i=1}^{N} \log f_X(x_i;\theta) \\
&= \sum_{i=1}^{N} \log\Big(\sum_{k=1}^{K} \alpha_k \; f_{X_k}(x_i|C=k;\beta_k)\Big)
\end{aligned} \tag{4.18}$$

and the goal of maximum likelihood estimation is to find the parameter θ that maximizes such a log-likelihood function. We immediately see the particularity of the functional form in the mixture models, where the maximization of the log-likelihood is now much more complex than with a standard procedure, due to the presence of the summation over k inside the logarithm. As a result, the maximum likelihood solution for the parameters no longer has a closed-form analytical solution.

4.2.3 Maximum likelihood for Gaussian mixtures

In order to give more insight about the problem concerning the maximum likelihood solution for the mixture parameters, it is worth to derive such an expressions in the context of the Gaussian mixture model (4.7), where the log-likelihood function is given by

$$\begin{aligned}\ell(\theta;\mathbf{x}) &= \log f_X(\mathbf{x};\alpha,\mu,V) \\ &= \sum_{i=1}^{N}\log\Big(\sum_{k=1}^{K}\alpha_k\ N(x_i;\mu_k,V_k)\Big)\end{aligned} \tag{4.19}$$

and the involving maximization of $\ell(\theta;\mathbf{x})$ is in order to find the mixing parameters $\alpha=(\alpha_1,\ldots,\alpha_K)$ and the component parameters $\mu=(\mu_1,\ldots,\mu_K)$ and $V=(V_1,\ldots,V_K)$.

Let us begin by writing down the conditions that must be satisfied at a maximum of the log-likelihood function. Setting the derivatives of $\ell(\theta;\mathbf{x})$ with respect to the means μ_k of the Gaussian components to zero, we can get

$$\begin{aligned}0 &= \frac{\partial\ell(\theta;\mathbf{x})}{\partial\mu_k} \\ &= -\sum_{i=1}^{N}\frac{\alpha_k\ N(x_i;\mu_k,V_k)}{\sum_{j=1}^{K}\alpha_j N(x_i;\mu_j,V_j)}V_k(x_i-\mu_k) \\ \mu_k &= \frac{1}{N_k}\sum_{i=1}^{N}\gamma(z_{ik})\ x_i\end{aligned} \tag{4.20}$$

where we assume that the covariance matrix V_k is nonsingular, so that multiplying by V_k^{-1} and rearranging we obtain the solution (4.20). The expressions $\gamma(z_{ik})$ and N_k are defined as

$$\gamma(z_{ik}) = \frac{\alpha_k\ N(x_i;\mu_k,V_k)}{\sum_{j=1}^{K}\alpha_j N(x_i;\mu_j,V_j)} \tag{4.21}$$

and

$$N_k = \sum_{i=1}^{N}\frac{\alpha_k\ N(x_i;\mu_k,V_k)}{\sum_{j=1}^{K}\alpha_j N(x_i;\mu_j,V_j)} = \sum_{i=1}^{N}\gamma(z_{ik}) \tag{4.22}$$

where we can interpret N_k as the effective number of points assigned to cluster k, whereas $\gamma(z_{ik})$ is the weighting factor that component k was responsible for generating the observation x_i. Note carefully that reading the form of the solution (4.20) we see that the mean μ_k for the k-th Gaussian component is obtained by taking a weighted mean of all of the points in the data set, according to the weighting factor.

To evaluate the maximum likelihood estimate for the covariance matrix, we set the derivative of $\ell(\theta; \mathbf{x})$ with respect to V_k equal to zero, and can get

$$\begin{aligned} 0 &= \frac{\partial \ell(\theta; \mathbf{x})}{\partial V_k} \\ V_k &= \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k)(x_i - \mu_k)' \end{aligned} \tag{4.23}$$

where we have followed a similar line of reasoning as above, making use of the result for the maximum likelihood solution for the covariance matrix of a single Gaussian. Note that solution (4.23) has the same form as the corresponding result for a single Gaussian fitted to the data, but again with each data point weighted by the corresponding weighting factor $\gamma(z_{ik})$, and the denominator given by the effective number of points associated with the corresponding component. If we consider the univariate case, $D = 1$, the solution for the variance is

$$\begin{aligned} \sigma_k^2 &= \frac{\sum_{i=1}^{N} \gamma(z_{ik}) \ (x_i - \mu_k)^2}{\sum_{i=1}^{N} \gamma(z_{ik})} \\ &= \frac{1}{N_k} \sum_{i=1}^{N} \gamma(z_{ik}) \ (x_i - \mu_k)^2 \end{aligned} \tag{4.24}$$

where the variance σ_k^2 for the k-th Gaussian component is obtained by taking a weighted mean of all the elements $(x_i - \mu_k)^2$, with respect to the weighting factor.

Finally, for the mixing coefficients we need to consider the constraint $\sum_{k=1}^{K} \alpha_k = 1$. Then, we can introduce the Lagrangian

$$L(\theta; \mathbf{x}) = \log f_X(\mathbf{x}; \alpha, \mu, V) + \lambda\Big(\sum_{k=1}^{K} \alpha_k - 1\Big) \tag{4.25}$$

and set the derivative of $L(\theta; \mathbf{x})$ with respect to α_k and λ equal to zero, so we get

$$\frac{\partial L(\theta; \mathbf{x})}{\partial \alpha_k} = \sum_{i=1}^{N} \frac{\gamma(z_{ik})}{\alpha_k} - \lambda = 0 \quad \Longleftrightarrow \quad \sum_{i=1}^{N} \gamma(z_{ik}) - \lambda \ \alpha_k = 0 \tag{4.26}$$

$$\frac{\partial L(\theta; \mathbf{x})}{\partial \lambda} = 1 - \sum_{k=1}^{K} \alpha_k = 0 \tag{4.27}$$

where again we see the appearance of the responsibilities. If we now sum with respect to K and make use of the constraint $\sum_{k=1}^{K} \alpha_k = 1$, we find

$$\sum_{k=1}^{K} \sum_{i=1}^{N} \gamma(z_{ik}) - \lambda \ \alpha_k = N - \lambda = 0 \quad \Longleftrightarrow \quad N = \lambda \tag{4.28}$$

and using this to eliminate λ and rearranging we get

$$\alpha_k = \frac{1}{N} \sum_{i=1}^{N} \gamma(z_{ik}) = \frac{N_k}{N} \tag{4.29}$$

so that the mixing coefficient for the k-th component is given by the average responsibility which that component takes for explaining the data points.

Now, it is easy to emphasize that the results (4.20), (4.23), and (4.29) do not constitute a closed-form solution for the parameters of the mixture model because the weighting factors $\gamma(z_{ik})$ depend on those parameters in a complex way through its expression (4.21). In particular, the problem is the presence of the responsibilities, $\gamma(z_{ik})$, which in turn refers to the mixing coefficients α_k, so that this is just a circular reference problem (i.e. chicken-and-egg dilemma). However, these results do suggest a simple iterative scheme for finding a solution to the maximum likelihood problem, which as we shall see turns out to be an instance of the EM algorithm for the particular case of the Gaussian mixture model.

4.3 Mixture models from the perspective of latent variables

As well as providing a framework for building more complex probability distributions, mixture models can also be used to cluster data. We therefore continue our discussion of mixture distributions by considering the problem of finding clusters in a set of data points. Since we are now interested in clustering it appears that one information is missing regarding the observed sample: the assignment of data points to the different clusters. In this section we introduce a formulation of mixture models in terms of latent variables, which will provide us with a deeper insight into this important models, and will also serve to motivate the Expectation-Maximization algorithm.

4.3.1 The responsibility role of the posterior probability

As we have just seen in the above section, the problem for the parameter estimates resulting from maximum likelihood solution is due to the presence of the factor $\gamma(z_{ik})$, which is interpreted as the weighting factor that mixture component k was responsible for generating the observation x_i. A key step is to know such a fundamental information.

Recalling the equation of finite mixture model, and using the sum and product rules for probability measures, we can rewrite the mixture density (4.4) in an equivalent form, as follows

$$f(\mathbf{x}) = \sum_{k=1}^{K} f(k)\ f(\mathbf{x}|k) \tag{4.30}$$

where we can view $\alpha_k = f(k)$ as the prior probability of picking the k-th component, and the density $N(\mathbf{x}|\mu_k, V_k) = f(\mathbf{x}|k)$ as the probability of observation $\mathbf{x}$ conditioned on k. Since we need to know the weighting factor, we have to complete our conditional reasoning. An important role is played by the posterior probabilities $f(k|\mathbf{x})$. Application of Bayes' theorem leads to

$$\begin{aligned} f(k|\mathbf{x}) &= \frac{f(k)\ f(\mathbf{x}|k)}{\sum_{j=1}^{K} f(j)\ f(\mathbf{x}|j)} \\ &= \frac{\alpha_k\ N(x_i;\mu_k,V_k)}{\sum_{j=1}^{K}\alpha_j N(x_i;\mu_j,V_j)} = \gamma_k(\mathbf{x}) \end{aligned} \tag{4.31}$$

where we have used the Gaussian densities as mixture components. The posterior probability can also be viewed as the responsibility that component k takes for explaining the observation $\mathbf{x}$. Note that by means of index j the denominator considers the data points belonging to the k-th cluster only, so as to satisfy normalizing requirements. In order to know which cluster k generates the observation $\mathbf{x}$, a new random variable is introduced.

4.3.2 Latent variables in the light of the unobserved heterogeneity

In the context of latent variables, the EM algorithm is a general method of finding the maximum likelihood estimate of the parameters of an underlying distribution from a given dataset when the data have missing values or are incomplete. The first occurs when the data indeed have missing values, due to problems with or limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable, but when the likelihood function can be simplified by assuming the existence of and values for additional but missing or hidden parameters. Note that it does not play a role if this unobserved data is technically unobservable or just not observed due to other reasons. Often, it is just an artificial construct in order to enable a numerical scheme.

Then we introduce the latent variable view of mixture distributions in which the latent variables can be interpreted as defining assignments of data points to specific components of the mixture. Let us denote this missing or hidden variable by Z, and the unobserved data by a vector $\mathbf{z}$. In our case, that is the context of mixture distributions, a good choice for Z is a K-dimensional binary random variable whose observed values have a $1 \times K$ representation given by

$$\mathbf{z} = (z_1, \ldots, z_k, \ldots, z_K)$$

where each entry z_k satisfies the following properties

$$z_k \in \{0, 1\} \qquad \sum_{k=1}^{K} z_k = 1 \tag{4.32}$$

so that the binary vector $\mathbf{z}$ is a sequence of the form

$$\mathbf{z} = (000 \ldots 0 \underset{z_k}{1} 0 \ldots 00) \tag{4.33}$$

in which a particular element z_k is equal to 1 and all other $K-1$ elements are equal to zero. Note that each row contains exactly one entry equal to one, which occurs if and only if component k produced measurement $\mathbf{x}$, otherwise it is zero. We see that there are K possible states for the vector $\mathbf{z}$ according to which k-th element z_k is nonzero.

In order to appreciate the role of latent variables, we need to visualize it in the light of the unobserved heterogeneity. By this choice, we introduce additional information into the process: the unobservable vector $\mathbf{z}$ tells us (in an oracular fashion) where the observation $\mathbf{x}$ comes from. Indeed, we will see that $\mathbf{z}$ tell us the probability density function that underlies a certain measurement. Nobody knows if we can really measure this hidden variable, and it is not important. We just assume that we have it, do some calculous, and see what happens. It may be surprising, but this problem of dealing with unobserved data in the end facilitates calculation of the maximum likelihood parameter estimate.

Having introduced in our model such an auxiliary variable $\mathbf{z}$, we are now interested to define the joint distribution of $(\mathbf{x}, \mathbf{z})$, and using the product rule it can be expressed as follows

$$f(\mathbf{x}, \mathbf{z}) = f(\mathbf{z})\ f(\mathbf{x}|\mathbf{z}) \tag{4.34}$$

that is in terms of a marginal distribution $f(\mathbf{z})$ and a conditional distribution of observation $\mathbf{x}$ given a particular value for $\mathbf{z}$. Concerning the marginal distribution over $\mathbf{z}$, we have just seen that the structure of this indicator variable inform us which component generate the observation $\mathbf{x}$, as follows

$$z_k = \begin{cases} 1 & \text{if } \mathbf{x} \text{ is generated from } k\text{-th component} \\ 0 & \text{otherwise} \end{cases} \qquad k = 1, \ldots, K \tag{4.35}$$

so that each entry z_k has the following probability distribution

$$f(z_k) = \begin{cases} f(z_k = 1) = \alpha_k \\ f(z_k = 0) = 0 \end{cases} \qquad k = 1, \ldots, K \tag{4.36}$$

where the marginal distribution over $\mathbf{z}$ is naturally specified in terms of the mixing coefficients, and we know that parameters $\{\alpha_k\}_{k=1}^K$ must satisfy the constraint $0 \leq \alpha_k \leq 1$ together with $\sum_{k=1}^K = 1$, in order to be valid probabilities. Like multinomial distribution, the marginal distribution over $\mathbf{z}$ can also be written in the form

$$f(\mathbf{z}) = \alpha_1^{z_1} \ \cdots \ \alpha_K^{z_K} = \prod_{k=1}^K \alpha_k^{z_k} \tag{4.37}$$

because the variable $\mathbf{z}$ uses a $1 \times K$ representation. According to this marginal distribution, it is easy to see that the conditional distribution of observation $\mathbf{x}$ given a particular value for $\mathbf{z}$ is exactly the k-th mixture component

$$f(\mathbf{x}|z_k = 1) = N_k(\mathbf{x}; \mu_k, \mathbf{V}_k) \qquad k = 1, \ldots, K \tag{4.38}$$

which can also be rewritten in the form

$$\begin{aligned} f(\mathbf{x}|\mathbf{z}) &= (N_1(\mathbf{x}; \mu_1, \mathbf{V}_1))^{z_1} \ \cdots \ (N_K(\mathbf{x}; \mu_K, \mathbf{V}_K))^{z_K} \\ &= \prod_{k=1}^K (N_k(\mathbf{x}; \mu_k, \mathbf{V}_k))^{z_k} \end{aligned} \tag{4.39}$$

due to the representation of variable $\mathbf{z}$. Therefore, the joint distribution (4.34) is given by

$$\begin{aligned} f(\mathbf{x}, \mathbf{z}) &= f(\mathbf{z}) \ f(\mathbf{x}|\mathbf{z}) \\ &= \prod_{k=1}^K \alpha_k^{z_k} \prod_{k=1}^K N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)^{z_k} \\ &= \prod_{k=1}^K \big(\alpha_k \ N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)\big)^{z_k} \end{aligned} \tag{4.40}$$

and the distribution of $\mathbf{x}$ is then obtained by summing the joint distribution over all possible states of variable $\mathbf{z}$ to give

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\mathbf{z}} f(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z}} f(\mathbf{z}) \ f(\mathbf{x}|\mathbf{z}) \\ &= \sum_{\mathbf{z}} \prod_{k=1}^K \alpha_k^{z_k} \ N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)^{z_k} \\ &= \sum_{k=1}^K \alpha_k \ N_k(\mathbf{x}; \mu_k, \mathbf{V}_k) \end{aligned} \tag{4.41}$$

where we have used the fact that the variable $\mathbf{z}$ is a K-dimensional vector with all elements equal to 0 except for a single element having the value 1, then $f(\mathbf{x})$ reduces to a sum of K contributions, one for each mixture component. Thus the distribution of $\mathbf{x}$ becomes a mixture model of the form (4.4). We have therefore found an equivalent formulation of the mixture model involving an explicit

laten variable. It might seem that we have not gained much by doing so. However, we are now able to work with the joint distribution $f(\mathbf{x}, \mathbf{z})$ instead of the whole distribution $f(\mathbf{x})$, and this will lead to significant simplifications, most notably through the introduction of the Expectation-Maximization algorithm.

For our clustering purposes, the quantity that will play an important role is the conditional probability of hidden variable $\mathbf{z}$ given a particular observation $\mathbf{x}$. In order to complete our conditional reasoning, we are interested to define the probability $f(z_k = 1|\mathbf{x})$, which is given by

$$\begin{aligned} f(z_k = 1|\mathbf{x}) &= \frac{f(z_k = 1)\ f(\mathbf{x}|z_k = 1)}{f(\mathbf{x})} \\ &= \frac{f(z_k = 1)\ f(\mathbf{x}|z_k = 1)}{\sum_{k=1}^{K} f(z_k = 1)\ f(\mathbf{x}|z_k = 1)} \\ \gamma(z_k) &= \frac{\alpha_k\ N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)}{\sum_{k=1}^{K} \alpha_k\ N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)} \end{aligned} \tag{4.42}$$

where we have used the Bayes theorem and the probability distribution (4.36) for the entry $z_k = 1$. We shall view the mixing coefficient α_k as the prior probability of variable $\mathbf{z}$, with $z_k = 1$. Whereas, the quantity $\gamma(z_k) = f(z_k = 1|\mathbf{x})$ is viewed as the corresponding posterior probability once we have observed $\mathbf{x}$. The quantity $\gamma(z_k)$ can also be viewed as the responsibility that component k takes for explaining the observation $\mathbf{x}$. In other words, under the distribution (4.36), for each entry z_k we can get

$$\begin{aligned} \mathbb{E}(z_k|\mathbf{x}) &= 0 \cdot f(z_k = 0|\mathbf{x}) + 1 \cdot f(z_k = 1|\mathbf{x}) \\ &= f(z_k = 1|\mathbf{x}) \end{aligned} \tag{4.43}$$

which is the expected value that component k produced observation $\mathbf{x}$, and this illustrate the responsibility role of the posterior probability.

4.4 The Expectation-Maximization algorithm

The Expectation-Maximization (EM) algorithm is an elegant and powerful statistical method for finding likelihood solutions when the likelihood functions have complex functional form and for handling models with latent variables, considered missing or incomplete. The EM algorithm formalizes an intuitive idea for obtaining parameter estimates: (i) initialize parameters, (ii) replace missing values by estimated values, (iii) estimate parameters, (iv) repeat step (ii) using estimated parameter values as true values, and step (iii) using estimated values as observed values, iterating until convergence. This idea has been in use for many years before the *missing information principle* provided the theoretical foundation of the underlying idea (Orchand and Woobdury 1972 [66]). The EM algorithm was first formulated by Dempster et al. (1977 [24]) in their seminal paper, where proof of general results about the behaviour of the algorithm was first given as well as a large number of applications. Considerable advances have been made since the introduction of the EM algorithm, and many works are devoted to its development and improvement. The EM algorithm has broad applicability in a variety of different contexts. The purpose of cluster analysis and the mixture model parameter estimation problem are probably the most widely used applications of the EM algorithm.

Suppose we have a dataset of observations, $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} = \{\mathbf{x}_i\}_{i=1}^{N}$. We can represent this dataset as an $N \times D$ matrix $\mathbf{X}$ in which the i-th row is given by $\mathbf{x}_i = (x_{i1}, \ldots, x_{id}, \ldots, x_{iD})$. We assume

that data $\{\mathbf{x}_i\}_{i=1}^N$ is observed and is generated by some distribution, and we wish to model this data using a mixture distribution. We shall refer to the actual dataset $\mathbf{x}$ as observed but incomplete data. Now suppose that, for each observation in $\mathbf{x}$, we were told the corresponding value of the hidden or missing variable $\mathbf{z}$. Similarly, the corresponding latent variables will be denoted by an $N \times K$ matrix $\mathbf{Z}$ with rows $\mathbf{z}_i = (z_{i1}, \ldots, z_{ik}, \ldots, x_{iK})$. We assume that a complete dataset exists, $\{\mathbf{x}, \mathbf{z}\} = \{\mathbf{x}_i, \mathbf{z}_i\}_{i=1}^N$, and that it consists of the incomplete observed data $\mathbf{x}$ and the unobserved data $\mathbf{z}$.

The goal of the EM algorithm is to find maximum likelihood parameter solutions for models having latent variables, and the observations can be viewed as incomplete data. Now consider the problem of maximizing the likelihood for the complete data $\{\mathbf{x}, \mathbf{z}\}$. If we have several observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$, it follows that for every observed data point $\mathbf{x}_i$ there is a corresponding latent variable $\mathbf{z}_i$. From joint distribution (4.40) the likelihood function takes the following form

$$\begin{aligned}\mathcal{L}_C(\theta; \mathbf{x}, \mathbf{z}) &= \prod_{i=1}^{N} f(\mathbf{x}, \mathbf{z}; \theta) \\ &= \prod_{i=1}^{N} \prod_{k=1}^{K} \alpha_k^{z_{ik}} \; N_k(\mathbf{x}; \mu_k, \mathbf{V}_k)^{z_{ik}}\end{aligned} \tag{4.44}$$

which is called the complete-data likelihood function, where z_{ik} denotes the k-th component of latent variable $\mathbf{z}_i$. Note that this new likelihood function is in fact a random variable since the missing information $\mathbf{z}$ is unknown, random, and presumably governed by an underlying distribution $Q(\mathbf{z})$. That is, we can think of $\mathcal{L}_C(\theta; \mathbf{x}, \mathbf{z}) = h_{\mathbf{x},\theta}(\mathbf{z})$ for some function $h_{\mathbf{x},\theta}$, where the observed data $\mathbf{x}$ and the parameter θ are constant, whereas $\mathbf{z}$ is a random variable. The original likelihood $\mathcal{L}(\theta; \mathbf{x})$ is referred to as the incomplete-data likelihood function. Taking the logarithm of (4.44), the log-likelihood function of the complete data is

$$\begin{aligned}\ell_C(\theta) &= \log p(\mathbf{X}, \mathbf{Z}; \theta) = \log \prod_{i=1}^{N} f(\mathbf{x}_i, \mathbf{z}_i; \theta) \\ &= \log \prod_{i=1}^{N} \prod_{k=1}^{K} f(z_{ik} = 1)^{z_{ik}} \; f(\mathbf{x}_i | \mathbf{z}_{ik} = 1; \theta)^{z_{ik}} \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \; \log f(z_{ik} = 1) + z_{ik} \; \log f(\mathbf{x}_i | \mathbf{z}_{ik} = 1; \theta) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \; \log \alpha_k + z_{ik} \; \log f(\mathbf{x}_n | \mathbf{z}_{ik} = 1; \theta)\end{aligned} \tag{4.45}$$

and this leads to a much simpler solution to the maximum likelihood problem. Indeed, comparison with the log-likelihood function (4.18) for the incomplete data shows that the summation over k and the logarithm have been interchanged. The logarithm now acts directly on the component distribution.

The purpose of the EM algorithm is the iterative computation of maximum likelihood estimators when observations can be viewed as incomplete data. The basic idea of the EM algorithm is to associate a complete data model to the incomplete structure that is observed in order to simplify the computation of maximum likelihood estimates. Similarly, a complete data likelihood is associated to the complete data model. In practice, however, we are not given the complete dataset $\{\mathbf{x}, \mathbf{z}\}$, but only the incomplete data $\{\mathbf{x}\}$. Our state of knowledge of the values of the latent variables in $\{\mathbf{z}\}$ is given only by the posterior distribution $Q(\mathbf{z}) = f(\mathbf{z}|\mathbf{x}; \theta)$. Because we cannot use the complete-data log-likelihood, we consider instead its expected value under the posterior distribution of the latent variable.

Compared to the maximum likelihood approach which involves just maximizing a log-likelihood, the EM algorithm makes one step in between: the calculation of the expectation. In the framework of maximum likelihood problem, we attempt to find the particular vector θ that maximizes the expectation of log-likelihood function such that

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}}\, \mathbb{E}_{Q(\mathbf{z})}(\ell_C(\theta, \theta^{old})) \tag{4.46}$$

and we wish to find the solution θ^{new}. Note that in order to calculate the expectation, an initial parameter estimate θ^{old} is needed. But still, the whole term will depend on a parameter θ.

EM algorithm is usually described as two steps (the E-step and the M-step), but let us break it down into four steps (Gupta and Chen 2011 [39]). The next section details the following steps: initialization of parameters, expectation step (E-step), maximization step (M-step), convergence check.

4.4.1 Step 1: Initialization of parameters

Before to start the iteration cycle where expectation step alternates with maximization step, we need an initial estimate of parameter θ. If we consider a predefined number of iterations $m = 0, 1, \ldots, M$, the algorithm is initialized by choosing some starting value for the parameter which is denoted by $\theta^{start} \equiv \theta^{(0)}$. For the m-th iteration, the current parameter value $\theta^{old} \equiv \theta^{(m)}$ is used to find the posterior distribution of the latent variables given by $Q(\mathbf{z}) = f(\mathbf{z}|\mathbf{x}; \theta^{(m)})$. The E-step find the expectation of the complete data log-likelihood evaluated for some general parameter value θ, whereas the subsequent M-step maximize this expectation. Then a pair of successive E and M steps gives rise to a revised estimate $\theta^{new} \equiv \theta^{(m)}$ which will be used in the $(m+1)$-th iteration.

It is not uncommon to initialize the iterative procedure by randomly choosing K of the N samples and making these the first estimates of the cluster means, setting the first estimate of the covariances to be identity matrices, and the first guess at the weights $\alpha_1 = \cdots = \alpha_K = 1/K$. Common wisdom is that initializing by first doing a cheaper clustering will generally produce more desirable results. Therefore, it is often used to run the K-means algorithm in order to find a suitable initialization for the mixture model that is subsequently adapted using EM algorithm. The covariance matrices can conveniently be initialized to the sample covariances of the clusters found by the K-means algorithm, and the mixing coefficients can be set to the fractions of data points assigned to the respective clusters.

4.4.2 Step 2: Expectation step (E-step)

The expectation step consists of calculating the expected value of the complete data log-likelihood function. Assuming that this function is governed by an underlying distribution $Q(\mathbf{z})$, the expected value is

$$\begin{aligned}\mathbb{E}_{Q(\mathbf{z})}(\ell_C(\theta)) &= \mathbb{E}_{Q(\mathbf{z})}\Big(\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\log\alpha_k + z_{ik}\log f(\mathbf{x}_i|\mathbf{z}_{ik}=1;\theta)\Big) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}_{Q(\mathbf{z})}(z_{ik})\log\alpha_k + \mathbb{E}_{Q(\mathbf{z})}(z_{ik})\log f(\mathbf{x}_i|\mathbf{z}_{ik}=1;\theta)\end{aligned} \tag{4.47}$$

where we have used the linearity of the expectation operator. In order to complete the present step, we have to consider how the expected value in (4.47) looks like, with respect to $Q(\mathbf{z})$. On the basis of the assumptions we made concerning the latent variable $\mathbf{Z}$, the underlying distribution is given by

$$Q(\mathbf{z}) = f(\mathbf{z}|\mathbf{x};\theta)$$

so that each expected value has to be evaluated as follows

$$\begin{aligned}\mathbb{E}_{Q(\mathbf{z})}(z_{ik}) &= 0 \cdot f(z_{ik}=0|\mathbf{x}_i;\theta) + 1 \cdot f(z_{ik}=1|\mathbf{x}_i;\theta) \\ &= f(z_{ik}=1|\mathbf{x}_n;\theta) \\ \gamma(z_{ik}) &= \frac{f(z_{ik}=1)\ f(\mathbf{x}_i|z_{ik}=1;\theta)}{\sum_{k=1}^{K} f(z_{ik}=1)\ f(\mathbf{x}_i|z_{ik}=1;\theta)}\end{aligned} \tag{4.48}$$

which is the probability that component k was active while measurement i was produced, that is the responsibility that component k takes for explaining the observation $\mathbf{x}_i$. All the terms on the right side can be easily calculated. According to the previous step, the current parameter value $\theta^{old} \equiv \theta^{(m)}$ is used to evaluate the posterior distribution of the latent variables. The quantity $f(z_{ik}=1)$ is the probability of choosing the k-th component, which is nothing else than the parameter $\alpha_k^{(m)}$. The quantity $f(\mathbf{x}_i|z_{ik}=1;\theta)$ just comes from the k-th component density function.

4.4.3 Step 3: Maximization step (M-step)

In the previous steps we have chosen some initial values for the parameters, $\theta^{old} \equiv \theta^{(m)}$, and used these to evaluate the responsibilities. In the maximization step we need to determine the revised parameter estimate $\theta^{new} \equiv \theta^{(m)}$. In order to do this computation, we then keep the responsibilities fixed and perform the optimization by the following maximization

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}}\ \mathbb{E}_{Q(\mathbf{z})}(\ell_C(\theta, \theta^{old})) \tag{4.49}$$

where the expected value of the complete data log-likelihood function is

$$\mathbb{E}_{Q(\mathbf{z})}(\ell_C(\theta)) = \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}_{Q(\mathbf{z})}(z_{ik}) \log \alpha_k + \mathbb{E}_{Q(\mathbf{z})}(z_{ik}) \log f(\mathbf{x}_i|\mathbf{z}_{ik}=1;\theta)$$

which has already been computed in the E-step. Setting the respective partial derivatives to zero, this leads to closed form solutions for $\mu_k^{(new)}$, $V_k^{(new)}$ and $\alpha_k^{(new)}$ given by (4.20), (4.23), and (4.29).

4.4.4 Step 4: Convergence check

The algorithm is initialized by choosing some starting value for the parameter, $\theta^{start} \equiv \theta^{(0)}$. On the basis of the current parameter values, $\theta^{old} \equiv \theta^{(m)}$, the E-step computes the expected value of the complete data log-likelihood function. In the subsequent M-step, this expectation is maximized, so that a new updated parameter values is found, $\theta^{new} \equiv \theta^{(m)}$. Then, a pair of successive E and M steps gives rise to a revised estimate θ^{new}. In the present step, we need to check for convergence of either the log-likelihood or the parameter values. Once we have fixed the convergence criterion, if this criterion is not satisfied, then let

$$\theta^{old} \quad \leftarrow \quad \theta^{new} \tag{4.50}$$

and return to Step 2. This whole procedure is repeated until the stopping criterion is satisfied, that is until the difference of change between the parameter updates become very small.

A numerical implementation of the Expectation-Maximization algorithm is reported in Chapter 6. The code has been developed using MATLAB, version `7.10.0.499 (R2010a)`.

Mixture models have experienced increased interest and popularity over last decades. Expectation-Maximization algorithm and its variants are regularly used to solve a broad range of today's estimation problems. Some useful references are Titterington et al. (1985 [85]), McLachlan and Krishnan (1997 [62]), McLachlan and Peel (2000 [63]), Bishop (2006 [11]), Fruhwirth-Schnatter (2006 [31]), Theodoridis and Koutroumbas (2008 [84]), Everitt et al. (2011 [29]), among others.

Chapter 5

Model Parameters: Identification and Estimation

In this chapter we provide the procedure for identifying and estimating the two main elements of the present model: the potential function U and the diffusion parameter κ. In order to estimate and fit the potential function, we consider a scaled potential and postulate a parametric form of the invariant distribution based on the framework of finite mixture model. The estimate of the diffusion parameter refers to a preliminary discretization of the equation model, together with a combination of the effects of time discretization and random disturbances. In such a way, the new unknown parameter can be estimated by means of a regression procedure. The estimated model is tested in various ways. The analysis of the model residuals is important in order to capture main dependence characteristics of the observed data, and it is related to the estimate of the diffusion parameter and the corresponding changes in volatility within the variance of the model residuals. The performance of the fitted model and its prediction accuracy is tested using the mean square prediction error in a context of cross-validation procedure. More important, the model is tested in terms of predicting the direction of the next price move, by means of the correct up-down moves. Concerning the volatility measures, the potential function model is able to provide an estimate of such an unobserved parameter, so we briefly address to this estimate which can be used as a valid alternative to more traditional techniques. The model is able to generate copies of the observed price series with the same distributional properties, which is useful for applications such as Monte Carlo analysis, scenario testing, and other studies that require a large number of independent price trajectories. We provide numerical schemes in order to simulate price process. Finally, by using an approach based on Monte Carlo simulations, we implement a goodness-of-fit procedure in order to access the validity of the model obtained by testing if there is a lack-of-fit between the stochastic differential equation model and the collection of data used to estimate the drift and the diffusion.

An underlying assumption of the present model is that the potential function and the long-term volatility do not change with time. As we have described in Chapter 2, global commodity markets have experienced significant price swings in recent years. In such a context of new market conditions, new attraction regions can form, changing the shape of the potential and the magnitude of the long-term volatility. In order to investigate further the behavior of the potential model, we have considered a price dataset to ensure the availability of as long a span of high quality data as possible. At first we employ daily spot price data concerning crude oil. Additionally, we have also tested the behaviour of the present model with an agricultural commodity, using soybean data.

In order to give some insight into the whole procedure and its empirical results, we estimate the unknown model parameters using historical daily spot prices of West Texas Intermediate (WTI) crude

oil. The full sample we have considered is a data period covering more than 25 years, and it consists in the daily closing prices from January 2, 1986 to August 13, 2012. For more intuition in describing the estimate procedure we use a sub-sample from January 4, 1993 to December 30, 1999, which yields 1795 observations. The WTI data consists of Spot Price FOB, and the prices are in US Dollars per Barrel ($/bbl). More details based on the full sample of data are discussed in Section 5.10.

5.1 Estimation of potential function

Consider a continuous time price process $(P(t))_{t\in\mathbb{T}}$, with the one-dimensional state variable, $P(t) \in \mathbb{R}$, where the time index set is $\mathbb{T} = [0, T]$. We postulate that the evolution of the state variable is described according to the potential model (3.46), so that the price value $P(t)$ evolves according to the stochastic differential equation model

$$dP(t) = -U'(P(t))\ dt + \kappa\ dB(t) \tag{5.1}$$

where $U : A \subseteq \mathbb{R} \to \mathbb{R}$ is a potential function that governs the evolution of the price process, the stochastic process $(B(t))_{t\in\mathbb{T}}$ denotes the standard Brownian motion, and $\kappa \in \mathbb{R}$ is the diffusion coefficient, a scalar factor that measures the magnitude of random fluctuations, that is the influence of the Brownian motion on the evolution of the price process.

Recall that, as we have seen in Section 3.5.2, for a continuous time process $(P(t))_{t\in\mathbb{T}}$ evolving according to equation model (5.1), the distribution of P approaches an equilibrium in a weak sense, which is an invariant distribution with the following density function

$$\begin{aligned} f(p;\kappa,U) &= \frac{1}{Z} e^{-\frac{2\ U(p)}{\kappa^2}} \\ &= e^{-\frac{2\ U(p)}{\kappa^2}} \end{aligned} \tag{5.2}$$

where the normalization constant can be taken as $Z = 1$, without loss of generality.

In order to model the price evolution by letting the price process be governed according to a diffusion model (5.1) with its invariant distribution (5.2), we need to know the potential function U and the diffusion parameter κ. Our approach is to learn such unknown elements starting from historical data. Now, we suppose that the observed price series $\{p(t_i)\}_{i=1}^n$ is a realization of the process (5.1) at discrete time points. In the conceptual framework of the potential, these data will be viewed as locations at successive times, $\{t_i\}_{i=1}^n$, of an object (a particle) moving along the path of the trajectory of the process (5.1). Starting from the dataset given by the empirical data, we are able to learn a potential U from such a trajectory, and subsequently to get an estimate for the diffusion parameter κ.

First step is to derive the structure of the potential function. Inverting equation (5.2), it is possible to express the potential function via the density of the invariant distribution as

$$U(p; f, \kappa) = -\frac{\kappa^2}{2} \ln(f(p)) \tag{5.3}$$

where the potential is the minus logarithm of the density value, scaled by the proper factor $\kappa^2/2$. Our goal is to know the potential and the parameter κ. First of all, it is easy to realize that the preliminary step is to identify the density function f. From the observed price series, we can estimate the invariant distribution and the corresponding density function. In such a way, assuming that we are able to get an estimate of the density, then an estimate of the potential is

$$\widehat{U}(p;\kappa) = -\frac{\kappa^2}{2} \ln(\widehat{f}(p)) \tag{5.4}$$

where $\widehat{f}$ is the empirical density of the observations. Even though the density of the invariant distribution f can be estimated from the data, this alone does not give us the estimate of the potential, since the representation (5.3) also involves the unknown parameter κ. As a preliminary step, to get around such a problem, we do not estimate the potential function itself, but the *scaled* potential defined as follows

$$U(p) = -\frac{\kappa^2}{2}\ln(f(p)) \quad \Longleftrightarrow \quad \frac{2}{\kappa^2}U(p) = -\ln(f(p)) = G(p) \tag{5.5}$$

where the function G is a preliminary version of the potential by means of a convenient scalar factor. The scaled potential G can be estimated as

$$\widehat{G}(p) = -\ln(\widehat{f}(p)) \tag{5.6}$$

where $\widehat{f}$ is some estimate of the empirical density of the given observations price series.

It is interesting to note that the above construction of potential function is exactly what we have addressed in Section 3.4, where potential and probability density are related each other in order to provide a way of viewing randomness from an equilibrium perspective, that is to view probability density from a perspective of forces/potentials, hidden behind it. This is an interesting aspect to study further.

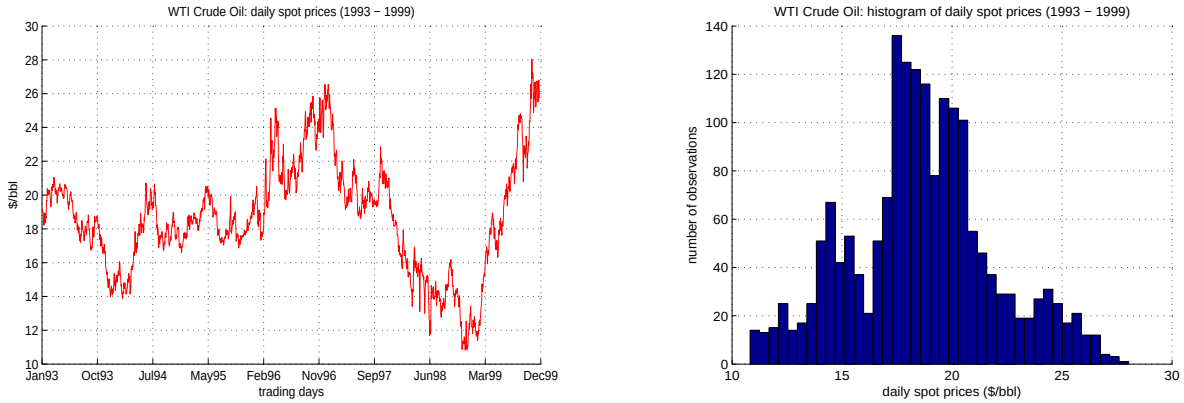
In order to provide an estimate of the density function, in the next Section 5.2 we adopt the framework of finite mixture models using a mixture of Gaussians (see Chapter 4).

5.2 Density estimation

In this section we discuss the fundamental step of fitting the multimodal density of the invariant distribution. Starting from the dataset of the observed price series, $\{p(t_i)\}_{i=1}^n$, the goal is to estimate the density function by means of fitting the resulting histogram of the historical data.

Recalling that the derivative of the potential needs to be quickly and accurately evaluated, the estimation method has to provide an analytical expression for the density, with the advantage that such a derivative can be computed directly from the expression provided by the method we have chosen.

The multimodal density can be estimated in numerous ways. As a first step, we might address ourselves to fit a polynomial of high degree, which is the fastest and most accurate way to apply the model, because such a method allows for fast calculation of the potential's derivative. In the present analysis, our main concern is to postulate a parametric form of the invariant price distribution in the framework of finite mixture models. In order to fit the potential we provide a numerical implementation of Expectation-Maximization algorithm for a finite mixture of Gaussians. This way of fitting the potential model extends the original approach. Mixture modeling is a rapidly developing area, with the range of applications exploding. Finite mixture models provide a straightforward, but very flexible extension of classical statistical models. There exist various features of finite mixture distributions that render them useful in statistical modeling. The most striking property of a mixture density is that the shape of the density is extremely flexible. Indeed, from the data-oriented perspective, it turns out that statistical models that are based on finite mixture distributions are able to capture many specific features of real data, such as multimodality, skewness, and kurtosis. Moreover, mixture modeling is particularly useful in the multivariate extension of the potential model.



(a) (b)

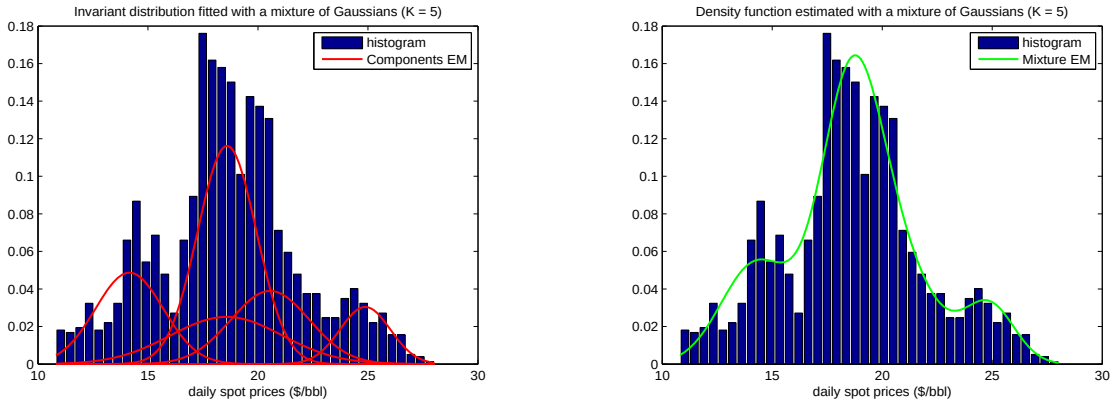
Figure 5.1: WTI crude oil daily spot prices over the trading days from January 4, 1993 to December 30, 1999. The observed price series shows (a) significant price swing, while (b) on the histogram can be clearly seen the phenomenon of price clustering. A summary of descriptive statistics is reported in Table 5.2.

5.2.1 Data

The dataset concerns the sub-sample of daily spot prices for WTI crude oil from January 4, 1993 to December 30, 1999, which yields 1795 observations. The WTI data consists of Spot Price FOB, and the prices are in US Dollars per Barrel ($/bbl). Figure 5.1 shows price series and the corresponding histogram, whereas Table 5.2 provides a summary of descriptive statistics. In the observed data series (a) prices regularly move between multiple attracting regions, and such a behaviour is related to the significant price swings which global commodity markets have experienced in recent years. The time spent at a given region is unpredictable and can be long or short. In this sub-sample period the preferred regions are around approximately 14, 18, 20 and 24 dollars per barrel. The price lies outside these stable regions only relatively briefly, and there it becomes rather unstable. When the price fluctuates in an unstable region, i.e. around 16 dollars per barrel, its evolution will likely either drop towards 14 dollars per barrel or rise towards 18, depending on whether the market is falling (bearish) or rising (bullish). On the histogram (b) can be clearly seen the so-called phenomenon of price clustering where the price concentrate in a number of attraction regions (or clusters). Such a behaviour of the price is a well-known phenomenon in commodity markets.

5.2.2 Mixture of Gaussians fit

To fit a polynomial of high degree can be a fast and accurate enough way to apply the model, and it also provides a simple analytical expression to directly compute the potential's derivative. However, this fitting method don't give us the flexibility needed in statistical modeling. The structure resulting from a finite mixture model is indeed more capable to face various statistical issues and practical problems in rather different applications. In Chapter 4 we have addressed to the framework of finite mixture models. In order to fit the potential function, we can postulate a parametric form of the invariant price



(a) (b)

Figure 5.2: An estimate of the mixture model (a) with $K = 5$ Gaussian components (b) and the corresponding mixture density function, according to the mixture model parameters in Table 5.1, using EM algorithm.

	K	1	2	3	4	5	Mixture	93-99
Mean	μ_k	14.1560	18.5646	18.5949	20.5521	24.8731	18.6443	18.6443
Variance	σ_k^2	2.3692	6.4880	1.7966	3.1845	1.3039	10.6545	10.6604
Mixing	α_k	0.1878	0.1605	0.3903	0.1743	0.0871	-	-

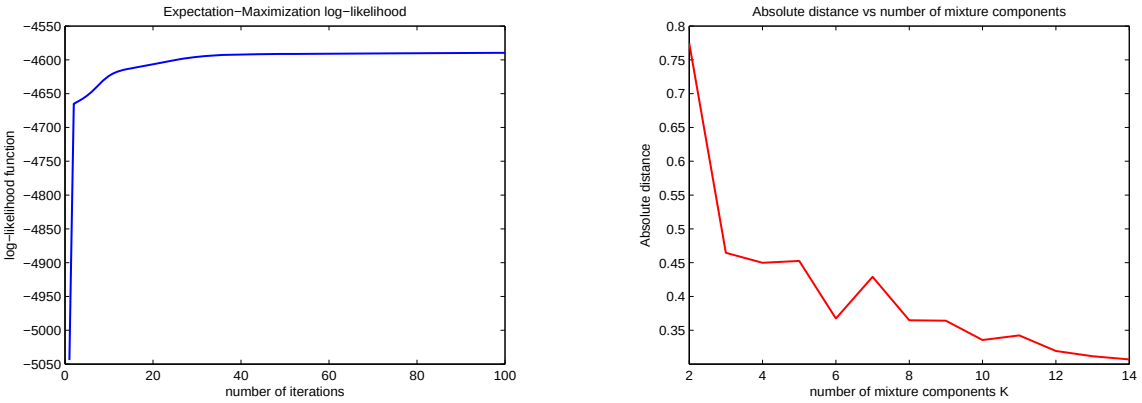
Table 5.1: An estimate of the mixture model parameters, with $K = 5$ Gaussian components, using EM algorithm. The estimated mixture components and the corresponding mixture density are reported in Figure 5.2.

distribution, $f_P(p;\theta)$, in the context of a finite mixture of Gaussians, with K components, as follows

$$f_P(p;\theta) = \sum_{k=1}^{K} \alpha_k \; f_{P_k}(p;\beta_k) = \sum_{k=1}^{K} \alpha_k \; N_k(p;\mu_k,\sigma_k^2) \tag{5.7}$$

where each k-th marginal component is a Normal distribution with mean μ_k and variance σ_k^2, whereas α_k are the mixing coefficients. In order to estimate the matrix parameters $\theta = (\mu, \sigma^2, \alpha)$, we have numerically implemented an algorithm of Expectation-Maximization (EM) method. In Section 6.2 we have reported the numerical code with a description of the listing.

Assuming that we have preliminarily specified the number of components, we consider a mixture model with $K = 5$ Gaussian components, then we are interested to fit the invariant distribution. According to the sub-sample dataset of observed price series, $\{p(t_i)\}_{i=1}^n$, the numerical implementation of EM algorithm give us an estimate of the parameters concerning means, variances and mixing coefficients. The Gaussian mixture components and the corresponding estimate of density function are illustrated in Figure 5.2. The parameter estimates are reported in Table 5.1. The column "Mixture" reports mean, 18.6443, and variance, 10.6545, of the estimated mixture model, the standard deviation is 3.2641, and all these parameters are in full agreement with the descriptive statistics of the dataset



(a) (b)

Figure 5.3: The behavior of (a) the convergence of the likelihood function for the EM algorithm and (b) the absolute distance between the histogram and the estimate of mixture density plotted against the corresponding number of mixture components K.

(Table 5.2). Figure 5.3 (a) shows the behavior of the likelihood function for the EM algorithm, which already has a fast convergence around 40 steps. Now, we want to consider the robustness of the estimation method with respect to the number of components. Modeling of data by a finite mixture model requires some specification of K, the number of components. In the above estimates, we assumed that the number of components is known, and we fixed it to $K = 5$. Then, it is interesting to investigate whether different values for K significantly affect the estimates of the density function. The goal of this section was to estimate the density function by means of fitting the resulting histogram of the historical data. In order to compare the different estimates corresponding to changes in the number of components K, we compute the distance between the histogram and the estimate of mixture density, by means of the sum of absolute errors. For that, we consider the distance for the number of components $K = 2, 3, \ldots, 14$. The results for different values of K are illustrated in Figure 5.3 (b), where we have plotted the computed absolute error against the corresponding number of components K. As expected, when the number of mixture components increases, the distance between the histogram and the estimate of mixture density tends to become small. In such a way, it is easy to say that a good fit can be obtained by increasing the number K. But, it is also important to avoid some complexity in the structure of mixture model, so the resulting mixture density can be modeled, for example, using mixture components belonging to different parametric family. Note that if we employ 7 components the error is greater than in the case of 6 components, which is equal to 8 or 9 components, then employing a mixture with 6 components yields a fast and accurate enough way to apply the model. How the different values for K can affect the estimates of the density function can be also seen in Figure 5.4 (b), where we consider the behaviour of scaled potential function with respect to the log-inverse of the histogram.

5.3 Estimation of the diffusion parameter

In this section our goal is to provide an estimate for the unknown diffusion parameter κ. Given the sub-sample dataset of observed price series, $\{p(t_i)\}_{i=1}^n$, the basic problem assumes the equation model (5.1) and seeks to learn the potential function U from the empirical trajectory of the historical data. In the previous Section 5.2, using some estimation method, with main focus on finite mixture of Gaussians, we provided an estimate of the scaled potential, $\widehat{G}$, so we can write

$$\widehat{U}(p;\kappa) = \frac{\kappa^2}{2}\widehat{G}(p) \tag{5.8}$$

that is an estimate of the potential U by means the estimated scaled potential $\widehat{G}$ with the scalar factor $\kappa^2/2$. Since we are interested in knowing the potential function, we need to estimate the parameter κ.

In order to provide an estimate of the diffusion parameter κ, we proceed as follows. Assuming regularity conditions on function U, and that the observation times are close together, so we can consider a small enough time interval, $h = t_{i+1} - t_i$, then one can set down an approximation to equation model (5.1), in terms of scaled potential G, by means of the following discretization

$$p(t_{i+1}) - p(t_i) = -\frac{\kappa^2}{2}G'(p(t_i))\ (t_{i+1} - t_i) + \kappa\ \sqrt{t_{i+1} - t_i}\ N(0,1) \qquad i = 1, \ldots, n-1 \tag{5.9}$$

where the process $(\sqrt{t_{i+1} - t_i}N(0,1))_{i=1}^{n-1}$ denotes the increments of the Brownian motion over the time intervals, $dB(t_i, t_{i+1})$, so its components are independent normally distributed random variables having mean $\mu = 0$ and variance $\sigma^2 = t_{i+1} - t_i$, with standard deviation $\sigma = \sqrt{t_{i+1} - t_i}$. In Section 5.8.1 we give more details for the above discretization (5.9).

Note that, since only observations at discrete times are available, it is in general impossible to separate the effects of time discretization and the random disturbances on fluctuations of the time series. Thus we combine these effects into a new parameter, $\gamma = \kappa^2(t_{i+1} - t_i)$, so we can rewrite the discretization equation (5.9) of the model as follows

$$p(t_{i+1}) - p(t_i) = -\frac{\gamma}{2}G'(p(t_i)) + \epsilon(t_i) \qquad i = 1, \ldots, n-1 \tag{5.10}$$

where the components of the process $(\epsilon(t_i))_{i=1}^{n-1}$ are

$$\epsilon(t_i) \sim N(0, \gamma) \qquad i = 1, \ldots, n-1 \tag{5.11}$$

that is independent normally distributed random variables with mean $\mu = 0$ and variance $\sigma^2 = \kappa^2(t_{i+1} - t_i) = \gamma$. Now, given an estimate of the scaled potential, $\widehat{G}$, we can consider the variables in equation (5.10) as follows

$$(p(t_{i+1}) - p(t_i))_{i=1}^{n-1} = Y \tag{5.12}$$

$$\Big(-\frac{G'(p(t_i))}{2}\Big)_{i=1}^{n-1} = X \tag{5.13}$$

where, for every $i = 1, \ldots, n-1$, the variable $Y_i = p(t_{i+1}) - p(t_i)$ denotes the price increments over the time interval $[t_i, t_{i+1}]$, and the variable $X_i = -G'(p(t_i))/2$ denotes the minus derivative of the scaled potential, with respect to $1/2$. Then one has the following linear regression model

$$Y = \gamma X + \epsilon \tag{5.14}$$

where we see that γ is the unknown parameter of a linear regression of the price increments Y (the response variable) on the scaled potential X (the predictor variable), without the intercept term (see Hahn 1977 [41], Turner 1960 [86], Eisenhauer 2003 [28]). In order to fit the regression line, the regression parameter γ can be estimated in a fast and accurate way by the method of least squares. For that, consider the objective function Q and the following minimization problem

$$\min_{\gamma} Q(\gamma) = \min_{\gamma} \sum_{i=1}^{n-1} \epsilon_i = \min_{\gamma} \sum_{i=1}^{n-1} (y_i - \gamma x_i)^2 \tag{5.15}$$

so that the objective is to minimize the sum of squared errors ϵ_i of the linear regression model (5.14). The first-order condition is

$$\frac{dQ}{d\gamma} = \frac{d}{d\gamma} \sum_{i=1}^{n-1} (y_i - \gamma x_i)^2 = 0 \tag{5.16}$$

and we get the normal equation as

$$\begin{aligned} \sum_{i=1}^{n-1} 2(y_i - \gamma x_i)(-x_i) &= 0 \\ -2\sum_{i=1}^{n-1} x_i y_i + 2\sum_{i=1}^{n-1} \gamma x_i^2 &= 0 \\ \gamma \sum_{i=1}^{n-1} x_i^2 &= \sum_{i=1}^{n-1} x_i y_i \\ \gamma &= \frac{\sum_{i=1}^{n-1} x_i y_i}{\sum_{i=1}^{n-1} x_i^2} \end{aligned} \tag{5.17}$$

and, hence, the derived second-order condition

$$\frac{d^2 Q}{d\gamma^2} = \frac{d^2}{d\gamma^2} \sum_{i=1}^{n-1} (y_i - \gamma x_i)^2 = 2\sum_{i=1}^{n-1} x_i^2 > 0 \tag{5.18}$$

clearly guarantees that the expression (5.17) is indeed a minimum for the objective function Q. Considering substitutions in (5.12) and (5.13) we get that the value of parameter γ can be expressed as

$$\begin{aligned} \widehat{\gamma} &= \frac{\sum_{i=1}^{n-1} x_i y_i}{\sum_{i=1}^{n-1} x_i^2} \\ &= \frac{\sum_{i=1}^{n-1} (p(t_{i+1}) - p(t_i))(-\frac{\widehat{G}'(p_i)}{2})}{\sum_{i=1}^{n-1} (-\frac{\widehat{G}'(p_i)}{2})^2} \\ &= \frac{-\frac{1}{2}\sum_{i=1}^{n-1} (p(t_{i+1}) - p(t_i))\widehat{G}'(p_i)}{\frac{1}{4}\sum_{i=1}^{n-1} (\widehat{G}'(p_i))^2} \\ &= -2\frac{\sum_{i=1}^{n-1} (p(t_{i+1}) - p(t_i))\widehat{G}'(p_i)}{\sum_{i=1}^{n-1} (\widehat{G}'(p_i))^2} \end{aligned} \tag{5.19}$$

which requires the evaluation of the derivative of estimated scaled potential, $\widehat{G}'(p(t_i))$, with respect to the current value of price process, $p(t_i)$.

5.3.1 Computation of the first derivative of scaled potential G

Recalling expression (5.5), the scaled potential G is the minus logarithm of the density function, then its first derivative is

$$G'(p) = \frac{dG(p)}{dp} = \frac{d[-\ln(f(p))]}{dp} = -\frac{1}{f(p)} f'(p) \tag{5.20}$$

that is the logarithmic derivative of the density function, which has to be evaluated with respect to the observed price series, $\{p(t_i)\}_{i=1}^n$. Because the derivative (5.20) depends on the density function, f, we have to take into account the different methods used for estimating the density and, hence, its different analytical structure.

When we adopt a fit by means of a finite mixture model with Gaussian components, then the mixture density is

$$f(p) = \sum_{k=1}^{K} \alpha_k \; N_k(p; \mu_k, \sigma_k) = \sum_{k=1}^{K} \alpha_k \; \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{p-\mu_k}{\sigma_k}\right)^2} \tag{5.21}$$

so that the derivative of the scaled potential is

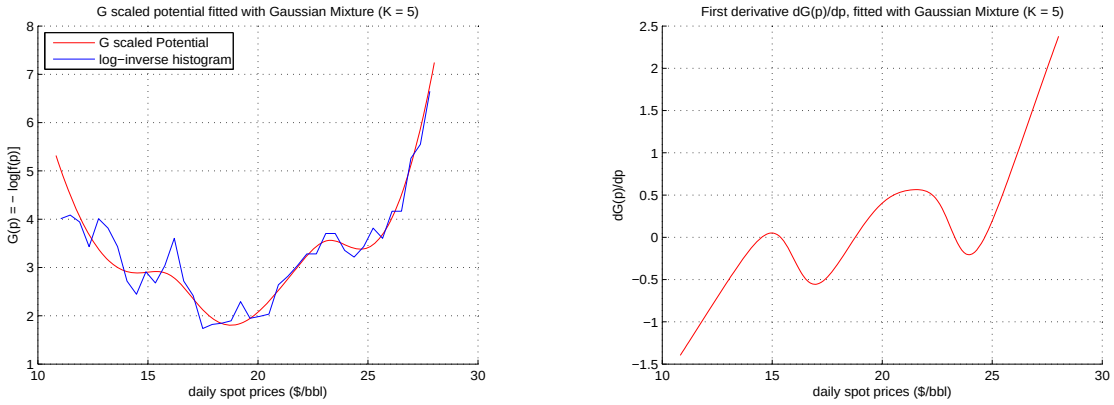
$$\begin{aligned} G'(p) &= [-\ln(f(p))]' = -\frac{1}{f(p)} f'(p) \\ &= -\frac{\sum_{k=1}^{K} \alpha_k \; \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{p-\mu_k}{\sigma_k}\right)^2} \; \frac{-(p-\mu_k)}{\sigma_k^2}}{\sum_{k=1}^{K} \alpha_k \; \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{p-\mu_k}{\sigma_k}\right)^2}} \\ &= -\frac{\sum_{k=1}^{K} \alpha_k \; N(p; \mu_k, \sigma_k) \; \frac{-(p-\mu_k)}{\sigma_k^2}}{\sum_{k=1}^{K} \alpha_k \; N(p; \mu_k, \sigma_k)} \end{aligned} \tag{5.22}$$

where, as we know, K is the number of mixture components, and the adjoint factor $-(p-\mu_k)/\sigma_k^2$ comes from the expression for the first derivative of a normal probability density function.

5.4 Empirical results: diffusion parameter and potential function

The diffusion parameter γ is estimated by equation (5.19). According to the dataset, the estimated value is $\widehat{\gamma} = 0.1374$. This is an estimate of the average volatility of daily price increments. More details about volatility estimation are given in Section 5.7.

Concerning the (scaled) potential function, an estimate is given in Figure 5.4 (a) where it is represented by the red line, whereas the blue line represents the corresponding values for the logarithmic inverse of the dataset histogram. In the present model, the typical shape of the potential's derivative is shown in Figure 5.4 (b). Recalling what we have said in Section 3.5, the potential's derivative is lower in the regions of "equilibrium" price level (attraction regions), whereas it is higher further away from these regions. Note that the usual mean-reversion corresponds to approximating the graph in Figure 5.4 (b) by a straight line. Indeed, in the mean-reverting models the influence of the drift term is the same across all price regions. In such a perspective, the potential model is considered as an extension of the mean-reverting class of models, because it allows for a "continuum" of regimes and a "continuum" of reversion rates, rather than a finite number of regimes, each with its own mean-reversion rate. Then, it is more versatile, as it allows modelling of multiple stable price equilibrium levels (attracting regions)



(a) (b)

Figure 5.4: An estimate of (a) the (scaled) potential function fitted with a mixture model of $K = 5$ Gaussian components and (b) the corresponding derivative of the potential (reversion rates).

and encompasses a mean-reverting models as special cases. Indeed, the potential model allows for a state-dependent reversion rate, so that it allows for a much richer model structure.

An underlying assumption of the present model is that the potential function and the long-term volatility do not change with time. As we have described in Chapter 2, global commodity markets have experienced significant price swings in recent years. In such a context of new market conditions, new attraction regions can form, changing the shape of the potential and the magnitude of the long-term volatility. In order to investigate further the behavior of the potential model, we have considered a price dataset to ensure the availability of as long a span of high quality data as possible. Section 5.10 reports more details about the functional form of the potential and changes in its shape, which reflects new price equilibrium levels (attraction regions) and hence new market conditions.

5.5 The estimated model

Now we can resume the procedure to estimate the model. Once we have provided an estimate of the scaled potential, $\widehat{G}$, and an estimate for the diffusion parameter, $\widehat{\gamma}$, then an estimate of the model (5.10) is given by

$$\widehat{p}(t_{i+1}) = p(t_i) - \frac{\widehat{\gamma}}{2}\widehat{G}'(p(t_i)) \qquad i = 0, 1, 2, \ldots \tag{5.23}$$

which can now be used for all possible applications that require to forecast the next price value, to predict the direction of the next move and to generate new independent price paths.

Concerning forecasting applications, the next day price value, $p(t_{i+1})$, can be evaluated from the actual observation, $p(t_i)$. More important, the model is able to take into account the evolution direction, in order to reduce uncertainty about the future behavior of the price move and thus allowing better predictions. The distributional characteristics of the prices are captured remarkably well from the resulting estimated model. By testing the model on our sub-sample dataset, we get the sample statistics

	observed	forecasted
Mean	18.6443	18.6406
Std.Dev.	3.2650	3.2399
Skewness	0.1218	0.1222
Kurtosis	2.9554	2.9574
Min	10.8200	10.9160
Max	28.0300	27.8682

Table 5.2: Descriptive statistics concerning the observed prices and the forecasted prices using the fitted model. Dataset: WTI crude oil daily spot prices over the trading days for the sub-sample period 1993-1999.

reported in Table 5.2. Comparing distributional moments of the observed and the forecasted prices confirms that the model is capable to capture the structure of the empirical data and its main features. More details are given in Section 5.6, where we point out the fitting of the model and its predictive power.

Furthermore, starting from the estimated model (5.23), adding an initial condition, $P(t_0) = p_0$, and an appropriate diffusion term, $\epsilon_i \sim N(0, \widehat{\gamma})$, the resulting model

$$p(t_{i+1}) = p(t_i) - \frac{\widehat{\gamma}}{2}\widehat{G'}(p(t_i)) + \sqrt{\widehat{\gamma}}\, N(0,1) \tag{5.24}$$

is able to generate copies of the observed price series with the same distributional properties. Knowing that $\gamma = \kappa^2(t_{i+1} - t_i)$, it is important to take into account the size of the time interval $h = t_{i+1} - t_i$ we consider. Equation (5.24) is useful for applications such as Monte Carlo studies, scenario testing, and other studies that require a large number of independent price trajectories. About numerical schemes and simulations of price process, we provide more details in Section 5.8.

5.6 Testing the model

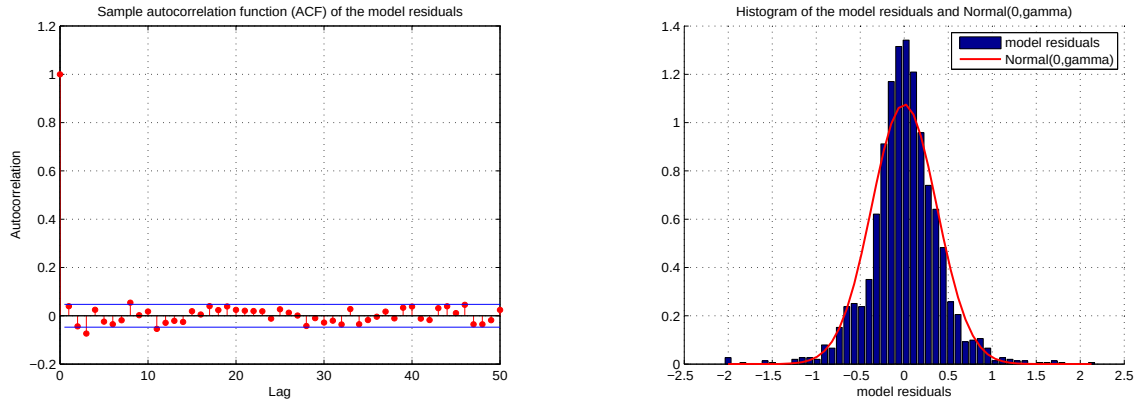
In order to analyze the model fit and investigate its performance, our main interest in this section is to provide some diagnostic analysis for the model we have estimated. As a first step, we consider an appropriate analysis of the model residuals. Then, for testing the predictive power of the model, we provide the mean square prediction error (MSPE). More important, we also consider the ability of the potential model in predicting the direction of the next period price move.

Once an estimate of the regression parameter, $\widehat{\gamma}$, has been provided, then we can consider the least squares regression line, which is given by

$$\widehat{Y} = \widehat{\gamma} X \tag{5.25}$$

where X denotes the predictor variable and Y represents the corresponding response variable. Using the substitutions (5.12) and (5.13), we can compute

$$\widehat{p}(t_{i+1}) = p(t_i) - \frac{\widehat{\gamma}}{2}\widehat{G'}(p(t_i)) \qquad i = 1, \ldots, n-1 \tag{5.26}$$



(a) (b)

Figure 5.5: Analysis of model residuals: (a) the autocorrelation function; (b) the histogram with a superimposed normal density of mean 0 and variance 0.1374. The sample statistics are reported in Table 5.3.

	Residual $\widehat{\epsilon}$	Error ϵ
Mean	0.0037	0.0000
Variance	0.1544	0.1374
Std. Dev.	0.3929	0.3706
Skewness	-0.1234	0.0000
Kurtosis	6.4800	3.0000
Exc Kurt	3.4800	0.0000

Table 5.3: Parameters for the error model, ϵ, and sample statistics for the corresponding model residuals, $\widehat{\epsilon}$. Figure 5.5 shows the autocorrelation function and histogram.

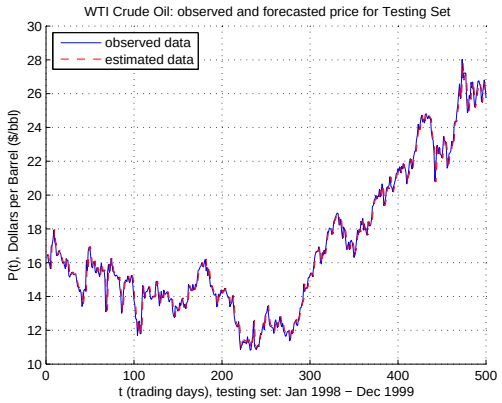
that is the fitted values, $\widehat{p}(t_{i+1})$, with respect to each previous i-th price observation, $p(t_i)$. In other words, since at the end of each time stepsize we know the current price value, $p(t_i)$, the next observation (the next fitted price value), $\widehat{p}(t_{i+1})$, is estimated as in expression (5.26), so it can be predicted from the previous one.

5.6.1 Diagnostics for the model fit

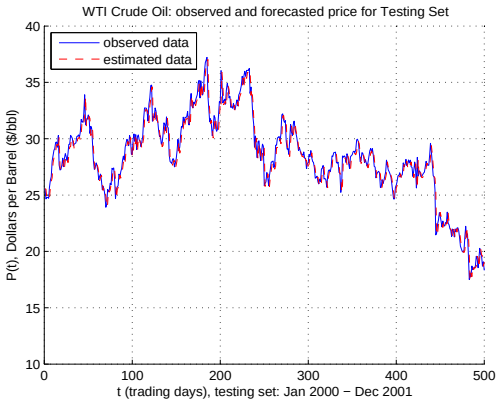
Using the fitted model (5.26) it is possible to compute the model residuals, as follows

$$\widehat{\epsilon}(t_i) = p(t_i) - \widehat{p}(t_i) = p(t_i) - \left(p(t_{i-1}) - \frac{\widehat{\gamma}}{2}\widehat{G'}(p(t_{i-1}))\right) \qquad i = 1, \ldots, n \tag{5.27}$$

which can be analyzed further. In order to test the model fit, we examine the residuals and perform standard diagnostics on them. If the model captures main dependence characteristics of the observed data, the residuals should be uncorrelated and approximately normally distributed. Figure 5.5 (a)



(a)



(b)

Figure 5.6: Testing the predictive power of the model for different testing set of $m = 500$ observations: (a) January 1998 - December 1999 with $MSPE = 0.2029$; (b) January 2000 - December 2001 with $MSPE = 0.6563$.

shows the sample autocorrelation function of the model residuals, which confirms that the residuals are uncorrelated. Figure 5.5 (b) shows the histogram of the model residuals with a superimposed normal density of mean 0 and variance 0.1374. A summary of sample statistics for the model residuals, $\widehat{\epsilon}$, and the error model, ϵ, are reported in Table 5.3. The model residuals have mean 0.0037 and variance 0.1544. Empirical results suggest little deviation from normality. Indeed, the histogram has a slightly more peaked shape than one would expect from a normal distribution. The variance of the residuals, 0.1544, is higher than the value estimated by the least squares, 0.1374, although the difference is not very large, especially if we consider the corresponding standard deviations. The variance of the residuals not always is in agreement with the diffusion parameter, then a further study of the residuals is needed, in order to improve the model fit and to capture the changes in volatility within the variance of the model residuals. In conclusion, we are able to state that the present analysis confirms a satisfactory model fit.

5.6.2 Testing the predictive power of the model

In this section we are interested to test the predictive power of the model. In order to measure the performance of the fitted model and assess its prediction accuracy, we use the mean squared prediction error ($MSPE$) defined as

$$MSPE = \mathbb{E}\Big(\sum_{i=1}^{m}\big(g(p_i) - \widehat{g}(p_i)\big)^2\Big) \tag{5.28}$$

which measures the expected squared distance between the predictor, $\widehat{g}$, and the true value, g, where the expectation is with respect to the true population. To estimate the error given by expression (5.28), we need to test our model using an independent set of data, say m independent observations. Then,

we estimate the above error measure using

$$\widehat{MSPE} = \frac{1}{m}\sum_{i=1}^{m}\big(p(t_i) - \widehat{p}(t_i)\big)^2 \tag{5.29}$$

which measures the average squared error between the predicted response, $\widehat{p}(t_i)$, obtained from the fitted model and the true measured response, $p(t_i)$.

To make the best use of the available data for evaluating the error in the model, the fundamental idea underlying this error analysis is a cross-validation procedure. How to determine the error depends on the application and the goal of analysis. In our case, we are interested in assessing the prediction accuracy of the fitted model. It is important to note that the response of the model depends on the estimated structure of the potential function, and in turn on the estimated invariant distribution. Therefore, in choosing the test sample we have to take into account a set of observations that reflects such a structure. Then, we partition the data into two folds. The first one is the training set and it is used to obtain an estimate for the model, so we choose the original dataset of the observed price series, $\{p(t_i)\}_{i=1}^{n}$, which yields $n = 1795$ observations. For the second one, we consider the last $m = 500$ observed values, and use it as testing set for estimating the prediction error. Note that the prediction, $\widehat{p}(t_i)$, comes from the model obtained by means of the fitted model (5.26), using the training set. Then, to asses the prediction accuracy, we calculate the squared error, $(p(t_i) - \widehat{p}(t_i))^2$, using the testing set as an independent test sample, where the observed value, $p(t_i)$, is the true measured response.

The mean squared prediction error computed on the basis of the testing set is 0.2029, and the difference with the variance of the residual remains acceptable (see Figure 5.6). If we consider a different testing set, i.e. the first $m = 500$ observed values for the subsequent period 2000-2006, the mean squared prediction error is 0.6563.

Obviously, different choices for training and testing set gives different prediction errors. More in general, differences in the error magnitude clearly show that the structure of the model is dependent on the structure of the estimated potential, which in turn is related to the estimated invariant distribution. Therefore, it is important to note that the estimated model is dependent on the accuracy in estimating the invariant distribution. For example, we pointed out the importance of the number of components in the mixture model and the more desired accuracy concerning the numerical algorithm implemented for the estimation of mixture parameters.

We have just seen that the resulting model is capable of reducing uncertainty about the future behavior of the price process and thus allowing better predictions. However, although the performance of the model is satisfactory in terms of the mean squared prediction error, better performances are also observed in predicting the direction of the next price move. It's worth pointing out that such a feature is actually more useful than predicting the actual value (i.e. for trading applications). The percentage of the correct up-down moves predicted by the potential model is 53.65%.

What has been said above for the error magnitude (i.e. the structure of the potential and the invariant distribution) is also valid for the predicting performance. Moreover, this performance can be improved if we only take into account up-down move forecasts when prices are away from the local minima of the potential. Indeed, we pointed out that at the local minima the derivative of the potential is nearly zero, thus having no effect on the price evolution. On the contrary, when the price is even slightly away from the minima, the influence of the potential prevails over the influence of the random fluctuations, thus making the evolution more predictable. As a consequence, we can consider as significant only the forecasts which have the absolute value of the potential's derivative larger than some parameter λ. Such a prediction scheme is capable of more correctly predicting performances in up-down moves.

5.7 The volatility in the potential function model

The volatility is an important parameter used for derivatives pricing, risk management, portfolio management and other areas of applications where volatility estimation and forecasting are required. This parameter is characterized by the fact that often it is the only one not directly observable in the market. For example, it is well known that in derivatives pricing the only ingredient of the Black and Scholes formula which is not observable is the parameter σ. In fact, the value of volatility used to price an option should be a forecast of the volatility over the entire period from the present time until the option's expiry.

The potential function model is able to provide an estimate of the volatility parameter. In the above sections we have outlined the procedure to estimate the (scaled) potential function and the diffusion parameter. Once we have got an estimate of these elements, we can now construct an estimate of the volatility parameter. According to the estimated model (5.23), the daily estimate of the volatility is provided as follows

$$\widehat{\sigma}^2(t_i) = \left(p(t_i) - p(t_{i-1}) + \frac{\widehat{\gamma}}{2}\widehat{G}'(p(t_{i-1}))\right)^2 \tag{5.30}$$

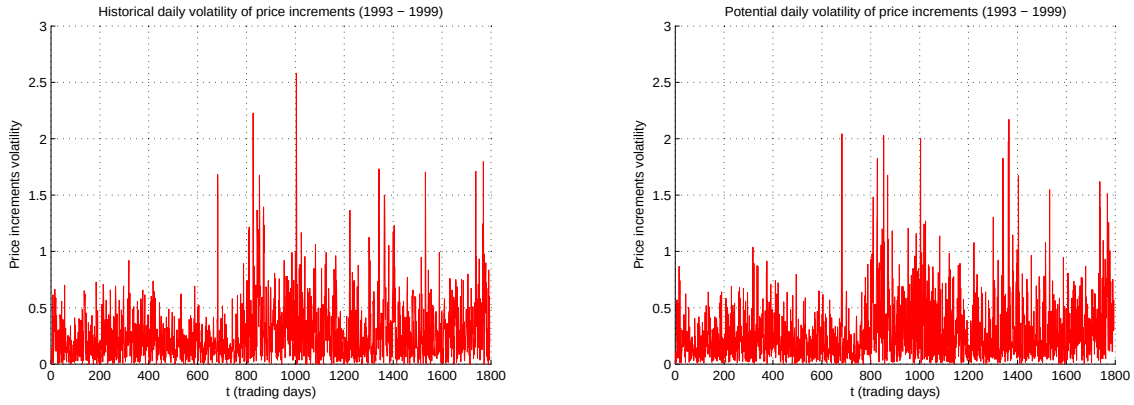
where $\widehat{\gamma}$ and $\widehat{G}$ are the estimates for the diffusion parameter and the scaled potential, $p(t_{i-1})$ is the current price and $p(t_i)$ is the daily price on subsequent time t_i. It is clear that the estimate (5.30) is a measure of the variability of daily price increment $\triangle p = p(t_i) - p(t_{i-1})$, whereas it is usually to consider the variability of the daily price's return.

In this way, the volatility estimated from the potential function model can then be used as a valid alternative to the historical volatility or others such as GARCH volatility models.

In the volatility measure (5.30) the presence of the potential's derivative play the role of a deterministic component and this represents an important feature which characterizes the structure of such a variability measure. For example, this is what distinguishes such a measure from the historical volatility, where it is replaced by a simple mean. Figure 5.7 shows the series of the realized volatility of daily price increments and the series of potential function volatility estimated by equation (5.30).

In our analysis of potential model, we have already pointed out that the evolution of the process is governed by the potential drift. According to equation model (5.1), the magnitude of the price increments and, hence, the spread of their distribution is in some part determined by the derivative of the potential and the remaining part by random fluctuations. This share-out is obviously dependent on how much feasible and accurate the structure of the potential function we are able to construct and estimate. In the context of variability measure, in an analogous manner the variability of price increments is partly explained in term of deterministic component, while another part is due to stochastic fluctuations. Since the derivative of the potential is evaluated with respect to the current price $p(t_{i-1})$, when the current price is far from one of the attracting points $p^\star$ (i.e. the equilibrium price levels), then the derivative of the potential is higher in absolute value, so the deterministic component plays its more significant role. On the other hand, the more the current price is near the attraction points (at equilibrium price), the more the deterministic component (via potential's derivative) is close to zero and the price movements are predominantly determined by random fluctuations. This equilibrium view, namely the convergence to an equilibrium price $p^\star$, is in good agreement with economic arguments and reflects the perceptions of the market evolution.

The importance to have an alternative method for estimating the volatility parameter is also linked to the availability and reliability of data. For example, in derivatives pricing to determine estimates of volatility based on implied volatility is a methodology often used. Given a finite set of maturities, the market option prices are observed, so the volatility is calculated by inverting the formula for the option price. This approach is feasible when the corresponding market for the asset or claim we are interested



(a) (b)

Figure 5.7: An estimate of (a) historical daily volatility of price increments and (b) the potential function daily volatility.

in pricing is developed and liquid enough, i.e. stock option markets. Especially for commodities options, such a markets are not as developed and as liquid as stock option markets, then we are facing to commodities options the implied volatility approach could be not always feasible. Also in the case of Over-The-Counter option markets, which constitute a large part of traded commodities options, the observed option prices are often either not available or not reliable.

5.8 Numerical schemes and simulations of price process

In this section we are interested to provide a numerical implementation of the model (5.1), which in its differential equation form is

$$dP(t) = -U'(P(t))\ dt + \kappa\ dB(t) \qquad t_0 \leq t \leq T \tag{5.31}$$

that is a scalar, autonomous stochastic differential equation. The dynamics of the price process $(P(t))_{t\in\mathbb{T}}$ over the time interval $\mathbb{T} = [t_0, T]$, expressed in integral equation form is

$$P(t) = P(t_0) + \int_{t_0}^{t} -U'(P(s))\ ds + \int_{t_0}^{t} \kappa\ dB(s) \qquad t_0 \leq t \leq T \tag{5.32}$$

where the initial condition $P(t_0)$ can be a random variable or a constant. We now need to allow us to solve the above stochastic differential equation numerically, so our goal is to apply a numerical method to equation (5.31) over the time interval $[t_0, T]$.

For computational purposes it is useful first to adopt a discretization of the time interval. We thus set a discrete number of t values as follows

$$\triangle t = \frac{T - t_0}{N} \qquad \triangle t = t_{k+1} - t_k \qquad t_k = t_0 + k\triangle t \qquad k = 0, 1, \ldots, N$$

where the positive integer N denotes the number of equally sized subintervals, whereas $\triangle t$ denotes the constant size of every k-th subintervals $[t_k, t_{k+1}]$. In such a way we get $N + 1$ time points, $\{t_k\}_{k=1}^{N}$, with respect to we will develop the numerical approximation.

The stochastic source in integral equation (5.32) is given by the stochastic process $(B(t))_{t\in\mathbb{T}}$ that we assume is a Brownian motion. We need to construct a discretized Brownian motion, where $B(t)$ is specified at discrete t values. Assuming the above time discretization, we define the stochastic source as

$$B(0) = 0 \tag{5.33}$$
$$B(t_k) = B(t_{k-1}) + dB(t_k) \qquad k = 1, 2, \ldots, N \tag{5.34}$$

meaning that we will generate a sequence of N random numbers that determine $B(t)$ at a discrete number of t values. In order to generate such a sequence we need to generate the increments $dB(t_k)$. From properties of Brownian motion, we can use the following relations

$$B(t_k) - B(t_{k-1}) = dB(t_k) \sim N(0, t_k - t_{k-1}) \tag{5.35}$$
$$= dB(t_k) \sim \sqrt{t_k - t_{k-1}}\, N(0, 1) \tag{5.36}$$

where each increment $dB(t_k)$ is a Normal random variable with mean $\mu = 0$ and variance $\sigma^2 = t_k - t_{k-1}$, and the standard deviation $\sigma = \sqrt{t_k - t_{k-1}}$ stands for a diffusion coefficient with respect to the standard Normal, $N(0, 1)$. We also use the independence of the increments of a Brownian motion.

5.8.1 The Euler-Maruyama method

The Euler-Maruyama method is constructed within the Itö integral framework. We will attempt to construct a numerical method from the integral form (5.32). We will compute our own discretized Brownian motion paths and use them to generate the needed increments $dB(t)$.

For convenience, we first set a new discretization of time interval as follows

$$\delta t = \frac{T - t_0}{L} \qquad \delta t = t_{j+1} - t_j \qquad t_j = t_0 + j\delta t \qquad j = 0, 1, \ldots, L$$

for some positive integer L. In such a way, we use a stepsize Dt for the numerical method to be an integer multiple $R \geq 1$ of the increment $\triangle t$ used for the Brownian path, $Dt = R \cdot \triangle t$. This ensures that the set of points $\{t_k\}_{k=0}^N$ on which the discretized Brownian path is based contains the points $\{\tau_k\}_{k=0}^N$ at which the numerical solution is computed.

Starting from integral form (5.32), we begin to define

$$P(t) \approx P(\tau_j) = P_j \qquad j = 0, 1, \ldots, L \tag{5.37}$$

where numerical approximation to $P(\tau_j)$ will be denoted as P_j. Now, setting successively the endpoints $t = \tau_{j-1}$ and $t = \tau_j$, we obtain

$$P(\tau_{j-1}) = P(\tau_0) + \int_{\tau_0}^{\tau_{j-1}} -U'(P(s))\, ds + \int_{\tau_0}^{\tau_{j-1}} \kappa\, dB(s)$$
$$P(\tau_j) = P(\tau_0) + \int_{\tau_0}^{\tau_j} -U'(P(s))\, ds + \int_{\tau_0}^{\tau_j} \kappa\, dB(s)$$

and subtracting the first equation from the second one, then we get

$$P(\tau_j) = P(\tau_{j-1}) + \int_{\tau_{j-1}}^{\tau_j} -U'(P(s))\, ds + \int_{\tau_{j-1}}^{\tau_j} \kappa\, dB(s) \tag{5.38}$$

so that we can now consider approximating each of the integral terms. Considering each subinterval $[\tau_{j-1}, \tau_j]$, the approximation of the corresponding term on the right-hand side of (5.38) is

$$\int_{t_{k-1}}^{t_k} -U'(P(s))\ ds \approx -U'(P(t_{k-1}))\ (t_k - t_{k-1}) \tag{5.39}$$

$$\int_{t_{k-1}}^{t_k} \kappa\ dB(s) \approx \kappa\ (B(t_k) - B(t_{k-1})) \tag{5.40}$$

where for the first integral we can use the conventional deterministic quadrature, whereas for the second integral we use the Itö formula. Combining these approximation together, we get the equation

$$\begin{aligned} P_0 &= p_0 \\ P_j &= P_{j-1} - U'(P_{j-1})\ (\tau_j - \tau_{j-1}) + \kappa\ (B_j - B_{j-1}) \end{aligned} \tag{5.41}$$

that is the formula for the Euler-Maruyama method.

5.8.2 The Milstein method

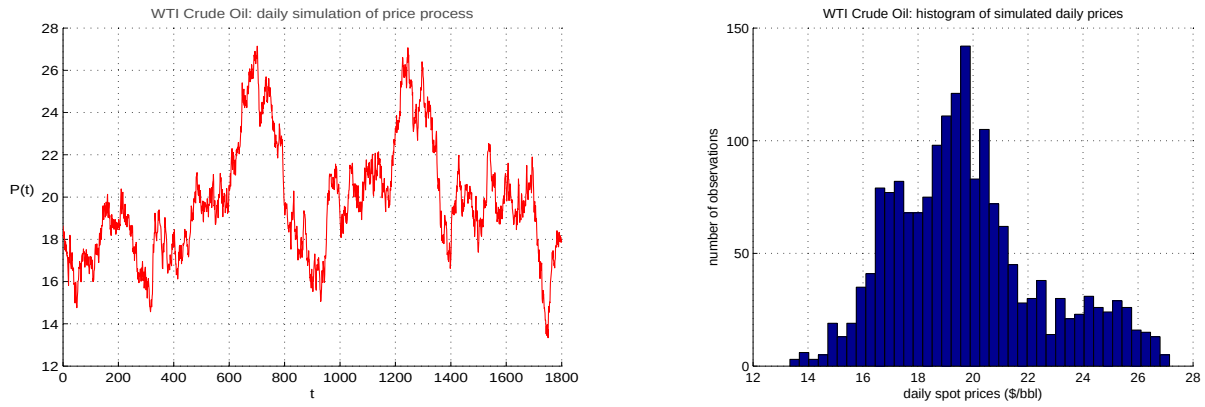
A stochastic differential equation is said to have *additive noise* if all entries of the diffusion coefficient matrix are either constant or functions of time only. If the diffusion coefficient contains entries that are functions of the state variables, then the stochastic differential equation is said to have *multiplicative* noise. The Euler-Maruyama scheme has strong convergence of order $\eta = 0.5$, but this increases to $\eta = 1.0$ if only additive noises are present. It becomes clear why this is so if we consider the higher order scheme, the Milstein scheme. The Milstein scheme can be shown to have order of strong convergence $\eta = 1.0$. If the noise is additive the last term in the Milstein scheme vanishes, thus reducing to the Euler-Maruyama scheme. Therefore, for additive noise the Euler-Maruyama scheme has order of strong convergence $\eta = 1.0$. It is important to note that if only additive noises are present, the order, and hence efficiency, of a numerical scheme increases. The potential model has additive noise, so our numerical scheme refers to the higher order convergence.

5.8.3 Simulation of price trajectories

Suppose that the continuous time stochastic process $(P(t))_{t\in\mathbb{T}}$ is a price process over the time interval $\mathbb{T} = [t_0, T]$, with $P(t) \in \mathbb{R}$. Now, consider we have a dataset, $p = \{p(t_i)\}_{i=1}^N$, of an observed price series which is a realization of the price process at discrete time points, $\{t_i\}_{i=1}^N$. We postulate that the dynamic of the process evolves according to the stochastic differential equation given by the equation model (5.1). In the framework of Euler-Maruyama method, having estimated the scaled potential $\widehat{G}$ and the parameter $\widehat{\gamma}$, we are able to give a numerical solution to the diffusion equation (5.1) by means of the following approximation

$$\begin{aligned} P(t_i) &= P(t_{i-1}) - \widehat{U}'(P(t_{i-1}))\ (t_i - t_{i-1}) + \widehat{\kappa}\ (B(t_i) - B(t_{i-1})) \\ &= P(t_{i-1}) - \frac{\widehat{\kappa}^2}{2}\widehat{G}'(P(t_{i-1}))\ (t_i - t_{i-1}) + \widehat{\kappa}\ (B(t_i) - B(t_{i-1})) \\ &= P(t_{i-1}) - \frac{\widehat{\gamma}}{2}\widehat{G}'(P(t_{i-1})) + N(0, \widehat{\gamma}) \end{aligned} \tag{5.42}$$

for some initial condition $P(t_0) = p_0$, with the discretization stepsize $h = t_i - t_{i-1}$, and $\gamma = \kappa^2(t_i - t_{i-1})$, where we have used the relation (5.36), so each Brownian increment is an independent Normal random variable with mean $\mu = 0$ and variance $\sigma^2 = \gamma$, whereas the standard deviation is $\sigma = \kappa\sqrt{t_i - t_{i-1}}$.



(a) (b)

Figure 5.8: Price path of length $N = 1800$ generated by the fitted model: (a) one simulated price trajectory and (b) the corresponding histogram.

Since in the present context we suppose to analyse daily price movements, we take the discretization step $h = 1$, corresponding to the price movement observed in one trading day. The above scheme becomes

$$\begin{aligned} P(t_i) &= P(t_{i-1}) - \frac{\widehat{\gamma}}{2}\widehat{G}'(P(t_{i-1})) + N(0,\widehat{\gamma}) \\ &= P(t_{i-1}) - \frac{\widehat{\kappa}^2}{2}\widehat{G}'(P(t_{i-1})) + \widehat{\kappa}\ N(0,1) \end{aligned} \tag{5.43}$$

where each Brownian increment becomes an independent Normal random variable with mean $\mu = 0$ and variance $\sigma^2 = t_i - t_{i-1} = 1$. Furthermore, in Section 5.3 we have pointed out that it is in general impossible to separate the effects of time discretization and the random disturbances on fluctuations of time series, thus we combined these effects into a new parameter, $\gamma = \kappa^2 h$. When we choose the discretization stepsize as $h = 1$, we get the estimation of the original parameter, $\widehat{\gamma} = \widehat{\kappa}^2$.

Starting from the fitted model, one simulated price trajectory of length $N = 1800$, and the corresponding histogram are shown in Figure 5.8. Therefore, the model is able to generate copies of the observed price series with the same invariant distribution, which is useful for applications such as Monte Carlo analysis, scenario testing, and other studies that require a large number of independent price trajectories.

5.9 A goodness-of-fit test for the SDE model

In this section, our interest is to access the validity of the model obtained by testing if there is a lack-of-fit between the stochastic differential equation model and the collection of data used to estimate the drift and diffusion. We adopt the approach of a goodness-of-fit procedure developed in Bak (1998 [6]) (see also Allen 2007 [2] and Bak et al. 2000 [7]). The approach considered in such a method is based on Monte Carlo simulation of trajectories between neighbour observations points and, thus, it does not rely on the availability of explicit expressions of the conditional densities.

Suppose that the price process $(P(t))_{t\in\mathbb{T}}$ is observed at discrete time points, $\{t_i\}_{i=1}^{N-1}$, where $t_i = i\triangle t$ for a constant stepsize $\triangle t = [t_{i-1}, t_i]$. The dataset $\{p(t_i)\}_{i=1}^{N-1}$ consists of the observations of the process. We assume that an estimate of the drift, $\widehat{\mu}(P(t)) = \widehat{G}(P(t))$, and the diffusion, $\widehat{g}(P(t)) = \widehat{\gamma}$, has been obtained. For each i-th time interval $[t_{i-1}, t_i]$, the procedure consider that simulations are performed to obtain M trajectories from time t_{i-1} until time t_i, starting at the observed value $p(t_{i-1})$. For example, consider that each i-th time interval $[t_{i-1}, t_i]$ is divided into K sub-intervals of the form

$$Dt = \frac{t_i - t_{i-1}}{K} \qquad Dt = t_{k+1,i} - t_{k,i} \qquad \text{con} \quad t_{k,i} = t_{0,i} + kDt \qquad k = 0, 1, \ldots, K$$

where for $k = 0$ we obtain the start time point $t_{0,i} = t_{i-1}$ and for $k = K$ the end time point $t_{K,i} = t_i$. Using the Euler-Maruyama method with K inter-steps, the update formula is

$$P_{k+1,i}^{(m)} = P_{k,i}^{(m)} - \frac{\widehat{\kappa^2}}{2}\widehat{G}'(P(t_{i-1}))\ Dt + \widehat{\kappa^2}\ \sqrt{Dt}\ \xi_{k,i}^{(m)} \tag{5.44}$$

for $k = 0, 1, \ldots, K-1$ and $m = 1, 2, \ldots, M$, where $\xi_{k,i}^{(m)} \sim N(0,1)$ for each i, k and m. In the update formula (5.44), for $m = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N-1$, the following values

$$k = 0 \qquad P_{0,i}^{(m)} = p_{i-1} \tag{5.45}$$

$$k = K \qquad P_{K,i}^{(m)} = P_i^{(m)} \tag{5.46}$$

are the starting observed value p_{i-1} and the m-th simulated value $P_i^{(m)}$ at endpoint interval time t_i, respectively.

In order to perform the test, we need to organize the sample dataset in a certain number of classes. We do this by means of the rank. To calculate the rank r_i of the observed value p_i as compared with the M simulated values $P_i^{(m)}$, now define the following indicator function

$$s_i^{(m)} = \begin{cases} 0 & \text{if} \quad P_i^{(m)} > p_i \\ 1 & \text{if} \quad P_i^{(m)} \leq p_i \end{cases} \tag{5.47}$$

and use this function to define

$$r_i = 1 + \sum_{m=1}^{M} s_i^{(m)} \qquad i = 1, 2, \ldots, N-1 \tag{5.48}$$

where the integer number r_i is the rank of the observed value p_i as compared to the endpoints values of the M simulated trajectories. Notice that

$$1 \leq r_i \leq M + 1 \qquad i = 1, 2, \ldots, N-1 \tag{5.49}$$

where $r_i = 1$ if the M simulated endpoint values are all above the observed value p_i, whereas $r_i = M+1$ if they are all below.

The present method use a χ^2 goodness-of-fit test to test the validity of the model (i.e. a distribution assumed for the random phenomenon). The test evaluates the null hypotheses $\mathcal{H}_0$ that model (5.1) describes the price process (i.e. the data are governed by the assumed distribution), against the alternative $\mathcal{H}_1$ that the process cannot be described by (5.1) (i.e. the data are not drawn from the assumed distribution). In general, the chi-square test statistic is of the form

$$\mathrm{X}^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \tag{5.50}$$

where the assumed model (expected or theoretical) is evaluated against the observed data. If the computed test statistic is large, then the observed and expected values are not close and the model is a poor fit to the data. To perform this test, the observed and expected frequencies are needed.

Concerning the *expected* frequencies, consider that R_i denote the random variable corresponding to the observation r_i. Under the null hypothesis $\mathcal{H}_0$, the probability that the rank r_i get the value q is

$$P_{\mathcal{H}_0}(R_i = q) = p_q = \frac{1}{M+1} \qquad q = 1, \dots, M+1 \tag{5.51}$$

that is the ranks r_i have equally likely values between 1 and $M+1$, so the distribution of each i-th random variable R_i is uniformly distributed. If the null hypothesis $\mathcal{H}_0$ is true, the expected frequency is

$$(N-1)p_q = \frac{N-1}{M+1} \tag{5.52}$$

that is the number of rank r_i that get value q.

In order to define the *observed* frequencies, at each i-th step we implement the following indicator function

$$\mathbf{1}_{i,q} = \begin{cases} 0 & \text{if} \quad r_i \neq q \\ 1 & \text{if} \quad r_i = q \end{cases} \qquad i = 1, 2, \dots, N-1 \tag{5.53}$$

and use it function to define

$$\Omega(q) = \sum_{i=1}^{N-1} \mathbf{1}_{i,q} \qquad q = 1, 2, \dots, M+1 \tag{5.54}$$

where $\Omega(q)$ is the observed frequency that the rank equals the value q. It is clear that, for every $i = 1, 2, \dots, N-1$, we assign only one rank, so that is $\sum_{q=1}^{M+1} \Omega(q) = N-1$.

In the spirit of the Pearson chi-square test the statistic for the hypothesis that $p_q = 1/(M+1)$ is therefore

$$\mathrm{X}^2_{(M)} = \sum_{q=1}^{M+1} \frac{(\Omega(q) - \frac{N-1}{M+1})^2}{\frac{N-1}{M+1}} \sim \chi^2_M \tag{5.55}$$

which, under the null hypothesis $\mathcal{H}_0$, is approximately distributed as a chi-square random variable with M degrees of freedom (Kendall and Stuart 1961 [56]). Namely, the degrees of freedom are the number of classes $M-1$ minus 1, so $(M+1)-1 = M$. A large value of calculated statistics X^2 indicates a lack-of-fit between the stochastic model and the data. Given that $\alpha \in (0,1)$ is the desired level of significance, the decision rule is

$$\begin{array}{lll} \mathrm{X}^2_{(M)} \geq \chi^2_{(M)}(\alpha) & \rightarrow & \text{reject } \mathcal{H}_0 \\ \mathrm{X}^2_{(M)} < \chi^2_{(M)}(\alpha) & \rightarrow & \text{do not reject } \mathcal{H}_0 \end{array}$$

that is reject $\mathcal{H}_0$ if the calculated value $\mathrm{X}^2_{(M)}$ exceeds the upper α critical value of the $\chi^2_{(M)}(\alpha)$ distribution. Often, the null hypothesis involves fitting a model with parameters estimated from the observed data. By estimating a parameter, we lose a degree of freedom in the chi-square test statistic. In general, if we estimate d parameters under the null hypothesis, the number of degrees of freedom for the associated chi-square distribution is adjusted to $M-d$.

It is important to note that the chi-square approximation fails when the expected frequencies under the null hypothesis are small (Kendall and Stuart 1961 [56] p. 440, Piccolo 2010 [68] p. 705). As a consequence, many researchers often applied the rule-of-thumb that the expected frequencies should be no less than 5. Applying this rule gives

$$5 \leq \frac{N-1}{M+1} \quad \Longleftrightarrow \quad M \leq \frac{N-6}{5} \tag{5.56}$$

yielding an upper bound and, in practice, the number of simulated trajectories will be well below that bound. For example, if the dataset has $N = 500$ observations, then the rule requires that the number M of inter-observation simulated trajectories should be no more than 98.

5.10 New market conditions and changes in the potential function

An underlying assumption of the present model is that the potential function and the long-term volatility do not change with time. As we have described in Chapter 2, global commodity markets have experienced significant price swings in recent years. In such a context of new market conditions, new attraction regions can form, changing the shape of the potential and the magnitude of the long-term volatility. In order to investigate further the behavior of the potential model, we have considered a price dataset to ensure the availability of as long a span of high quality data as possible. At first we employ daily spot price data concerning crude oil. Additionally, we have also tested the behaviour of the present model with an agricultural commodity, using soybean data.

5.10.1 Crude oil

As a first step we consider the daily crude oil price. Significant changes affected the oil market in the last two decades. Moreover, it is a key input to the world's economy, and it is the most considered commodity in literature as well.

West Texas Intermediate (WTI) crude oil and North Sea Brent are two major oil types often considered in the literature. North Sea Brent is a high quality crude oil produced in the Norwegian/British North Sea, also used as global crude benchmark, and its contract are traded on the Intercontinental Exchange (ICE) in London. In this work, we have examined the WTI crude oil spot price, which is referenced as the the most widely followed benchmark of crude oil in the energy complex. It is quoted in the financial press and in analyst reports and, more importantly, it is used by the industry players as a benchmark for global oil prices. WTI is of very high quality, and is a light and sweet crude, lighter and sweeter than Brent Crude. WTI is excellent for refining a larger portion of gasoline. Currently, most heavy, sour crudes are priced relative to their lighter and sweeter counterparts. WTI crude oil is traded in the U.S. spot market at Cushing, Oklahoma and is also the underlying commodity of New York Mercantile Exchange's oil futures contracts (NIMEX). Concerning the dataset, the data period covered is more than 25 years, consisting in the daily closing prices from January 2, 1986 to August 13, 2012. The full sample yields 6813 observations. The WTI data consists of Spot Price FOB, and the prices are in US Dollars per Barrel ($/bbl). The data are taken from the Energy Information Administration (EIA), which is the statistical branch of the Department of Energy (DOE), and it details activity in the country's energy sector providing official energy statistics for the U.S. government (see url: http://www.eia.doe.gov/). The release date is August 15, 2012 (Source: Thomson Reuters).

Figure 5.9 shows the daily spot prices of WTI crude oil since 1986. In order to investigate the evolution in the structure of potential function over a long observation interval, we have considered as

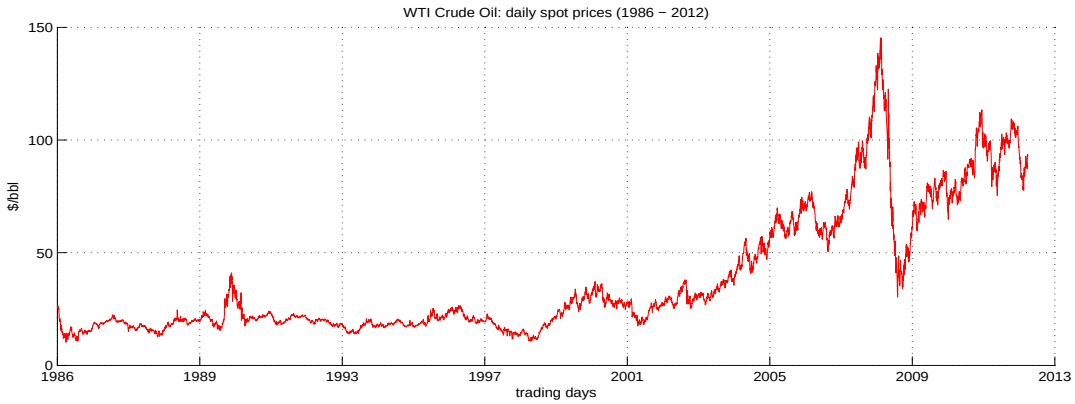


Figure 5.9: WTI crude oil daily spot prices over the trading days from January 2, 1986 to August 13, 2012. A summary of descriptive statistics is reported in Table 5.4 (upper section).

Period	86-12	86-92	93-99	00-06	07-12
Mean	37.8898	19.5122	18.6443	39.6800	83.2840
Std.Dev. (SD)	27.8366	4.2530	3.2650	15.3915	20.6677
Skewness (Sk)	1.3437	1.5377	0.1218	0.7847	0.1420
Kurtosis (Ku)	3.7769	7.8255	2.9554	2.2748	3.1562
Min	10.25	10.25	10.82	17.50	30.28
Max	145.31	41.07	28.03	77.05	145.31
Observations	6813	1794	1795	1808	1416
Mixture Mean	37.8898	19.5122	18.6443	39.6800	83.2840
Mixture SD	27.8345	4.2518	3.2641	15.3872	20.6604
Diffusion $\widehat{\gamma}$	0.1564	0.1300	0.1374	0.4525	4.5048
$V(\widehat{\epsilon})$	1.2536	0.4094	0.1544	0.8507	4.2128
Forecasted Mean	37.8799	19.5156	18.6406	39.6605	83.2625
Forecasted SD	27.8262	4.2402	3.2399	15.3723	20.5665
Forecasted Sk	1.3451	1.5509	0.1222	0.7895	0.1462
Forecasted Ku	3.7810	7.8751	2.9574	2.2773	3.1587
$\widehat{MSPE}(1)$	3.1472	0.1309	0.2029	1.4164	3.1308
$\widehat{MSPE}(2)$	-	0.0844	0.6563	6.4206	-
Up-Down	52.30%	55.24%	53.65%	53.98%	50.85%

Table 5.4: Descriptive statistics (upper section) and estimates of potential model features (lower section). Dataset: WTI crude oil spot price data concerning the full sample period 1986-2012 and the four sub-samples for the periods 1986-1992, 1993-1999, 2000-2006, 2007-2012.

sub-samples the following four periods of seven years each: 1986-1992, 1993-1999, 2000-2006, 2007-2012, where the last sub-sample is of about five years and a half. A summary of descriptive statistics for spot prices of full sample and related sub-samples are reported in Table 5.4 (upper section). The increasing mean price shows the changes in oil price starting from the year 1999, where the volatility has increased as well. In Table 5.4 (lower section) are also reported the main features of the potential model. An estimate of the (scaled) potential function fitted with a mixture model of $K = 5$ Gaussian components and the corresponding derivative of the potential (reversion rates) are reported in Figure 5.10. Together with the increasing mean price, the estimates of diffusion parameter, $\widehat{\gamma}$, show the increasing size of the volatility starting from the period 2000-2006, with the highest levels in the period 2007-2012. The variance of the residuals, $V(\widehat{\epsilon})$, not always is in agreement with the diffusion parameter, then a further study of the residuals is needed, in order to improve the model fit. Comparing distributional characteristics of the observed prices and the forecasted ones confirms that the behaviour of the price evolution is captured remarkably well. Concerning the functional form of the potential, changes in shape reflect new price equilibrium levels (attraction regions) and hence new market conditions. If the underlying price time series is daily, the model can be regularly refitted every few months (i.e. every 6-12 months), in order to capture the changes in the market within the potential function and the changes in volatility within the variance of the model residuals. In terms of the mean square prediction error, the resulting model is capable of reducing uncertainty about the future behavior of the price process and thus allowing better predictions. The estimates $\widehat{MSPE}(1)$ stand for a testing set which considers the last 500 observed values for the same period, whereas the estimates $\widehat{MSPE}(2)$ stand for a testing set which considers the first 500 observed values for the subsequent period. More in general we recall that differences in the error magnitude clearly show that the structure of the model is dependent on the structure of the estimated potential, which in turn is related to the estimated invariant distribution. Therefore, it is important to note that the estimated model is dependent on the accuracy in estimating the invariant distribution. We pointed out the importance of the number of components in the mixture model and the more desired accuracy concerning the numerical algorithm implemented for the estimation of mixture parameters. In particular, better performances are observed in predicting the direction of the next price move in terms of up-down moves.

Finally, the application of the potential model to the crude oil prices shows that the essential characteristics of the data are captured remarkably well. In particular, the model is able to take into account new attraction regions arising from new market conditions and changes in the variables (forces) acting on the market.

5.10.2 Soybean

In order to investigate the behaviour of the potential model, we have also tested its main characteristics with an agricultural commodity, using soybean data.

Soybean is one of the major agricultural commodities. The cultivation of soybeans, which developed in Asia, has been taking place for centuries. Soybeans are a vital crop for the world economy and have a growing number of uses and applications, ranging from food products to the creation of biofuel to agricultural feedstock. The world soybean production is about 260 million tonnes (2011-2012 forecasts), and the major producers are the United States, Brazil, Argentina, China, India and Paraguay. The United States dominates the soybean market, accounting for over 35 percent of total global production, together with the increasing production of Brazil, with about 25 percent of the market. The different soybean extracts traded on the market are soybeans themselves, soybean oil, and soybean meal. Concerning recent price development, spot and futures prices in the soy complex continue to be very high in historic terms. Marked production shortfalls in the 2011/12 season - first in

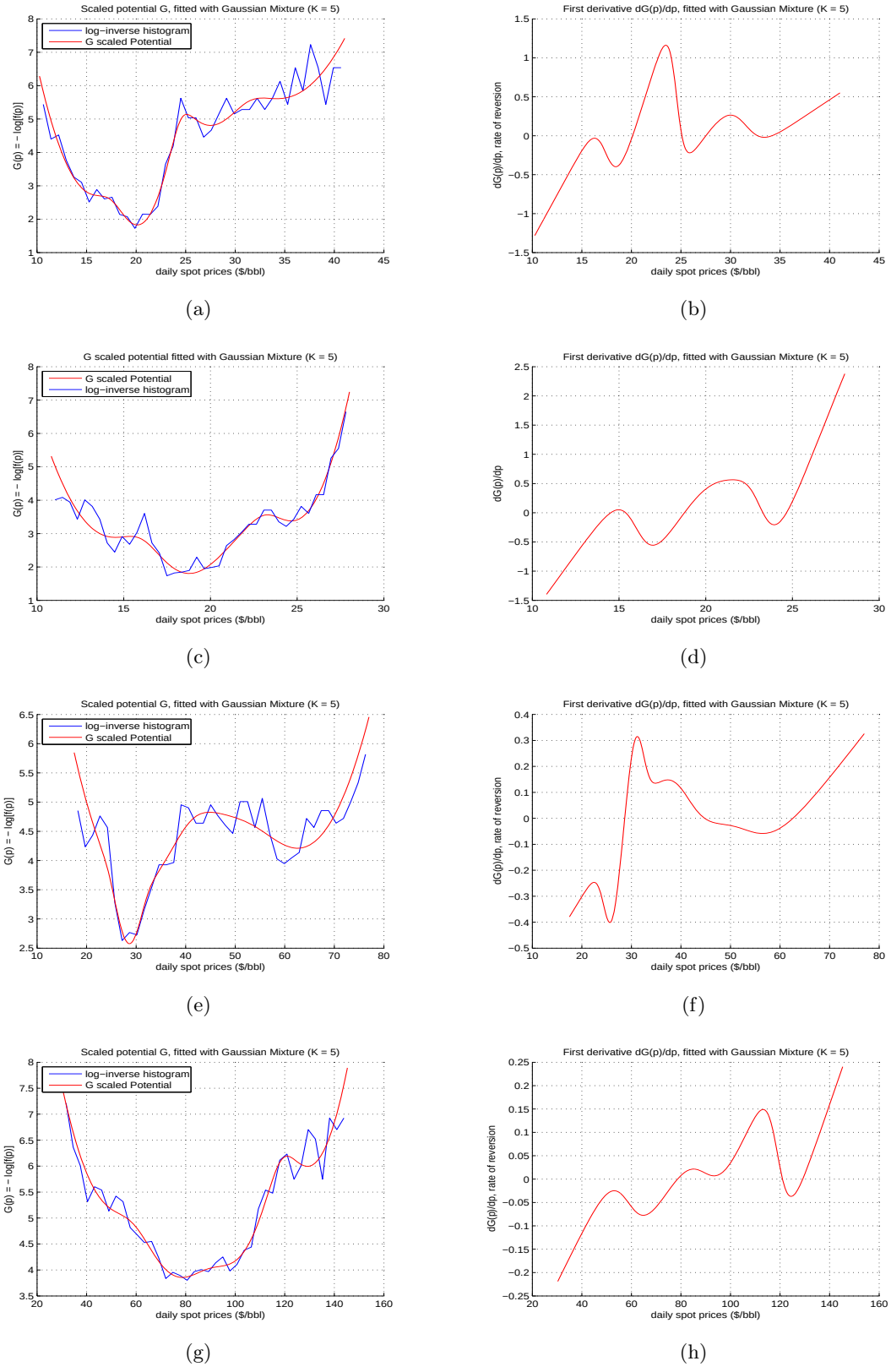


Figure 5.10: An estimate of the (scaled) potential function fitted with a mixture model of $K = 5$ Gaussian components and the corresponding derivative of the potential (reversion rates). Dataset: WTI crude oil spot price data concerning the four sub-samples for the periods (a)-(b) 1986-1992, (c)-(d) 1993-1999, (e)-(f) 2000-2006, (g)-(h) 2007-2012.

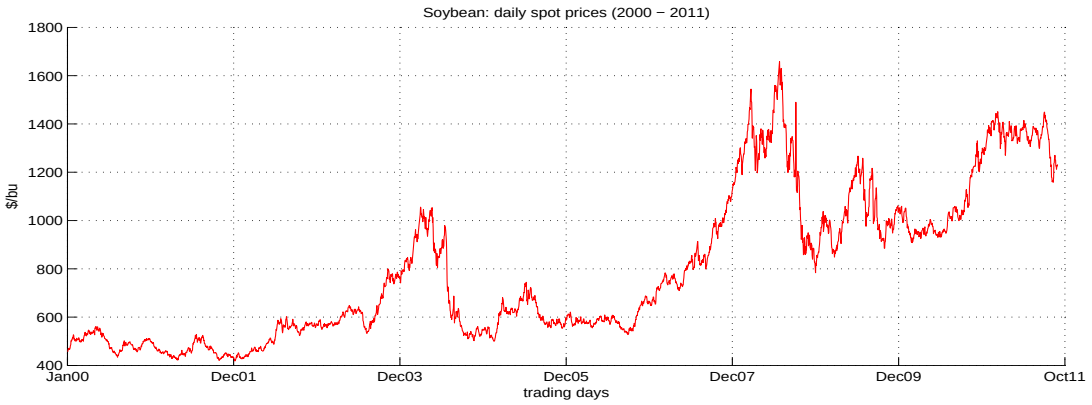


Figure 5.11: Soybean daily spot prices over the trading days from January 3, 2000 to October 26, 2011. A summary of descriptive statistics is reported in Table 5.5 (upper section).

Period	00-11	00-02	03-05	06-08	09-11
Mean	791.0049	490.7719	665.4679	896.6878	1130.9283
Std.Dev. (SD)	305.3175	45.0439	135.5040	304.4114	174.3742
Skewness (Sk)	0.7300	0.4839	1.2290	0.6478	0.3258
Kurtosis (Ku)	-0.6758	2.1269	3.5439	2.1843	1.5760
Min	418.00	418.00	499.50	527.25	848.50
Max	1658.00	602.00	1055.75	1658.00	1451.00
Observations	2978	755	756	756	711
Mixture Mean	791.0016	490.7719	665.4679	896.6878	1130.9283
Mixture SD	305.0314	45.0141	135.4143	304.2100	174.2515
Diffusion $\widehat{\gamma}$	147.4054	29.2658	86.4654	84.7257	211.3012
$V(\widehat{\epsilon})$	307.1695	38.6573	200.9154	611.1364	379.9789
Forecasted Mean	790.7441	490.6219	665.4472	896.2190	1130.5603
Forecasted SD	305.0314	44.6498	135.1976	304.4306	173.8405
Forecasted Sk	0.7335	0.4988	1.2376	0.6512	0.3338
Forecasted Ku	-0.6749	2.1136	3.5496	2.1833	1.5585
$\widehat{MSPE}(1)$	275.5529	39.2358	256.4807	897.6573	274.9810
$\widehat{MSPE}(2)$	-	261.2142	106.5576	401.1555	-
Up-Down	52.79%	52.85%	51.06%	50.93%	58.37%

Table 5.5: Descriptive statistics (upper section) and estimates of potential model features (lower section). Dataset: Soybean spot price data concerning the full sample period 2000-2011 and the four sub-samples for the periods 2000-2002, 2003-2005, 2006-2008, 2009-2011.

the United States and then, even more, in South America - resulted in a pronounced tightness of global soybean supplies, while import demand continued to grow unabatedly, notably in China. In the recent weeks prices have risen above the peak levels recorded during the crisis in 2007-2008, thus marking new historic records. On this occasion some of the structural and cyclical factors that triggered the previous food crisis are not in evidence. Indeed, it is weather conditions that are pushing up prices, these corresponding to good and bad harvest years whose occurrence is not periodic.

Concerning the dataset, the data period covered is about 12 years, consisting in the daily closing prices from January 3, 2000 to October 26, 2011. The full sample yields 2978 observations. The soybean data refers to the Chicago Board of Trade, and the prices are in US Dollars per Bushel ($/bu). The full sample period 2000-2011 has been divided into four sub-samples corresponding to the periods 2000-2002, 2003-2005, 2006-2008, 2009-2011.

Figure 5.11 shows the daily spot prices of soybean since 2000. In order to investigate the evolution in the structure of potential function over a long observation interval, we consider the four mentioned sub-samples of three years each. A summary of descriptive statistics for spot prices of full sample and related sub-samples are reported in Table 5.5 (upper section). The increasing mean price shows the changes in soybean market starting from the period 2003-2005, and especially in the period 2006-2008 corresponding to the crisis, where the volatility has increased as well. The estimates of diffusion parameter, $\widehat{\gamma}$, in the periods 2003-2005 and 2006-2008 show the increasing magnitude of the volatility, whereas the highest level in the period 2009-2011 is related to the variance of the residuals, $V(\widehat{\epsilon})$, which not always is in agreement with the diffusion parameter, thus requiring a further study of the residuals, in order to improve the model fit. Comparing distributional characteristics of the observed prices and the forecasted ones confirms that the behaviour of the price evolution is captured remarkably well. New market conditions are reflected by the functional form of the potential, where changes in shape capture new price equilibrium levels (attraction regions). An estimate of the (scaled) potential function fitted with a mixture model of $K = 5$ Gaussian components and the corresponding derivative of the potential (reversion rates) are reported in Figure 5.12. In order to detect the changes in the market, the model can be regularly refitted every few months (i.e. every 6-12 months), since the underlying price time series is daily. In terms of the mean square prediction error, the resulting model is capable of reducing uncertainty about the future behavior of the price process and thus allowing better predictions. The estimates $\widehat{MSPE}(1)$ and $\widehat{MSPE}(2)$ have the same meaning adopted for the case of crude oil (testing set: the last 500 observed values for the same period and the first 500 observed values for the subsequent period). In this case as well, and even more, differences in the error magnitude clearly show the need for a greater accuracy in estimating the structure of the potential function and the corresponding invariant distribution. In the case of soybean, it would be useful to consider a shorter period and to refit the model on it. In this case as well, better performances are observed in predicting the direction of the next price move in terms of up-down moves.

In conclusion, when the potential model is applied to the soybean prices the main features of the data are detected at a satisfactory level. A more accurate estimate of the potential function (i.e. by means of a different mixture model) would be suitable for improving the model fit. Indeed, new attraction regions arising from new market conditions and changes in the variables (forces) acting on the market are detected remarkably well by the potential model.

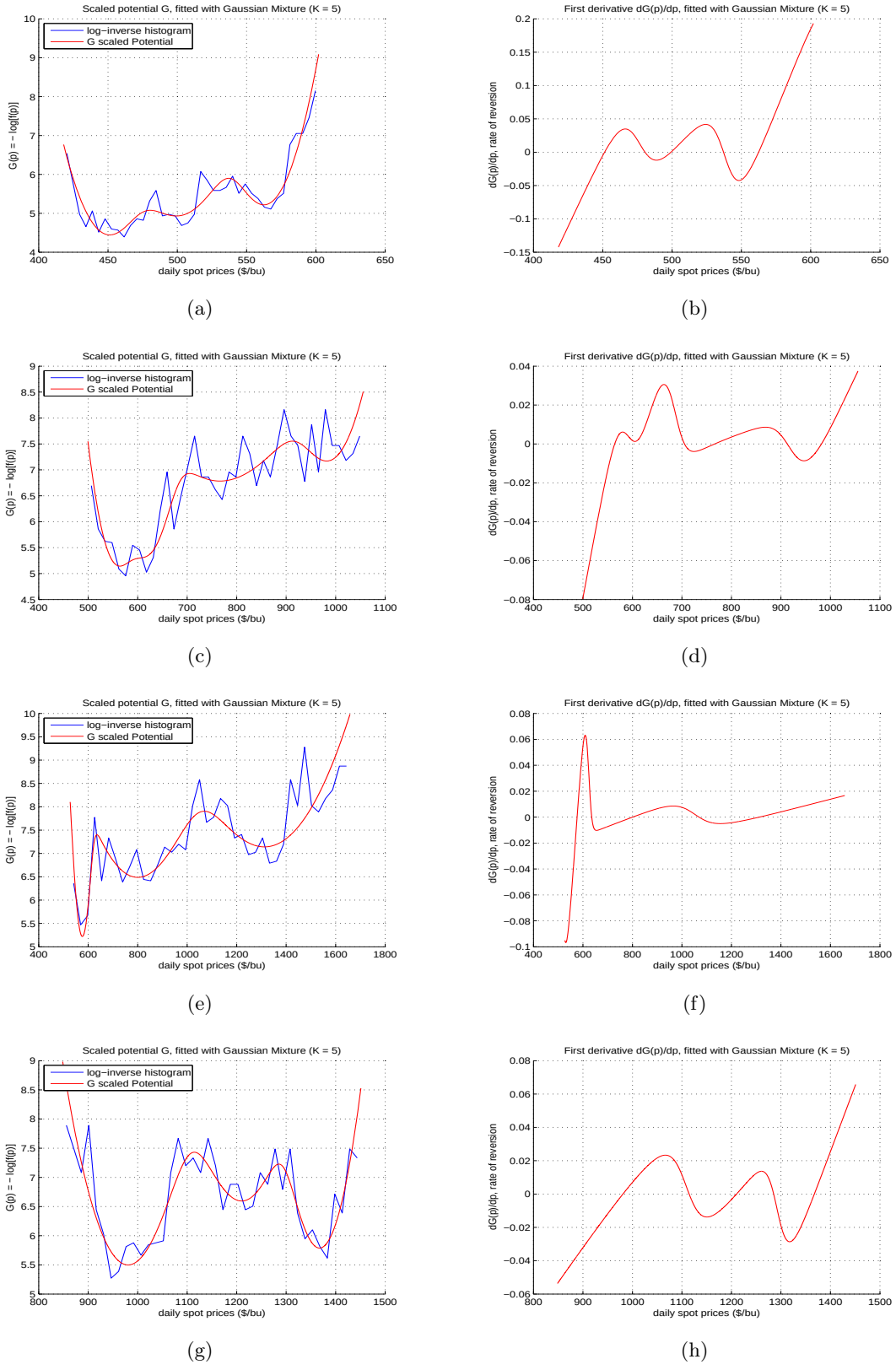


Figure 5.12: An estimate of the (scaled) potential function fitted with a mixture model of $K = 5$ Gaussian components and the corresponding derivative of the potential (reversion rates). Dataset: Soybean spot price data concerning the four sub-samples for the periods (a)-(b) 2000-2002, (c)-(d) 2003-2005, (e)-(f) 2006-2008, (g)-(h) 2009-2011.

Chapter 6

A Numerical Implementation

A numerical implementation of the present analysis is provided. The implemented code concerns the potential function model, the Expectation-Maximization algorithm, the testing performance of the estimated model, the numerical schemes for simulating price trajectories, the goodness-of-fit test for the stochastic differential equation model.

The code has been developed using MATLAB, version `7.10.0.499 (R2010a)`.

6.1 A preliminary data analysis

```
% ------------------------------------------------------------------------
% Description: The present listing loads the dataset observations, p(i=1:N),
%              computes usual descriptive statistics with some tests, and
%              plots data time series and data histogram.
%
%  References: [] ...
%
%       Filename: EM1_DataStatisticsStart1D.m
% MATLAB version: 7.10.0.499 (R2010a)
%          Author: ...
%       Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

clear all                          % -------- % ...let's do some cleaning
close all                                     %
%clc                                           %

                                   % -------- % --- INPUT DATA
                                               % --- WTI crude oil daily spot prices
load -ascii RWTC19862012.txt                   % load 1986-2012 sample
                                               % --- select sub-samples
%RWTC19861992 = RWTC19862012(1:1794,:);        % Jan1986-Dec1992
RWTC19931999 = RWTC19862012(1795:3589,:); % Jan1993-Dec1999
%RWTC20002006 = RWTC19862012(3590:5397,:); % Jan2000-Dec2006
%RWTC20072012 = RWTC19862012(5398:6813,:); % Jan2007-Dec2012

                                               % --- Soybean daily spot prices
%load -ascii SOY20002011.txt                   % load 2000-2011 sample
                                               % --- select sub-samples
%SOY20002002 = SOY20002011(1:755,:);           % Jan2000-Dec2002
%SOY20032005 = SOY20002011(756:1511,:);        % Jan2003-Dec2005
%SOY20062008 = SOY20002011(1512:2267,:);       % Jan2006-Dec2008
%SOY20092011 = SOY20002011(2268:2978,:);       % Jan2009-Oct2011
```

```
                                % ------- % --- RESET THE DATA
data = RWTC19931999;            % set 'data'
N = length(data);               % count the number of observations
t = [1:N];                      % set 't' as time vector
p = data(:,2);                  % set 'p' as price vetcor

                                % ------- % --- DATA STATISTICAL ANALYSIS
                                % --- Descriptive Statistics
Pmu = mean(p);                  % mean (as initial condition)
StatsP = DataStatistics(p);     % compute sample statistics: mean, var, std,
                                % skewness, kurtosis, exkur, min, max
%disp('')                       % display sample statistics

                                % ------- % --- GRAPHICS OF OBSERVED DATA
figure(1)                       % --- plot closing prices over time
%subplot(1,2,1)
%axis([0 1570 5 30])
xlabel('trading days')
ylabel('$/bbl')
%ylabel('$/Gallon')
%ylabel('$/bu')
title('WTI Crude Oil: daily spot prices (1993 - 1999)')
%title('Heating Oil No.2: daily spot prices (2000 - 2006)')
%title('Soybean: daily spot prices (2009 - 2011)')
grid on
hold on
plot(t,p,'r-')                  % time series of observed data
set(gca,'Fontsize',10)
%set(gca,'xtick',index);
%set(gca,'xticklabel', ...      % WTI crude oil
%       'Jan93|Oct93|Jul94|May95|Feb96|Nov96|Sep97|Jun98|Mar99|Dec99'),
%set(gca,'xticklabel', ...      % Soybean
%       'Jan00|Dec01|Dec03|Dec05|Dec07|Dec09|Oct11'),
%set(gca,'xlim',[1 rows(t')]);

figure(2)                       % --- histogram closing prices
%subplot(1,2,2)
%axis([]);
xlabel('daily spot prices ($/bbl)');
%xlabel('daily spot prices ($/Gallon)');
%xlabel('daily spot prices ($/bu)');
ylabel('number of observations');        % absolute frequencies
title('WTI Crude Oil: histogram of daily spot prices (1993 - 1999)');
%title('Heating Oil No.2: histogram of daily spot prices (2000 - 2006)');
%title('Soybean: histogram of daily spot prices (2009 - 2011)');
grid on
hold on
hist(p,50);                     % set the number of 'bins'
```

6.2 EM algorithm for mixture of Gaussians

```
function [par,loglike] = EM_MixtureGaussian_Algorithm1D(X,K)

% ------------------------------------------------------------------------
%
% Description: This function implements an Expectation-Maximization method for
%              estimating the parameters (means, variances, mixing) of a Finite
%              Mixture model of Gaussians, in 1-dimensional case (1D), with
%              K components
%
%       INPUT:
%           X  dataset of observations (vector)
%           K  number of components
%
%      OUTPUT:
```

```
%     par(:,k)  matrix Kx3 of mixture parameters
%                 (rows) K = number of components
%               (columns) 3 = means, variances, mixing
%      loglike  log-likelihood function w.r.t. number of iterations
%
%  References: [] ...
%
%        Filename: EM_MixtureGaussian_Algorithm1D.m
% MATLAB version: 7.10.0.499 (R2010a)
%          Author: ...
%       Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------------


                                % ------- % --- SET PARAMETERS
%K = 5;                           % set the number of mixture components
max_iter = 100;                 % max number of iterations
N = length(X);                  % total number of observations
Xmin = min(X);                  %
Xmax = max(X);                  %
                                % --- Tolerance Levels
tol_var = 1e-8;                 % variances
tol_mix = 1e-6;                 % mixing coefficients

                                % ------- % --- E-M ALGORITHM
                                % --- Data Preparation
X = X(:)';                      % set data in 'row' vector
X_rep = repmat(X,K,1);          % generate KxN matrix

                                % --- Step 1: initialitation of parameters
                                % choose initialitation (in 'column' vector)
mu = linspace(Xmin,Xmax,K+2)';  %
mu = mu(2:(K+1));               %
var = ones(K,1) * cov(X');      % variances of the same value
mix = ones(K,1) / K;            % mixing of the same value

for i=1:max_iter                % --- Start Iterations
                                % ------- % --- E step (expectation)
                                % --- Prepare data for computing Gaussian PDF
    mu_rep = repmat(mu,1,N);    % creo matrici KxN (come dimensioni X_rep)
    var_rep = repmat(var,1,N);  %
    mix_rep = repmat(mix,1,N);  %

    qform = (X_rep - mu_rep).^2 ./ var_rep;      % quadratic form of Gaussian
    normfactor = 1 ./ sqrt(2 * pi * var_rep);    % normalizing factor
    gaussian = normfactor .* exp(-0.5 * qform); % computes PDF Gaussian

    w = mix_rep .* gaussian;    % numeratore of E(z_nk) = priori * gaussian

                                % --- compute log-likelihood (i-th iteration)
    sum_K = sum(w,1);           % sum w.r.t. K (elements of each column)
    loglike(i) = sum(log(sum_K)); % sum w.r.t. N, computes log-likelihood

    sum_K_rep = repmat(sum_K,K,1); % denominatore of E(z_nk)
    w = w ./ sum_K_rep;         % compute E(z_nk) = posterior
                                % (they are the weights of the joint
                                % log-likelihood (X,Z))

                                % ------- % --- M step (maximization)
    post_n = sum(w,2);          % sum the posterior w.r.t. N

                                % --- Update Parameters
    mix = 1/N * post_n;         % update mixing
    mu = 1./post_n .* sum(X_rep .* w,2); % update means
    var = 1./post_n .* sum(X_rep.^2 .* w,2) ...
          - mu.^2 + tol_var;    % update variances
```

```
                                % ------- % --- Control of Tolerance Levels
    if (sum(mix < tol_mix) >= 1)          %
        index = 0;
        for k=1:K
            if (mix(k) > tol_mix) && (var(k) > small)
                index = index + 1;
                mix(index) = mix(k);
                mu(index) = mu(k);
                var(index)= var(k);
                w(index,:) = w(k,:);
            end
        end
                                          % ---
        mix = mix(1:index);               %
        mu = mu(1:index);                 %
        var = var(1:index);               %
        w = w(1:index,:);                 %
        K = index;                        %
        X_rep = repmat(X,K,1);            %
    end                                   % ---
end                                       % --- End iterations

                                          % --- Estimated Parameters
par = zeros(K,3);                         % pre-allocation
par(:,1) = mu;                            %
par(:,2) = var;                           %
par(:,3) = mix;                           %

return;                                   %
```

```
% ------------------------------------------------------------------------
% Description: The present listing estimates the multimodal probability density
%              function, f(p), by means of fitting the resulting histogram of
%              the historical data, p(i=1:N). In the framework of Finite Mixture
%              models, we postulate a parametric form of the invariant price
%              distribution, f(p), as follows
%
%              f(p) = sum(k=1:K) alfa(k) * N(p;mu(k),var(k))
%
%              that is in the context of mixture of Gaussians, N(p;mu,var), with
%              K components. In order to estimate the mixture parameters, we use
%              an Expectation-Maximization (EM) algorithm.
%
%       Used
%   functions: EM_MixtureGaussian_Algorithm1D(X,K).m
%              MixtureGaussianPDF1D(X,par).m
%
%  References: [] ...
%
%       Filename: EM1_MixtureGaussian1D.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                % ------- %  --- EM ALGORITHM
K = 5;                                    % set the number of mixture components

[par,like] = EM_MixtureGaussian_Algorithm1D(p,K); % ...

%disp('')                                 % display the estimated parameters

                                % ------- % GRAPHICS OF EM ESTIMATED DATA
                                          % --- Prepare data for 'plot'
pMin = StatsP(7);                         %
```

```
pMax = StatsP(8);                          %
pp = pMin:0.1:pMax;                        % set the price axis (abscissa)
%pp = linspace(Pmin,Pmax,...);

MixturePDF = zeros(K+1,length(pp));        % pre-allocation K components on 'pp'
%Mixture_pdf = ...

for k=1:K                                  % evaluate each k-th component on 'pp'
    MixturePDF(k,:) = MixtureGaussianPDF1D(pp,par(k,:));
end
MixturePDF(K+1,:) = MixtureGaussianPDF1D(pp,par); % evaluate Mixture on 'pp'

                                           % --- Prepare data for 'histogram'
n_int = 40;                                % set the number of subintervals ('bins')
[fqz_ass,bc] = hist(p,n_int);              % compute absolute frequencies
                                           % ...and center of each 'bin'
b_bar = bc(2) - bc(1);                     % set the lenght of histogram bar
fqz_rel = fqz_ass / N;                     % compute relative frequencies
                                           % ..., i.e., area = lenght x height
h_bar = fqz_rel / b_bar;                   % compute the height of each bar
                                           % i.e. height = area / lenght

figure(3)                          % -------- % --- Plot Components of mixture
bar(bc,h_bar);                             % histogram of observed data
for k=1:K
    hold on
    plot(pp,MixturePDF(k,:), ...           % plot each k-th component
        'LineWidth',1.5, ...
        'Color',[1,0,0])
end
hold on
xlabel('daily spot prices ($/bbl)');
%xlabel('daily spot prices ($/bu)');
%ylabel('relative frequencies');            % relative frequencies
title('Invariant distribution fitted with a mixture of Gaussians (K = 5)');
legend('histogram','K components EM')
hold off

figure(4)                          % -------- % --- Plot Mixture
bar(bc,h_bar);                             % histogram of observed data
hold on
plot(pp,MixturePDF(K+1,:), ...             % plot Mixture
    'LineWidth',1.5, ...
    'Color',[0,1,0])
xlabel('daily spot prices ($/bbl)');
%xlabel('daily spot prices ($/bu)');
%ylabel('relative frequencies');            % relative frequencies
title('Density function estimated with a mixture of Gaussians (K = 5)');
legend('histogram','Mixture EM')
hold off

figure(5)                          % -------- % --- Log-likelihood function
plot(like,'-', ...                         % versus number of iterations
    'LineWidth',1.5)
xlabel('number of iterations');
ylabel('log-likelihood function');
title('Log-likelihood convergence for EM iterations')
```

6.3 Computing the (scaled) potential function

```
% ------------------------------------------------------------------
% Description: The present listing computes the scaled potential function, G(p),
%              as follows
%
%              G(p) = - log[f(p)]
%
```

```
%               where the probability density function, f(p), has been estimated
%               by means of a Finite Mixture model of Gaussians, with K components.
%               In order to compare the behaviour of the estimated mixture, we
% 10            use also Gbar = - log(h_bar), i.e. the scaled potential w.r.t.
%               the histogram of historical data.
%               The listing also computes the first derivative of scaled potential,
%               G'(p), as follows
%
% 15            dG(p)                          1
%               ------ = [-log(f(p))]' = - ------ * f'(p)
%                dp                         f(p)
%
%  References: [] ...
% 20 %
%       Filename: EM2_G_potential.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
% 25 %
% NOTES: ...
% ------------------------------------------------------------------------

                                    % ------- % --- G POTENTIAL MODEL (RESCALED)
30 Gbar = - log(h_bar);             % estimate G from height bar histogram
   Gmix = - log(MixturePDF(K+1,:)); % estimate G from Gaussian Mixture
                                    % --- Further Possible Estimates
   %Gfqz = - log(fqz_rel);          % estimate G from relative frequencies
   %Gpol = - log(y_hat);            % estimate G from polynomial fit
35
                                    % --- Graphics of 'G' scaled potential
   figure(6)                        % plot G (scaled) potential: we want
   %axis([]);                       % to see what is the behaviour of the
   xlabel('daily spot prices ($/bbl)'); % estimate by mixture of Gaussians
40 %xlabel('daily spot prices ($/bu)'); % estimate by mixture of Gaussians
   ylabel('G(p) = - log[f(p)]');    % w.r.t. the log-inverse of the histogram
   title('Scaled potential G, fitted with Gaussian Mixture (K = 5)');
   grid on
   hold on
45 plot(bc,Gbar);                   % plot 'Gbar' against 'bc'
   plot(pp,Gmix,'r');               % plot 'Gmix' against 'pp'
   legend('log-inverse histogram', ...
          'G scaled Potential', ...
          'Location','NW')          % NW = North West
50 hold off

                                    % ------- % --- G' FIRST DERIVATIVE
                                    % compute the first derivative of G
                                    % evaluated over 'pp'
55 GfirstNum = MixtureGaussianPDFfirst(pp,par); % compute Numerator of Gfirst
   GfirstDen = MixtureGaussianPDF1D(pp,par);    % compute Denominator of Gfirst
   Gfirst_pp = - (GfirstNum ./ GfirstDen);      % compute the first derivative of G

                                    % --- GRAPHICS OF G'(p)
60 figure(7)                        % plot G'(p): we want to see what is the
   %axis([]);                       % behaviour of the first derivative of
   xlabel('daily spot prices ($/bbl)'); % the scaled potential, G'(p), estimated
   %xlabel('daily spot prices ($/bu)'); % the scaled potential, G'(p), estimated
   ylabel('dG(p)/dp, rate of reversion'); % by means of a mixture of Gaussians
65 title('First derivative dG(p)/dp, fitted with Gaussian Mixture (K = 5)');
   grid on
   hold on
   plot(pp,Gfirst_pp,'r');          % plot 'Gfirst' against 'pp'
   %legend('...', ...
70 %       '...', ...
   %       'Location','NW')         % NW = North West
   hold off
```

6.4 Computing the diffusion parameter

```
% ------------------------------------------------------------------------
% Description: The present listing implements a regression by means of the
%              the Least Square (LS) method. The linear regression model is
%
%              Y = gamma * X + epsilon
%
%              where, for i=1,2,...,n-1
%
%              Y = p(t(i+1)) - p(t(i))     <- the price increments
%
%                     G'(p(t(i)))
%              X = - ---------------       <- the scaled first derivative of G
%                          2
%
%              so that 'X' requires the evaluation of the first derivative
%              of the scaled potential, G'(p(t(i))), w.r.t. the current price
%              value, p(t(i)). The regression parameter 'gamma' is
%
%              gamma = kappa^2 * (t(i+1)-t(i))
%
%              which has to be estimated.
%
%  References: [] ...
%
%       Filename: EM3_regression_LS.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                % --------- % --- Y
                                          % --- Price Increments
p = p(:);                                 % reset prices in 'column' vector
Pstart = p(1:end-1,1);                    % extract p(1) to p(n-1)
Pend = p(2:end,1);                        % extract p(2) to p(n)
Y = Pend - Pstart;                        % compute price increments

                                % --------- % --- X
Gfirst = MixtureGaussianGfirst(p,par);    % compute first derivative of G
                                          % evaluated over 'p'

                                          % --- Data preparation for 'Gfirst'
Gfirst = Gfirst(:);                       % reset G' in 'column' vector
Gfirst_Pstart = Gfirst(1:end-1,1);        % extract G'(p(1)) to G'(p(n-1))
                                          % ...according to size of 'Pstart'

                                % --------- % --- LEAST SQUARE REGRESSION
                                          % --- Computation of 'gamma'
gammaNum = sum(Y .* Gfirst_Pstart);       % sum over N for Numerator
gammaDen = sum(Gfirst_Pstart .^2);        % sum over N for Denominator
gamma = -2 * (gammaNum / gammaDen)        % compute estimate of 'gamma'

                                % --------- % --- LS REGRESSION LINE
X = -0.5 .* Gfirst_Pstart;                % rescale w.r.t. (-1/2)
Xreg = min(X):0.01:max(X);                % set the X axis
Yhat = gamma * Xreg;                      % compute the fitted values

                                % --------- % --- GRAPHICS
figure(8)                                 %
scatter(X,Y,3,[1,0,0])                    % scatter plot of observations
hold on
%scatter(Xreg,Yhat,3,[0,1,0])              % scatter plot of fitted values
plot(Xreg,Yhat,'b')                       % plot of fitted values
xlabel('X, scaled first derivative of G');
```

```
ylabel('Y, price increments');
title('Regression model: Price increments vs. Scaled first derivative of G');
legend('dataset of observations', ...
       'LS regression line', ...
       'Location','NW')                    % NW = North West
hold off
```

6.5 The estimated model and diagnostics

```
% ------------------------------------------------------------------------
% Description: The present listing implements the estimated model, as follows
%
%              Pnextday(t(i)) = p(t(i-1)) - 0.5 * gamma * G'(p(t(i-1)))
%
%              P(0) = Pzero,                    i = 1,...,n
%
%              where gamma = kappa^2 * (t(i)-t(i-1)), that is on the
%              basis of the estimated scaled potential function, G, and the
%              estimated diffusion parameter kappa^2 (or gamma).
%              The listing also provides a standard diagnostics on the basis
%              of the residuals
%
%              e(t(i)) = p(t(i)) - Pnextday(i)          N(0,gamma)
%
%              which should be uncorrelated and approximately normally
%              distributed with mean mu = 0, and variance var = gamma.
%              The listing introduce an error measure to test the difference
%              between observed data and estimated data.
%              All the graphics are provided.
%
%  References: [] ...
%
%       Filename: EM5_EstimatedModel.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                % ------- % --- DATA & PARAMETERS
                                % --- Discretization
T = length(p);                  % set time interval [t(0),T] with t(0)=0
%N = N;                          % set number of steps to compute next price
dt = T/N;                       % compute time stepsize
                                %
                                % --- ReSet Parameters
h = dt;                         % consider time stepsize
kappa_squared = gamma / h;      % compute 'kappa^2' diffusion parameter
%alpha = 0.5 * kappa_squared;    % compute drift parameter

                                % ------- % --- ESTIMATED MODEL
Pnextday = zeros(N,1);          % incoming forecasted next day price (PfE)
Pnextday(1) = p(1);             %

for i = 1:N-1
    Gfirst_P = MixtureGaussianGfirst(p(i),par);
    Pnextday(i+1) = p(i) - 0.5 * kappa_squared * Gfirst_P * dt;
end

                                % ------- % --- DATA STATISTICAL ANALYSIS
StatsPnd = DataStatistics(Pnextday);  % compute sample statistics
%disp('')                       % display sample statistics

                                % ------- % --- GRAPHICS OF ESTIMATED DATA
figure(9)                       % --- plot price process over time
```

```
%subplot(1,2,1)
%axis([]);                                  % ...
xlabel('t (trading days)')
ylabel('P(t), Dollars per Barrel ($/bbl)')
%ylabel('P(t), Dollars per Bushel ($/bu)')
title('WTI Crude Oil: observed and forecasted price (1993 - 1999)')
%title('Soybean: observed and forecasted price (2009 - 2011)')
grid on
hold on
plot(t,p,'b-')                              % observed price
plot(t,Pnextday,'r--')                      % forecasted next day price
set(gca,'Fontsize',10)
%set(gca,'xtick',index);
%set(gca,'xticklabel', ...
%        'Jan93|Oct93|Jul94|May95|Feb96|Nov96|Sep97|Jun98|Mar99|Dec99'),
%set(gca,'xlim',[1 rows(t')]);
legend('observed data', ...
       'estimated data', ...
       'Location','NW')                     % NW = North West
hold off

figure(10)                                  % --- histogram forecasted prices
%subplot(1,2,2)
%axis([]);
xlabel('daily spot prices ($/bbl)');
%xlabel('daily spot prices ($/bu)');
ylabel('number of observations');           % absolute frequencies
%title('WTI Crude Oil: histogram of forecasted daily price (1986 - 2012)');
title('Soybean: histogram of forecasted daily price (2009 - 2011)');
grid on
hold on
hist(Pnextday,50);                          % set the number of 'bins'
hold off

                                % ------- % --- DIAGNOSTICS ON RESIDUALS
residuals = p - Pnextday;                   % compute residuals
                                            % --- Descriptive Statistics
StatsErr = DataStatistics(residuals);       % compute sample statistics
                                            % mean, var, std, skewness,
                                            % kurtosis, excess kurtosis
%disp('')                                   % display sample statistics

                                            % --- Prepare data for 'histogram'
                                            % --- Normal PDF for errors
nbins = 50;                                 % set the number of 'bins'
[f_abs,bcenter] = hist(residuals,nbins);    % compute absolute frequencies
                                            % ...and center of each 'bin'

Bbar = bcenter(2) - bcenter(1);             % set the lenght of histogram bar
f_rel = f_abs / length(residuals);          % compute relative frequencies
                                            % ..., i.e., area = lenght x height
Hbar = f_rel / Bbar;                        % compute the height of each bar
                                            % i.e. height = area / lenght

errorPDF = normpdf(bcenter,0,sqrt(gamma));% compute Normal PDF

%figure(11)                                 % --- histogram residuals
%subplot(1,2,2)
%axis([]);
%xlabel('model residuals');
%ylabel('');                                % absolute frequencies
%title('Histogram of the model residuals');
%grid on
%hold on
%hist(residuals,nbins);                     % set the number of bins
%hold off

figure(11)                                  % --- plot error and residuals
```

```
%subplot(1,2,2)
%axis([]);
xlabel('model residuals');
%ylabel('');                                % absolute frequencies
title('Histogram of the model residuals and Normal(0,gamma)');
grid on
hold on
bar(bcenter,Hbar);                         % histogram of residuals
plot(bcenter,errorPDF, ...                 % normal PDF superimposed
     'LineWidth',1.5, ...
     'Color',[1,0,0]);
legend('model residuals', ...
       'Normal(0,gamma)')
%       'Location','NW')                    % NW = North West
hold off

                                           % --- AutoCorrelation Function (ACF)
lags = 50;                                 % set the number of lags

figure(12)                                 % plot sample ACF
%grid on
autocorr(residuals,lags)                   % compute sample ACF
hold on
axis([0 50 -0.2 1.2]);
xlabel('Lag');
ylabel('Autocorrelation');                 %
title('Sample autocorrelation function (ACF) of the model residuals');
hold off
```

6.6 Testing the predictive power of the model

```
% ------------------------------------------------------------------------
% Description: The present listing implements the estimation and evaluation
%              of the predictive power of the model.
%
%
%
%  References: [] ...
%
%       Filename: EM6_PredictivePower.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                % ------- % --- TEST MODEL'S PREDICTIVE POWER
                                          % --- TESTING SET
%Ptest = p(1:500,1);                       % the first 500 observations IN
Ntest = N - 499;                          %
Ptest = p(Ntest:end,1);                   % the last 500 observations IN
%Ptest = SOY20092011(1:500,2);             % the first 500 observations OUT
m = length(Ptest);                        %

                                % ------- % --- PREDICTIONS
                                          % Use parameter estimates coming
                                          % from the training set.
Phat = zeros(m,1);                        % incoming predictors (PfE)
Phat(1) = Ptest(1);                       %

for i = 1:m-1
    Gfirst_Ptest = MixtureGaussianGfirst(Ptest(i),par);
    Phat(i+1) = Ptest(i) - 0.5 * kappa_squared * Gfirst_Ptest * dt;
end
```

```
                                % ---------- % --- MEAN SQUARED PREDICTION ERROR
PE = Ptest - Phat;              % compute Prediction Error
MSPE = sum(PE .^ 2) / m         % compute MSPE

                                % ---------- % --- GRAPHICS OF ESTIMATED DATA
figure(13)                      % --- plot price process over time
%subplot(1,2,1)
%axis([1 500 10 40]);           % ...
xlabel('t (trading days), testing set: Jan 1998 - Dec 1999')
ylabel('P(t), Dollars per Barrel ($/bbl)')
%ylabel('P(t), Dollars per Bushel ($/bu)')
title('WTI Crude Oil: observed and forecasted price for Testing Set')
%title('Soybean: observed and forecasted price for Testing Set')
grid on
hold on
plot(1:m,Ptest,'b-')            % observed price
plot(1:m,Phat,'r--')            % forecasted next day price
set(gca,'Fontsize',10)
%set(gca,'xtick',index);
%set(gca,'xticklabel', ...
%       'Jan98|Jun98|Oct98|Mar99|Aug99|Dec99'),
%set(gca,'xlim',[1 rows(t')]);
legend('observed data', ...
       'estimated data', ...
       'Location','NW')         % NW = North West
hold off

                                % ---------- % --- TEST EVOLTUION DIRECTION
p = p(:);                       % reset prices in 'column' vector
                                % --- Compute Price Increments
Pinc = p(2:end,1) - p(1:end-1,1);% true over true
Pinc = [0;Pinc];                %
PincND = Pnextday(2:end,1) - p(1:end-1,1);% next day over true
PincND = [0;PincND];            %

Pdir = zeros(N,1);              % incoming Directions (PfE)
Pdir(1) = 1;                    %

for i = 1:N-1
    if (Pinc(i+1)>=0 && PincND(i+1)>=0) || (Pinc(i+1)<=0 && PincND(i+1)<=0)
       Pdir(i+1) = 1;           % same sign = same direction
    else
       Pdir(i+1) = 0;           % ...otherwise
    end
end

DIRpc = (sum(Pdir)/N) * 100     % compute DIRection in percentage
```

6.7 Numerical schemes and simulations of price process

```
% ------------------------------------------------------------------
% Description: This listing implements the Euler-Maruyama method for an
%              approximation solution of a Stochastic Differential Equations
%              (SDEs). The SDE model is
%
%              dP(t) = - U'(p(t)) * dt + kappa * dB(t)
%
%              and the approximation solution is
%
%              P(k) = P(k-1)
%                     - 0.5 * kappa_squared * Gfirst(k-1) * (t(k)-t(k-1))
%                     + kappa * (B(k)-B(k-1))
%
%              and P(0) is the initial condition. The listing also provides
%              for simulations of price process trajectories.
%
```

```
% References: [1] Higham D.J. (2001) An Algorithmic Introduction to Numerical
%                  Simulation of Stochastic Differential Equations,
%                  SIAM Review, 43(3), pp. 525-546
%
%       Filename: EM4_sdeEulerMaruyama.m
% MATLAB version: 7.10.0.499 (R2010a)
%          Autor: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                % ------- % --- Brownian Motion (BM)
                                % --- Discretization
T = 1800;                       % set time interval [t(0),T] with t(0)=0
N = 1800;                       % set number of steps to compute BM over [t(0),T]
dt = T/N;                       % compute time stepsize

                                % --- Simulated Trajectories
%randn('state',100)              % set the state of 'randn' (repeatable trials on/
    off)
M = 1;                          % set numbers of trajectories
dW = sqrt(dt) * randn(M,N);     % Brownian increments
%W = cumsum(dW);                % Brownian path(s), trajectories (cumulated
    increments)

                                % ------- % --- Set Parameters of SDE
                                % --- Computation of 'kappa_squared'
h = dt;                         % consider time stepsize
kappa_squared = gamma / h;      % compute 'kappa_squared'
                                %
%Pzero = Pmu;                    % initial condition: mean price
Pzero = p(1);                   % initial condition: p(1)
%alpha = 0.5 * kappa_squared;    % drift parameter
kappa = sqrt(kappa_squared);    % diffusion parameter

                                % ------- % --- Euler-Maruyama (E-M) method
                                % --- Discretization
R = 1;                          % set the value of R inter-steps
L = N/R;                        % set number of steps to compute E-M iterations
Dt = R*dt;                      % compute time inter-stepsize for E-M

                                % ------- % --- Preallocate for Efficiency (PfE)
Pem = zeros(1,L);               % incoming price path(s)
Ptemp = Pzero;                  % 'temporary' variable for iterations

for k = 1:L
    dWem = sum(dW(R*(k-1)+1:R*k));          % sum over R the original dW increments
    Gfirstdot = MixtureGaussianGfirst(Ptemp,par);
    Ptemp = Ptemp - 0.5 * kappa_squared * Gfirstdot * Dt + kappa * dWem;
    Pem(k) = Ptemp;             % store incoming price path(s)
end

                                % ------- % --- RESET THE DATA
tem = 0:Dt:T;                   % reset time axis
Ptem = [Pzero,Pem];             % reset trajectories

                                % ------- % --- DATA STATISTICAL ANALYSIS
StatsPtem = DataStatistics(Ptem);   % compute sample statistics
                                % mean, var, std, skewness,
                                % kurtosis, excess kurtosis
StatsPtem = StatsPtem(:);       % reset stats in 'column' vector
%disp('')                       % display sample statistics

                                % ------- % --- GRAPHICS OF SIMULATED DATA
figure(16)                      % --- plot simulated price trajectory
%subplot(1,2,1)
%axis([0 1 0 10]);              % ...
```

```
xlabel('t', ...
       'fontsize',10);
ylabel('P(t)', ...
       'fontsize',10,...
       'rotation',0,...
       'HorizontalAlignment','right');
title('WTI Crude Oil: daily simulation of price process',...
      'fontsize',10,...
      'Color',[.3 .3 .3]);
grid on
hold on
plot(tem,Ptem,'r-')                          % ...

figure(17)                                   % ---- histogram simulated data
%subplot(1,2,2)
%axis([]);
xlabel('daily spot prices ($/bbl)');
ylabel('number of observations');            % absolute frequencies
title('WTI Crude Oil: histogram of simulated daily prices');
grid on
hold on
hist(Ptem,40);                               % set the number of 'bins'
```

6.8 A goodnes-of-fit test for the SDE model

```
% ------------------------------------------------------------------------
% Description: A goodness-of-fit test for a SDE model
%
%                                                        [(Omega(q)-((N-1)/(M+1)))]^2
%     test statistic: X^2(M) = sum(q=1:M+1) ------------------------------------------
%                                                              [(N-1)/(M+1)]
%        rank values: q = 1,2,...,M+1
% observed frequency: Omega(q)
% expected frequency: (N-1)/(M+1)
%
%  References: [1] Pearson K. (1900) On a criterion that a given system of derivations
%                  from the probable in the case of a correlated system of variables
%                  is such that it can be reasonably supposed to have arisen in random
%                  sampling, Phil. Mag., Ser. 5, 157-172
%              [2] Bak J. (1998) Nonparametric Methods in Finance
%              [3] Allen E. (2007) Modelling with Ito Stochastic Differential Equations
%
%       Filename: EM9_sdeChiSquareTest.m
% MATLAB version: 7.10.0.499 (R2010a)
%         Author: ...
%      Update to: ...
%
% NOTES: ...
% ------------------------------------------------------------------------

                                     % ------- % ---- SDE Parameters
Pzero = 1;                           % initial condition
alpha = 2;                           % drift parameter
beta = 1;                            % diffusion parameter

                                     % ------- % ---- Brownian Motion (BM)
                                     % ---- Discretization
T = 1;                               % set time interval [t(0),T] with t(0)=0
N = 10;                              % set number of steps to compute BM over [t(0),T]
dt = T/N;                            % compute time stepsize

                                     % ---- Simulated Trajectories
%randn('state',100)                   % set the state of 'randn' (repeatable trials on/
      off)
M = 5;                               % set numbers of trajectories (rule-of-thumb: M=(N
      -6)/5)
```

```
40
                                % ---------- % --- Euler-Maruyama (E-M) method
                                               % --- Discretization
R = 1;                                         % set the value of R inter-steps
L = N/R;                                       % set number of steps to compute E-M iterations
45 Dt = R*dt;                                  % compute time inter-stepsize

                                % ---------- % --- Preallocate for efficiency (PfE)
Pchi = zeros(M,L); % Pem                       % create MxL array for trajectories
Prank = zeros(M,1);                            %
50 %rank = zeros(1,N-1);

Puno = 10;

%for i = 1:N-1
55     for m = 1:M                             % --- Iteration for m-th trajectory
           dW = sqrt(dt) * randn(1,N);         % generate 1xN array of Brownian increments
           Ptemp = Pzero;                      % initial condition for temporary variable
           for k = 1:L                         % --- Iteration for k-th step
               dWchi = sum(dW(R*(k-1)+1:R*k)); % depending on R, sum over [] the original dW
                     increments
60             Gfirstdot = MixtureGaussianGfirst(Ptemp,par);
               Ptemp = Ptemp ...
                       - 0.5 * kappa_squared * Gfirstdot * Dt ...
                       + kappa * dWem;
               Pchi(m,k) = Ptemp;              % update Ptemp for m-th trajectory and k-th step
65         end
           if Pchi(m,L) > Puno                 % ---------- % --- Rank
              Prank(m) = 0;                    % indicator function for rank
           else
              Prank(m) = 1;
70         end
       end
       %rank(i) = 1 + sum(Prank)               % compute rank
%end

75                              % ---------- % --- Chi-square goodness-of-fit test
                                               % --- Observed frequencies
I = zeros(M+1,N-1)                             % create M+1xN-1 array for rank indicator
for i = 1:N-1
       for q = 1:M+1
80         if rank(i) == q                     % indicator function for observed frequencies
              I(q,i) = 1;
           else
              I(q,i) = 0;
           end
85     end
end
Fobs = sum(I(1,:));                            % sum over row q for observed frequency Omega(q)

                                               % --- Expected frequencies, under H0
90 Fexp = (N-1)/(M+1);                         % set the expected frequencies

                                               % --- Chi-square Test Statistic
Qi = ((Fobs - Fexp).^2)./Fexp;                 % generate the argument of sum
Q = sum(Qi);                                   % calculate the test statistic
95

                                               % --- Critical value
dof = 4;                                       % set the degrees of freedom
cl = 0.01;                                     % set the confidence level 'cl=(0,1)'
cv = chi2inv(1-cl,dof);                        % calculate the critical value
100

                                               % --- Decision Rule
if Q < Chi                                     % ...
   disp('The null hypothesis H0 is not rejected');
   else
105   disp('The null hypothesis H0 is rejected');
end
```

```
pv = 1 - chi2cdf(Q,dof);                    % calculate p-value

                                    % ------- % --- Graphics of Simulated Data
                                            % --- plot
tem = 0:Dt:T;                               %
Ptchi = [Pzero,Pchi(1,:)];                  %

%figure(7)
%subplot(1,2,1)
%axis([0 1 0 10]);                           % ...
xlabel('t', ...
       'fontsize',10);
ylabel('P(t)', ...
       'fontsize',10,...
       'rotation',0,...
       'HorizontalAlignment','right');
title('Simulation of Chi^2 test',...
      'fontsize',10,...
      'Color',[.3 .3 .3]);
grid on
hold on
plot(tem,Ptchi,'r --*')                     % ...

figure(8)                                   % --- histogram
%subplot(1,2,2)
%axis([]);
xlabel('daily spot prices ($/bbl)');
ylabel('number of observations');                % frequenze assolute
title('Crude Oil: histogram of daily spot prices (simulation)');
grid on
hold on
hist(Ptem,40);                                   % bins = 40
```

Chapter 7

Summary and Conclusions

Global commodity markets have experienced significant price swings in recent years. Analysts offer two general explanations: market forces and speculative expectations. These two explanations are not mutually exclusive, and both market forces and speculative expectations may be responsible. Since the behavior of commodity prices is different from that of more traditional financial assets, then analytical and modelling tools that take into account specific features of commodity prices are needed.

Recent developments on the significant and sharp rises and declines in commodity prices seem to indicate that various factors are acting in a very complex way. In particular, the present analysis starts from one specific characteristic feature, that is the tendency of many commodity prices to concentrate in a number of attraction regions, preferring some values over others, which is the price clustering phenomenon. Commodities are in the process of becoming mainstream. The mean-reverting class of diffusion models have been widely used to model commodity prices. However, these techniques of analysis are not able to model the phenomenon of multiple attraction regions. In order to overcame such limitations, we discuss the idea concerning the potential function approach, a nonlinear model where the evolution of the price process is governed by the potential function.

The present approach of potential model has a fundamental step on fitting the multimodal density of the invariant distribution. Our main contribution in the present analysis has been to extend the original approach. We postulate a parametric form of the invariant price distribution in the framework of finite mixture models and fit the potential by means of the maximum likelihood method with a numerical implementation of Expectation-Maximization algorithm for a finite mixture of Gaussians. Finite mixture models provide a straightforward, but very flexible extension of classical statistical models. There exist various features of finite mixture distributions that render them useful in statistical modeling. The most striking property of a mixture density is that the shape of the density is extremely flexible. Indeed, from the data-oriented perspective, it turns out that statistical models that are based on finite mixture distributions are able to capture many specific features of real data, such as multimodality, skewness, and kurtosis.

The application of the potential model to the price data shows that the essential characteristics of the data are captured remarkably well. In particular, the model is able to take into account new attraction regions arising from new market conditions and changes in the variables (forces) acting on the market. The variance of the residuals not always is in agreement with the diffusion parameter, then a further study of the residuals is needed, in order to improve the model fit.

In terms of the mean square prediction error, the resulting model is capable of reducing uncertainty about the future behavior of the price process and thus allowing better predictions. In particular, better performances are observed in predicting the direction of the next price move in terms of up-down moves.

We recall that differences in the error magnitude clearly show that the structure of the model

is dependent on the structure of the estimated potential, which in turn is related to the estimated invariant distribution. Therefore, it is important to note that the estimated model is dependent on the accuracy in estimating the invariant distribution. We pointed out the importance of the number of components in the mixture model and the more desired accuracy concerning the numerical algorithm implemented for the estimation of mixture parameters.

Concerning the functional form of the potential, changes in shape reflect new price equilibrium levels (attraction regions) and hence new market conditions. If the underlying price time series is daily, the model can be regularly refitted every few months (i.e. every 6-12 months), in order to capture the changes in the market within the potential function and the changes in volatility within the variance of the model residuals.

The model is able to generate copies of the observed price series with the same invariant distribution, which is useful for applications such as Monte Carlo analysis, scenario testing, and other studies that require a large number of independent price trajectories.

It is important to note that the main forecasting power of the model lies in its improved ability to predict the direction of the next move, once evolution departs from a local minimum. There the influence of the deterministic potential field prevails over random fluctuations, while at a local minimum the derivative of the potential field is close to zero and the evolution is largely determined by random forces. This is in agreement with economic arguments: if the price is far from an equilibrium price, external forces of the market drive the price towards nearest equilibrium, while at equilibrium price fluctuations are largely due to random shocks. A interesting extension of the potential model would be to look more closely at the behaviour of the price near equilibria.

An important question when fitting the potential model is the extrapolation of the potential into the regions of greater price moves. In the framework of finite mixture models, a possible extension of the potential approach consists in building a mixture model which is more capable of taking into account such a behaviour of the price dynamics.

In the context of finite mixture models and the Expectation-Maximization algorithm, the present model could be extended to the multivariate case in a natural way. However, some computational difficulties can arise. This extension to multivariate potential model and estimation procedure is an issue which has already been started and we will address in the future.

Bibliography

[1] M. Adelman. World oil production and prices. *The Quarterly Review of Economics and Finance*, 42:169–191, 2002.

[2] E. Allen. *Modeling with Ito Stochastic Differential Equations*. Springer, 2007.

[3] P. Ao. Potential in stochastic differential equations: novel construction. *Journal of Physics A: Mathematical and General*, 37(3):L25–L30, 2004.

[4] P. Ao. Boltzmann-Gibbs distribution of fortune and broken time reversible symmetry in econodynamics. *Communications in Nonlinear Science and Numerical Simulation*, 12(5):619–626, 2007.

[5] P. Ao, C. Kwon, and H. Qian. On the existence of potential landscape in the evolution of complex systems. *Complexity*, 12(4):19–27, 2007.

[6] J. Bak. *Nonparametric Methods in Finance*. Department of Mathematical Modelling, Technical University of Denmark, Lyngby, 1998.

[7] J. Bak, H. A. Nielsen, and H. Madsen. Goodness of fit of stochastic differential equations, 2000.

[8] M. Baker, S. Mayfild, and J. Parsons. Alternative models of uncertain commodity prices for use with modern asset pricing methods. *The Energy Journal*, 19(1):115–148, 1998.

[9] D. Bertsimas and J. Tsitsiklis. Simulated annealing. *Statistical Science*, 8(1):10–15, 1993.

[10] H. Bessembinder, J. F. Coughenour, S. Paul, and M. M. Smoller. Mean-reversion in equilibrium asset prices: Evidence from the futures term structure. *Journal of Finance*, 50:361–375, 1995.

[11] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[12] Z. Bodie. Commodity futures as a hedge against inflation. *The Journal of Portfolio Management*, 9(3):12–17, 1983.

[13] S. Borovkova, H. Dehling, J. Renkema, and H. Tulleken. A potential-field approach to financial time series modelling. *Computational Economics*, 22(2–3):139–161, 2003.

[14] M. Brennan. The supply of storage. *American Economic Review*, 48:50–72, 1958.

[15] M. Brennan and E. Schwartz. Evaluation natural resource investments. *Journal of Business*, 58(2):135–157, 1985.

[16] D. R. Brillinger, H. K. Preisler, and M. J. Wisdom. Modelling particles moving in a potential field with pairwise interactions and an application. *Brazilian Journal of Probability and Statistics*, 25(3):421–436, 2011.

[17] W. A. Brock and C. H. Hommes. Models of complexity in economics and finance. 1997.

[18] W. A. Brock and C. H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22:1235–1274, 1998.

[19] W. A. Brock, C. H. Hommes, and F. O. O. Wagener. Evolutionary dynamics in financial markets with many trader types. *CeNDEF Working Paper Series*, WP 01-01, 2001.

[20] A. Brook, R. Price, D. Sutherland, N. Westerlund, and C. Andre. Oil price developments: drivers, economic consequences and policy responses. *OECD Economics Department Working Papers, Paris*, 2004.

[21] B. Buyuksahin, M. S. Haigh, and M. A. Robe. Commodities and equities: Ever a market of one? *Journal of Alternative Investments*, 12(3):76–95, 2010.

[22] P. Cashin, C. J. McDermott, and A. Scott. Booms and slumps in world commodity prices. *Journal of Development Economics*, 69(1):277–296, 2002.

[23] V. Cherny. Thermodynamic approach to the travelling salesman problem: an efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45:41–51, 1985.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algoritm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[25] EC. High prices on agricultural commodity markets: situation and prospects. a review of causes of high prices and outlook for world agricultural markets. 2008.

[26] ECB. Oil prices - their determinants and impact on euro area inflation and the macroeconomy. monthly bulletin. 2010.

[27] F. Edwards and J. Park. Do managed futures make good investments? *Journal of Futures Markets*, 16(5):475–517, 1996.

[28] J. G. Eisenhauer. Regression through the origin. *Teaching Statistics*, 25(3):76–80, 2003.

[29] B. S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. Wiley, 2011.

[30] M. Farooki and R. Kaplinsky. *The Impact of China on Global Commodity Prices: The Global Reshaping of the Resource Sector*. 2011.

[31] S. Fruhwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.

[32] P. M. Garber. Famous first bubbles. *Journal of Economic Perspectives*, 4(2):35–54, 1990.

[33] A. Gaunersdorfer, C. H. Hommes, and F. O. O. Wagener. Bifurcation routes to volatility clustering. Preprint, 2000.

[34] S. Geman and C. R. Hwang. Diffusions for global optimization. *Journal on Control and Optimization*, 24:1031–1043, 1986.

[35] R. Gibson and E. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3):959–976, 1990.

[36] C. L. Gilbert. Commodity speculation and commodity investment. *FAO Commodity Market Review 2009-2010*, pages 26–46, 2010.

[37] G. Gorton and K. G. Rouwenhorst. Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62(2):47–68, 2006.

[38] M. Jr. Grendar and M. Grendar. Randomness as an equilibrium: otential and probability density. In *Maximum Entropy and Bayesian Methods: Proceedings of the 21-st International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis, Baltimore, Maryland, 4-9 August 2001*, pages 405–410, Melville, New York, 2002. American Institute of Physics.

[39] M. R. Gupta and Y. Chen. Theory and use of the em algorithm. *Foundations and Trends in Signal Processing*, 4(3):223–296, 2011.

[40] R. Gustafson. Carryover levels for grains: A method for determining amounts that are optimal under specified conditions. *USDA Technical Bulletin 1178*, 1958.

[41] G. J. Hahn. Fitting regression models with no intercept term. *Journal of Quality Technology*, 9(2):56–61, 1977.

[42] J. D. Hamilton. A new approach to the economic analysis of non-stationary time series and the bussiness cycle. *Econometrica*, 57(2):357–384, 1989.

[43] J. D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

[44] A. Heap. China - the engine of a commodities super cycle. *Citygroup Global Markets Inc.*, 2005.

[45] M. Henn. The speculator's bread: what is behind rising food prices. *Outlook, EMBO Reports*, 2011.

[46] M. W. Hirsch. The dynamical systems approach to differential equations. *Bulletin of the American Mathematical Society*, 11(1):1–64, 1984.

[47] M. W. Hirsch, S. Smale, and R. L. Devaney. *Differential Equation, Dynamical Systems and An Introduction to Chaos*. Elsevier, Amsterdam, 2004.

[48] C. H. Hommes. Financial markets as noninear adaptive evolutionary systems. *Quantitative Finance*, 1:149–167, 2001.

[49] H. Hoteling. The economics of exhaustible resources. *The Journal of Political Economy*, 39(2):137–175, 1931.

[50] IMF. *World Economic Outlook*. Washington, DC, 2009.

[51] S. Irwin and D. Sanders. Index funds, financialization, and commodity futures markets. *Applied Economic Perspectives and Policy*, 33(1):1–31, 2011.

[52] E. T. Jaynes and R. D. Rosenkrantz. *E.T. Jaynes: Papers on Probability, Statistics, and Statistical Physics / edited by R. D. Rosenkrantz*. D. Reidel Publishing Co., Dordrecht, 1983.

[53] N. Kaldor. Speculation and economic stability. *Review of Economic Studies*, 7:1–27, 1933.

[54] R. Kaplinsky. Asian drivers, commodities and the terms of trade. 2010.

[55] R. K. Kaufmann. The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy*, 39(1):105–115, 2011.

[56] M. Kendall and A. Stuart. *Inference and Relationship*. Charles Griffin, London, 1961.

[57] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[58] C. Kwon, P. Ao, and D. J. Thouless. Structure of stochastic dynamics near fixed points. In *Proceedings of The National Academy of Sciences of the USA*, volume 102, pages 13029–13033, 2005.

[59] M. W. Masters and A. K. White. How institutional investors are driving up food and energy prices. *The accidental hunt brothers*, Special Report July 31, 2008.

[60] B. J. Matkovsky and Z. Schuss. Eigenvalues of the Fokker-Planck operator and the approach to equilibrium for diffusions in potential fields. 81(2):387–394, 1981.

[61] J. Mayer. The growing interdependence between financial and commodity markets. Discussion paper 195, 2009.

[62] G. J. McLachlan and T. Krishnan. *The EM Algorithm and its Extensions*. Wiley, 1997.

[63] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, 2000.

[64] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1983.

[65] M. Nissanke. Commodity markets and excess volatility: Sources and strategies to reduce adverse development implications. 2011. Paper prepared for the Common Fund for Commodities (CFC).

[66] T. Orchard and M. A. Woodbury. A missing information principle: theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 697–715, 1972.

[67] K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, A(185):71–110, 1894.

[68] D. Piccolo. *Statistica*. Il Mulino, 2010.

[69] R. Pindyk. Inventories and the short-run dynamics of commodity prices. *RAND Journal of Economics*, 25(1):141–159, 1994.

[70] R. Pindyk. The dynamics of commodity spot and futures markets: a primer. *The Energy Journal*, 22(3):1–29, 2001.

[71] C. Pirrong. *Commodity Price Dynamics: A Structural Approach*. Cambridge University Press, 2011.

[72] M. Radetzki. The anatomy of three commodity booms. *Resources Policy*, 31(1):56–64, 2006.

[73] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.

[74] D. Sanders and S. H. Irwin. A speculative bubble in commodity futures prices: cross-sectional evidence. *Agricultural Economics*, 41(1):25–32, 2010.

[75] J. Scheinkman and J. Schectman. A simple competitive model of production with storage. *Review of Economic Studies*, 50:427–441, 1983.

[76] E. Schwartz. The stochastic behaviour of commodity prices: implications for valuation and hedging. *Journal of Finance*, 53(3):923–973, 1997.

[77] J.C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley-Interscience, 2003.

[78] A. Stuart and J.K. Ord. *Kendall's Advanced Theory of Statistics, Vol. I: Distribution Theory*. Edward Arnold, 1994.

[79] M. Takayasu, T. Mizuno, and H. Takayasu. Potential force observed in market dynamics. *Physica A*, 370(9):91–97, 2006.

[80] C. M. Tan. *Simulated Annealing*. I-Tech Education and Publishing, Vienna, Austria, 2008.

[81] K. Tang and W. Xiong. Index investment and financialization of commodities. Working Paper 16385, National Bureau of Economic Research. Princeton University, Cambridge (Massachusetts), September 2010.

[82] J. R. Taylor. *Classical Mechanics*. University Science Books, Sausalito, 2004.

[83] L. Telser. Futures trading and the storage of cotton and wheat. *Journal of Political Economy*, 66:233–255, 1958.

[84] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2008.

[85] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, 1985.

[86] M. E. Turner. Straight line regression through the origin. *Biometrics*, 16(3):483–485, 1960.

[87] UNCTAD. *Price Formation in Finalcialized Commodity Markets: The role of Information*. Geneva, 2011.

[88] K. Watanabe, H. Takayasu, and M. Takayasu. Random walker in temporally deforming higher-order potential forces observed in a financial crisis. *Physical Review E*, 80, 2009.

[89] WB. *Global Economic Prospects: Commodities at the Crossroads*. 2009.

[90] W. F. R. Weldon. Certain correlated variations in *Crangon vulgaris*. *Proceedings of the Royal Society of London*, 51:2–21, 1892.

[91] W. F. R. Weldon. On certain correlated variations in *Carcinus moenas*. *Proceedings of the Royal Society of London*, 54:318–329, 1893.

[92] J. Williams and B. Wright. *Storage and Commodity Markets*. Cambridge University Press, Cambridge, 1991.

[93] H. Working. Price relations between july and september wheat futures in chicago since 1885. *Wheat Studies*, 9:186–238, 1933.

[94] H. Working. The theory of the price of storage. *American Economic Review*, 39:1254–1262, 1949.