POLITECNICO DI MILANO
DEPARTMENT OF MATHEMATICS
PH.D. COURSE IN
MATHEMATICAL MODELS AND METHODS IN ENGINEERING

A PARTICLE FILTER APPROACH TO PARAMETER ESTIMATION IN STOCHASTIC VOLATILITY MODELS WITH JUMPS FOR CRUDE OIL MARKET

Doctoral Dissertation of:
Gaetano Fileccia

Advisor:**Prof. Carlo Sgarra**

Tutor: **Prof. Emilio Barucci**

February 28th, 2012 – Cycle XXV

Contents

Introduction

This dissertation deals with the inference in a crude oil market, the West Texas Intermediate (WTI) crude oil, whose futures are quoted on the New York Mercantile Exchange Market (NYMEX). The oil is sold at Cushing, Oklahoma and the spot data are collected by the U.S. government website eia.com (whose acronym means energy information administration), where are reported the data regarding energy markets.
The purpose of the research underlying this dissertation is twofold: first, we were looking for an appropriate model for the WTI crude oil market. Second, we tried to develop a new technique for parameter estimation suitable for the models considered, allowing to get inference results when the price dynamics is described both under the historical and the risk-neutral probability measures.
Since the data available for our study come from the spot market and the futures prices market, the model choice will need to be performed according to satisfy the requirement to catch both the markets structures. Hence the choice to consider in models proposed a variance latent process (like in well known Heston model) besides the usual convenience yield process, typical of the storable commodity financial markets. After the financial theory, that traces the evolution of oil modeling, have been recalled, we proposed different models to be compared in catching the observed data time series. The different models are selected in order to resume the different possible modeling characteristics of commodity markets, as the presence of seasonality or the possibility to observe jumps in spot dynamics.
These models, always belonging to the family of the affine models, to ensure admitting the affine futures price closed formula, have been tested and compared. Since the data comes from two different financial world, the spot market and the futures market, we needed to analyze our model for two different measures, under which we observed the two different sources of data: the historical measure

for spot data and the risk neutral measure for futures prices. A definition of risk premia to allow switching from one measure to another (according to Girsanov theorem) have been provided besides the description of any model proposed.
To conduct the inference we resorted to the Particle Filter algorithms, since this family of inference tools have demonstrated to be particularly suitable when non gaussian or non linear latent processes are included in dynamics. Hence, in our research study, they are preferable to other techniques since the presence of not linear terms (and, moreover, the eventual presence of jumps processes in our models). The algorithm family is thoroughly discussed, and main alternatives presented in the Chapter 2. Then it is discussed and implemented the specific algorithm; more specifically, the algorithm engine to carry out the inference comes from a recent evolution in the Markov chain Monte Carlo (MCMC) and link together the two bayesian techniques which had greatest relevance in the recent stochastic volatility econometrics literature: MCMC and Particle Filters. This algorithms, introduced by Andrieu Doucet and Holenstein [10] show particularly good behaviour when Stochastic volatility models (eventually with jumps) are analyzed, since the flexible structure of Particle Filters well fits the need to filter the non-linear (and in some cases non-gaussian) latent processes and the MCMC gives a well-known and tested technique to sample from the distribution of parameter set of the model.
Finally, we analyzed the performances of the model studying both the ability of the model in properly representing the data analyzed (by studying the property of the residuals got in our model inference) and the error produced in catching the futures prices curve for different maturities. Model including jumps showed the best performances in both the two analysis, among the models considered.
The thesis is divided in two parts: the first part deals with the market, the models and the algorithms to get inference. In the second part two articles are reported in which the models proposed to describe the data for WTI spot and WTI futures are analyzed.
In the first part the analyzed market and the main theories to describe it are presented. The attention is focused on reduced form models since in the second part we dealt with this kind of model. In the first chapter, it is also explained how prices for futures and options on futures can be computed when the model belongs to the affine family.
In the second chapter the topic moves from models to the description of algorithms to get inference. Obviously, significant attention is devoted to the algorithms related to the one used in our inferences studies, whose results are reported in the second part of the thesis. Since the algorithms belongs to Markov Chain Monte Carlo and Particle Filter algorithms, the Monte Carlo techniques are introduced together with a short survey of bayesian statistics, hence the families of algorithms used are deeply discussed and pseudo-codes provided. At the end of the chapter are discussed the main alternatives used in literature to tackle the

same inference problem we dealt with.
The second part consists in three chapters: the third chapter and the fourth one are the two articles resulting from the research work underlying this thesis, while the last chapter reports the concluding remarks and describes the interesting research lines to be pursued in the future.
In the third chapter three models with both stochastic representation for convenience yield and volatility dynamics are introduced, restricting to the case of spot process correlated just with volatility. Estimation inference is conducted using Particle Markov Chain Monte Carlo Methods. The first model introduced is a stochastic volatility model in which both convenience yield and volatility process are modeled by CIR dynamics; the other two models are variants of the first one: one includes Merton jumps in spot dynamics, while the second one allows for seasonality factor. In the fifth chapter two more models are introduced (which differ among themselves for the presence of jump activity in the spot) that allow also for correlation among convenience yield and spot dynamics. The two model performances are compared with the performances of the model proposed by Liu and Tang in [39], that we considered as the benchmark model for this commodity market. In both cases analyzed, with and without taking into account the correlation of the spot process with the convenience yield process, it is shown that models with volatility process and jumps overperform the other possible models (also the benchmark one) in catching the dynamics of the observed data analyzed.
In the fifth chapter some concluding remarks are provided and possible future developments of this work are outlined.

CHAPTER 1

Financial Models for Oil Market

1.1 Introduction

In this chapter it is presented the market, object of the analysis in the second part of this thesis: data, structure of the market and models to describe them are introduced. In the first section are outlined the main references in the literature, tackling the modeling issue for storable commodity markets and it is presented a quick description of the evolution of the theory of inventory, that solved the puzzle of contango and backwardation regimes. In the second section it is provided a survey of the main models for this specific market in the family of reduced form models; these models constitute the background for the models proposed and discussed in the second part of the thesis. In the last section, the tools to get prices for derivatives (when the model is affine) are presented.

1.2 The crude oil market

Among commodities we can recognize at least two main families: storable and non storable commodities, oil clearly belongs to the first family. Since our goal is describing the crude oil market, and in particular the WTI[1] market, we restricted

[1] West Texas Intermediate: the crude stream produced in Texas and southern Oklahoma which serves as a reference or "marker" for pricing a number of other crude streams and which is traded in the spot market at Cushing, Oklahoma (definition from US government website eia.gov)

our study to the storable commodity family. The description of these markets coincides in the financial literature with the theory of storage and inventories. The inventories are the reserve quantities for a specific commodity and the goal of this branch of financial literature is to study how the inventory levels affect the supply-demand equilibria and how these equilibria influence the futures[2] market. These relationships can be summarized by a return on storage, that could assume both positive and negative values. The main issue faced in the early financial commodity literature was the attempt to explain the curve of futures for different maturities: this curve can exhibit a peculiar behaviour (with respect to other asset class markets). Two futures markets settings are possible: decreasing futures prices for longer time to maturity, with the futures prices below the spot price level (this state is referred as inverted future market or backwardation regime) or the opposite setting: increasing futures prices for longer maturities and prices above the corresponding spot prices, this state is indicated as normal future market structure or contango regime (see Pindyck [47] or Liu and Tang [39]). This two possible states are not explainable with any observable variable in the market, hence the difficulties in history of commodities financial literature to model this market catching this peculiar dynamics. An extended dissertation of these arguments can be found in [24] or [28].
The theory of storage was outlined for the first time by Working in 1949 in [57] analyzing the wheat market: the quotes for the futures prices are described in terms of the "price of storage" and return required by the suppliers to store the commodity till delivery; the article discussed thoroughly the case of positive returns for storage and observed the possibility for negative returns associated to the cases of large stored stocks. The cases of negative prices of storage are the main focus of [58] by Wright and Williams, where it was proposed a model that allows for physical transformation of one product in another (like in a refining process) and it was shown that it is possible to allow for both positive and negative return for storage, with the simple representation of two object (where the second one is a refined product of the first one), even omitting the cost for storing products. This return for storage is synthesized in a convenience yield: the yield gives a measure of how profitable is owning the commodity, and can be thought, as in [47] with the expected flow of benefits for the holder of commodity inventory, benefits arising from the opportunity to use inventories to reduce production and marketing costs, and in case of refinable products from the opportunity to transform the product.
Pindyck in 1994 [46] discussed the convenience yield and described it as a discount factor that includes both the benefits and the costs due to storing process, and introduced, by no-arbitrage argument, in a seminal paper, the representation

[2] A futures contract is an agreement that postpone the delivery of the commodity in the future, within a short period after the expiry of the contract. The contract is marked to market and it is quoted on a regulate market.

of the convenience yield in terms of marginal value of storage defined by the difference between spot and futures prices observed at a given time. The marginal value of storage $\psi_{t,T}$, following [46], is:

$$\psi_{t,T} = (1 + r_{t,T})S_t - F_{t,T} \tag{1.1}$$

where the marginal value of storage, at time t, for the maturity T is defined by the spread between spot value S_t (capitalized at time T by the risk free factor $r_{t,T}$) and the futures prices quoted at time t with maturity T; the marginal value of storage was further analyzed and discussed, after the possible unpredictable revenues from storage have been separated from the costs, considered as fixed. This definition of the value of storage has been the background for the convenience yield modeling in the successive literature.

1.3 Models in crude oil market

Till the work by Pindyck [46], the model used to describe the spot price process S_t for the commodity prices was usually a geometric Brownian motion, as in Black Scholes model, in presence of a continuous discount yield(that, in this market, represents the convenience yield δ), that was a constant or a deterministic function of time:

$$dS_t = (r - \delta)S_t dt + \sigma S_t dW_t \tag{1.2}$$

The paper who moved on from these simplified assumptions and modeled the convenience yield as a stochastic process has been published by Gibson and Schwartz in 1997 [30], opening a whole literature sector of financial economics focusing on the reduced form models[3] to describe the commodity market. Their model assumed that the convenience yield δ_t is ruled by Ornstein Uhlenbeck process:

$$\begin{cases} dS_t &= (\mu - \delta_t)S_t dt + \sigma S_t dW_t \\ d\delta_t &= \lambda(\bar{\delta} - \delta_t)dt + \sigma_2 dW_t \\ & dW_1 dW_2 = \rho dt \end{cases} \tag{1.3}$$

where the spot process S_t, is described by a dynamics similar to Black Scholes: S_t is log-normally distributed conditionally to a given value for convenience yield δ_{t^-} (here $t^- = lim_{\epsilon \to 0} t - \epsilon$). The process δ_t is mean-reverting with a long run yield parameter $\bar{\delta}$ and mean-reverting speed λ. The two stochastic processes are

[3] reduced form models describe the futures curve dynamics modeling the dynamics of the spot process and, eventually, of latent processes, that affect the dynamics of spot and futures prices, but are not directly observable. The main alternatives to this model family is the market models family, for a description of this models in commodity market it is possible to refer to Eydeland and Wolyniec [**?**]

correlated by a constant parameter ρ. The dynamical system (1.3) is the expression of the model under the historical measure[4]. Using a constant risk premium definition, Gibson and Schwartz [30] moved to a descriptive model under the risk neutral measure:

$$\begin{cases} dS_t &= (r - \delta_t)S_t dt + \sigma S_t dW_t \\ d\delta_t &= [\lambda(\bar{\delta} - \delta_t) - \eta]dt + \sigma_2 dW_t \\ & dW_1 dW_2 = \rho dt \end{cases} \tag{1.4}$$

where η is the risk premia in the convenience yield process.
This model resume some of the peculiarities of the storable commodities market noticed by Pindyck in [47]:

- the convenience yield is strongly relevant
- both the two regimes (contango and backwardation) are possible
- the convenience yield fluctuates considerably over time
- convenience yield and spot price process are positively correlated

The model (1.3) was extended in [53]: in this article, Schwartz compared the performances in catching the futures curve for three different models: the standard model (1.2), the model (1.4) and a new model, based on (1.4), extended to incorporate a third factor to describe stochastic interest rates. Under the risk neutral measure:

$$\begin{cases} dS_t &= (r_t - \delta_t)S_t dt + \sigma S_t dW_1 \\ d\delta_t &= [\lambda(\bar{\delta} - \delta_t) - \eta]dt + \sigma_2 dW_2 \\ dr_t &= a(\bar{r} - r_t)dt + \sigma_3 dW_3 \\ & dW_1 dW_2 = \rho_1 dt \\ & dW_2 dW_3 = \rho_2 dt \\ & dW_1 dW_3 = \rho_3 dt \end{cases} \tag{1.5}$$

Schwartz's analysis [53] showed that the performances of the models including a stochastic representation for the convenience yield improved significantly the performances in describing the futures market (results were compared on the basis of the rooted mean square error) with respect to the model (1.2). The two models with stochastic convenience yield showed similar results, and the best performing model varied from case to case, showing significant differences in

[4]Historical measure is the probability measure under which are observed the spot prices; the futures prices, instead, are observed under a risk neutral probability measure. It is possible to move from one measure to another defining risk premia for the spot market. A dissertation of how it is possible to link the two measure via the risk premia can be find in almost any introductive financial mathematics book, as [54]

the estimated future curves only when futures with maturity beyond the 7 years were analyzed.
These models have been the main reference models in literature for the family of reduced form, till now. Some variants have been studied and tested, one of them was the possibility to include a seasonality term in the dynamics, as proposed by Hikspoors in [36]. Another interesting variant was proposed by Ribeiro and Hodges in [50]: their model separates the two contributes in the convenience yield, the positive return associated to the physical owning of the commodity and the costs, due mainly to storage and insurance, hence associated a fixed parameter, since their fluctuation are much less frequent than the ones associated to positive returns (the separation is similar to the one proposed by Pindyck [46]):

$$\begin{cases} dS_t &= (r + c - \delta_t)S_t dt + \sigma\sqrt{\delta_t}S_t dW_t \\ d\delta_t &= [\lambda(\bar{\delta} - \delta_t) - \eta]dt + \sigma_2\sqrt{\delta_t}dW_t \\ & dW_1 dW_2 = \rho dt \end{cases} \tag{1.6}$$

The convenience yield in this model is given by $\delta_t - c$ and the net convenience yield δ_t, separated from the cost of storage c, can assume just positive values, hence it has been modeled with a CIR dynamics, that cannot assume any negative values[5].
This structure in the dynamics is similar to the model proposed by Liu and Tang in [39]. Liu and Tang started their work from some econometrics results: the influence of the convenience yield risk premia in the futures risk premium discussed by Gorton, Hayashi and Rouwenhorst [33], and the heteroskedasticity of the convenience yield process (discussed and shown in the same article). After an introductive discussion on the effect of heteroskedastic behaviour of the convenience yield, it is introduced a three factor model that includes both stochastic convenience yield and stochastic interest rates:

$$\begin{cases} dx_t &= [r + c - \delta_t - \frac{1}{2}(\sigma_{x\delta}^2\delta_t + \sigma_{xr}^2 r_t + V_0 + v_{x\delta}\delta_t + v_{xr}r_t)]dt + \\ & \sigma_{x\delta}\sqrt{\delta_t}dW_\delta + \sigma_{xr}\sqrt{r_t}dW_r + \sqrt{V_0 + v_{x\delta}\delta_t + v_{xr}r_t}dW_x \\ d\delta_t &= \lambda_\delta(\bar{\delta} - \delta_t)dt + \sigma_\delta\sqrt{\delta_t}dW_\delta \\ dr_t &= \lambda_r(\bar{r} - r_t)dt + \sigma_r\sqrt{r_t}dW_r \end{cases} \tag{1.7}$$

where the x_t represents the log-spot process. The model (1.7) allows for correlation between the log-spot process and convenience yield, it is stochastically, depending on the state of the convenience yield. The system (1.7) is under the risk neutral measure; differently from the previous model presented, Liu and Tang proposed to define the risk premia not independently from the processes. Following the work by Dai and Singleton [18], they proposed risk premia are proportional to the square root of the variables, preserving in this way the structure of

[5]When the condition $2\lambda\alpha > \sigma^2$ is met the CIR process can assume just positive values

of the dynamics of the model in (1.7) and allowing the dependence of the futures risk premia on the convenience yield risk premia. Under the historical measure, the (1.7) becomes:

$$\begin{cases} dx_t &= \left[r + c - \delta_t - \frac{1}{2}(\sigma_{x\delta}^2 \delta_t + \sigma_{xr}^2 r_t + V_0 + v_{x\delta}\delta_t + v_{xr} r_t) \right. \\ &\quad \left. + \frac{\eta_\delta \sigma_{x\delta}}{\sigma_\delta}\delta_t + \frac{\eta_r \sigma_{xr}}{\sigma_r} r_t + \eta_x (V_0 + v_{x\delta}\delta_t + v_{xr} r_t) \right] dt \\ &\quad \sigma_{x\delta}\sqrt{\delta_t} dW_\delta + \sigma_{xr}\sqrt{r_t} dW_r + \sqrt{V_0 + v_{x\delta}\delta_t + v_{xr} r_t} dW_x \\ d\delta_t &= [\lambda_\delta(\bar{\delta} - \delta_t) + \eta_\delta \delta_t]dt + \sigma_\delta \sqrt{\delta_t} dW_\delta \\ dr_t &= [\lambda_r(\bar{r} - r_t) + \eta_r r_t]dt + \sigma_r \sqrt{r_t} dW_r \end{cases} \tag{1.8}$$

where $[\eta_\delta, \eta_r, \eta_x]$ are the three risk premia factor. The model under both the measure are analyzed in [39] and the parameter set estimated for oil and copper markets. Both the markets showed similar estimation results, in particular for the oil market, Liu and Tang found that the correlation of the interest rates with log-spot process is not significative for the data analyzed, while the heteroskedastic parameter is significatively positive, confirming the initial modeling choice of using a CIR process to model the convenience yield process.
Differently from the previous model discussed, the one proposed by [59] started from observation the skews and smiles in option on futures market, instead of starting from the observation of the future market structure. This results in proposing a different reduced form model, that includes an additional latent process to the usual dynamics for spot process: the volatility process V_t. His description is restricted to the risk neutral measure case (since Yan is interested in describing just futures and options of futures):

$$\begin{cases} \frac{dS_t}{S_t} &= [r - \delta_t - \lambda_J \mu_J^*]dt + \sigma dW_1 + \sqrt{V_t} dW_2 + Jdq \\ d\delta_t &= \lambda_\delta(\bar{\delta} - \delta_t)dt + \sigma_\delta dW_\delta \\ dr_t &= \lambda_r(\bar{r} - r_t)dt + \sigma_r \sqrt{r_t} dW_r \\ dV_t &= \lambda_V(\bar{V} - V_t)dt + \sigma_V \sqrt{V_t} dW_V + J_V dq \\ & dW_2 dW_V = \rho_V dt \\ & dW_1 dW_\delta = \rho_\delta dt \end{cases} \tag{1.9}$$

The use of two Wiener processes in the spot dynamics is needed to keep the affine structure of the model, allowing to find a closed formula for the pricing problem of futures and options on futures contracts. The process Jdq is a jump process (whose compensator is $\lambda_J \mu_J^*$), with exponentially distributed number of jump arrivals in the unit of time and exponentially distributed jump size. The volatility process and spot process are restricted to jump together.

1.4 Appendix: Futures and Options on Futures

Provided that the model has an affine structure (see Duffie, Pan and Singleton [17] for a discussion on the affine structure of models including jumps and Grzelak [34] for the limits in the correlation structure for multivariate process to keep the model affine) the future prices can be found as the solution of the backward Kolmogorov equation (KBE).
If X is the vector of the processes, such that each component follows the dynamics

$$dX_i = \mu_i(X)dt + \Sigma_i(X)dW_t^{(i)}$$

with Wiener processes, driving the stochastic behaviour, possibly correlated as:

$$dW_t^{(i)}dW_t^{(j)} = \rho_{ij}dt$$

Hence, the Feynman-Kac theorem [54] leads to the backward Kolmogorov equation:

$$\frac{\partial f}{\partial t} + \sum_i \left[\frac{\partial f}{\partial X_i}\mu_i(X) + \frac{1}{2}\frac{\partial^2 f}{\partial^2 X_i}\Sigma_i(X) + \sum_{i<j} \frac{\partial^2 f}{\partial X_i \partial X_i}\sqrt{\Sigma_i(X)\Sigma_j(X)}\rho_{i,j} \right] = 0 \tag{1.10}$$

subject to the terminal condition $f(T,T) = S(T)$. The solution can be find imposing the affine structure for the solution (the logarithm of the solution is a linear combination of the processes involved in the dynamics). For any contingent claim the KBE is a slight modification of (1.10):

$$\frac{\partial f}{\partial t} + \sum_i \left[\frac{\partial f}{\partial X_i}\mu_i(X_i) + \frac{1}{2}\frac{\partial^2 f}{\partial^2 X_i}\Sigma_i(X) + \sum_{i<j} \frac{\partial^2 f}{\partial X_i \partial X_i}\sqrt{\Sigma_i(X)\Sigma_j(X)}\rho_{i,j} \right] = rf \tag{1.11}$$

To find the prices for European call option on futures it is possible to follow the Bakshi and Madan method [9], adapted by Yan [59] to the case of European option on futures. The following methods is from [59]. If we denote by $C(t,\tau)$ the price the European option with time to maturity τ and strike K on a futures contract with maturity $\tilde{\tau} > \tau$:

$$C(t,\tau) = \mathbb{E}\left[\exp\left\{ -\int_t^{t+\tau} r(s)ds \right\} \max\{H(t+\tau, \tilde{\tau}-\tau) - K, 0\} \right] \tag{1.12}$$

Hence by [9] (explicit computations are in [59]) it is possible to decompose the solution:

$$C(t,\tau) = G(t,\tau)\Pi_1(t,\tau) + KB(t,\tau)\Pi_2(t,\tau)$$

with:

$$\begin{cases} B(t,\tau) & = f(t,\tau,0) \\ G(t,\tau) & = f\left(t,\tau,\frac{1}{i}\right) \\ f_1(t,\tau,\phi) & = \frac{1}{G(t,\tau)} f\left(t,\tau,\frac{1}{i}+\phi\right) \\ f_2(t,\tau,\phi) & = \frac{1}{B(t,\tau)} f\left(t,\tau,\phi\right) \end{cases}$$

where f_1 and f_2 are the characteristic function of Π_1 and Π_2 and $f(t,\tau,\phi)$ is the discounted characteristic function of logarithm of futures prices:

$$f(t,\tau,\phi) = \mathbb{E}\left[\exp\left\{-\int_t^{t+\tau} r(s)ds\right\}\exp\left\{i\phi \ln H(t,\tilde{\tau}\right\}\right] \tag{1.13}$$

where $f(t,\tau,\phi)$ solves the (1.11) with terminal condition

$$f(t+\tau,0,\phi) = \exp\{i\phi \ln H(t+\tau,\tilde{\tau}-\tau)\}$$

Hence, it is possible to find the price for european call options on futures, solving the 1.11

1.5 Appendix: the available data

The crude oil market analyzed in the second part of the thesis is the WTI market, the oil extracted by North America oil companies and traded on the NYMEX[6] market. On the NYMEX the only quoted contracts are the futures, the other contracts are traded over the counter and are less liquid financial instruments. For this reason, and since the lack of available quotes for spot market in literature to approximate the spot values in was used a rolling argument, using the future with shortest maturities available to approximate spot quotes (this is also the quotes published by Bloomberg platform) as it has been done by Schwartz [53][7]. Spot quotations used in this thesis comes from the US governmental site eia.gov[8], where are collected data from Reuters provider and from the spot market at Cushing in Oklahoma. These quotes are used as index for the futures quotations (and are, obviously, strongly correlated with spot quotes derived by a rolling process, as Bloomberg quotes).

[6] New York Mercantile Exchange Market

[7] An other possibility could be considering the spot prices as a latent process to be estimate from the model, observing just the futures prices

[8] Energy Information Administration

CHAPTER 2

Particle Filters

In this chapter it is presented a review of the algorithms used for the inference with the models presented in the Chapter 1. Inference with these techniques is carried out in the second part of the thesis. At the beginning of this chapter it is included an essential introduction to Monte Carlo integration and Bayesian inference. Then the algorithms are introduced. The main families of these algorithms are the Markov chain Monte Carlo and the Particle filters techniques. Both the techniques are presented in single sections devoted to them. Then it is discussed the simultaneous inference issue on the model parameters and the latent processes, and Particle Markov chain Monte Carlo are presented (the technique, based on MCMC and PF, it is the one used in the second part of the thesis). Last section is devoted to shortly resume the main alternatives to Monte Carlo estimates.

2.1 A short introduction to Monte Carlo algorithms

Monte Carlo methods were born as integration algorithms based on analogy between probability and volumes [32]. Basically, Monte Carlo methods allow to compute integrals

$$I = \int_{\alpha}^{\beta} f(x)dx$$

computing the expectations $\mathbb{E}[f(U)(\beta-\alpha)]$ with $U \sim \text{Unif}(\alpha,\beta)$. The core of the algorithm is based on the law of large numbers, and the central limit theorem. It is possible to build an estimator of the value of the integral, $\hat{I}$, relying on a sample $\{x_1, x_2, \ldots, x_n\}$ extracted from the uniform distribution (that is x_i are extractions independent from the uniform distribution).

$$\hat{I}_n = \frac{\beta-\alpha}{n}\sum_{i=1}^{n} f(x_i)$$

If function $f(x)$ is integrable in $[\alpha,\beta]$, by the law of large numbers, it follows that [32]:

$$\hat{I}_n \to I \text{ with probability } 1 \text{ as } n \to \infty$$

Moreover, if $f(x)$ is square integrable: in formulae it exists the integral

$$\sigma_f^2 = \int_\alpha^\beta (f(x)-I)^2 dx$$

then [32], asymptotically, $\hat{I}_n - I \sim \mathcal{N}\left(0, \frac{\sigma_f^2}{n}\right)$. Since σ_f^2 is usually unknown it is replaced by its sample estimate:

$$s_f^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(f(x_i)-\hat{I}_n\right)^2$$

It is remarkable that the error of the Monte Carlo estimate scales as the square root of n, that is the convergence rate of the estimate is $\mathcal{O}\left(n^{\frac{1}{2}}\right)$. This means that other quadrature algorithms (like trapezoidal rule) converges much more faster, when we compute integrals in one dimension; Monte Carlo becomes a much more useful technique when we move to multidimensional integration, since the estimation error does not depend on the dimension of the integral (instead of other integration algorithm families). This makes Monte Carlo so popular in scientific sector where state space dimension is usually high, like in finance.

2.1.1 Variance reduction: Importance sampling

Monte Carlo goal is to get an estimate for the expectation value of a function of a random value X (with density function $p(x)$):

$$I = \mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)p(x)dx \tag{2.1}$$

by a sample $\{x_i\}_{i=1,\ldots,N}$ drawn independently and uniformly from the distribution X:

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} f(x_i)\delta(x-x_i) \tag{2.2}$$

As already stated, the convergence rate is fixed, but there exist some techniques to reduce the estimator variance by a factor, increasing the precision of the estimate keeping fixed the largeness of the sample. The three main technique are stratified sampling, control variates and important sampling.
Stratified sampling suggests to divide the sample space in non-overlapping subsets $\{A_i\}_{i=1,\dots,k}$ such that $\mathbb{P}(x \in \cup_i A_i) = 1$. Then it is possible to prove [32] the estimator is still unbiased

$$\mathbb{E}[f(X)] = \sum_{i=1}^{k} \mathbb{P}(Y \in A_i)\mathbb{E}[X|Y \in A_i]$$

and we estimate the integral (2.1) getting Monte Carlo estimates for each $\mathbb{E}[X|Y \in A_i]$.
A particular case of stratified sampling is antithetic sampling (with two strata with equal probabilities).
Control variates is a technique which pairs up a correlated random variables Y to the r.v. X, whose we know the mean μ_Y. According to this technique, we sample together the two random variables and use what we know about Y to minimize the error in the estimate on $\mathbb{E}[X]$:

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \lambda(Y_i - \mu_y))$$

This estimator is sill unbiased [32]. An optimal choice for λ is $\lambda = -\dfrac{\mathbb{C}\text{ov}(X,Y)}{\mathbb{V}\text{ar}(Y)}$: effectively, we are reducing the variance of the estimator on the expected value of X by the part of variance that is possible to explain with Y (like in linear regression), using the information available on Y (that is μ_Y).
Important sampling allows to sample from a different distribution with respect to the one we want to integrate. As always, our final goal is to estimate $\mathbb{E}[f(X)]$ Instead of making use of (2.2) we sample independently and uniformly from a distribution Y (with density function $q(y)$) whose support includes the support of the random variable X:

$$I = \mathbb{E}[f(X)] = \int_R f(y)\frac{p(y)}{q(y)}q(y)dy$$

Hence the (2.2) becomes [32]:

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} f(y_i)\frac{p(y_i)}{q(y_i)}\delta_y(y_i)$$

$h(y) = \dfrac{p(y)}{q(y)}$ is called importance weight function and $\dfrac{p(y_i)}{q(y_i)}$ importance weights.
The estimator is still unbiased under the hypothesis that $\text{supp}(X) \subseteq \text{supp}(Y)$ (we

are just computing the integral by substituting the variable). Since we are free in choosing the importance function $q(y)$ it is critic to choose an appropriate important function. The following example clearly shows how delicate is the choice of the importance function. Let us consider the simple case in which we want to estimate the mean of $X \sim \mathbb{T}(\nu = 12)$ using two important functions: a gaussian distribution (with zero mean and variance $\frac{\nu}{\nu-2}$) and a Cauchy distribution. In Figure2.4 are shown the probability density functions of the three distribution

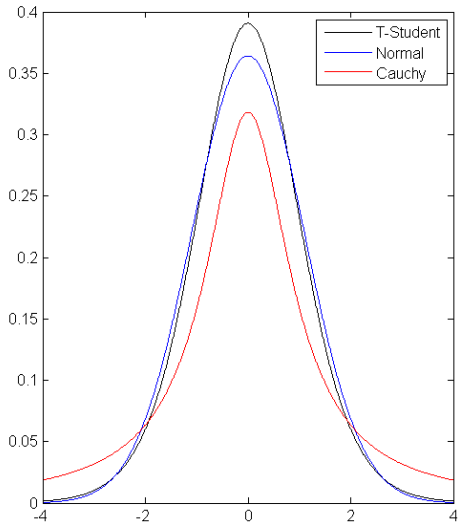

Figure 2.1: *probability density functions for three distribution: T-student (with parameter $\nu = 12$), a gaussian (with zero mean and variance equal to $\frac{\nu}{\nu-2}$) and a Cauchy distribution.*

Even if the normal density is much more close to the t-student density than the Cauchy distribution, it is remarkable that extreme events have higher probability in t-student random variable than in normal r.v.. This potentially leads to highly inhomogeneous importance weights when we run a MC simulation. In Figure2.2 and in Figure2.3 there is an example with two MC runs (each one with 10000 sampled values) for the estimation of the mean of the t-student r.v.. For the first run we drew the values form the importance normal r.v. $\mathbb{N}\left(0, \frac{\nu}{\nu-2}\right)$, in the second the values are drawn from a Cauchy distribution. In Fig.2.2 for each value sampled it is represented the importance weight value with a bar.

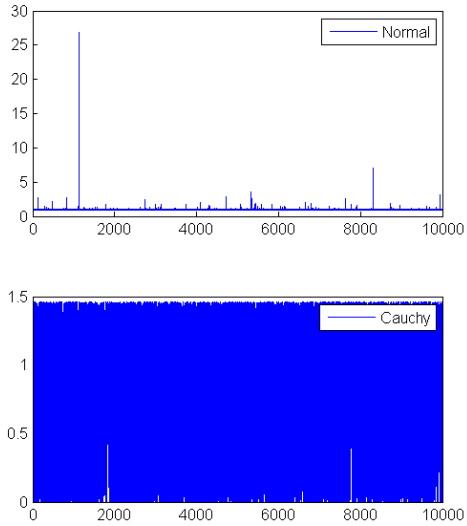

Figure 2.2: *importance weights for a sample generate from normal distribution and from Cauchy distribution to estimate the expected value of a random variable distributed accordingly to a t-student distribution.*

It is evident the lack of homogeneity when we use the normal r.v.; this lack in homogeneity in weights values (and in particular extremely high values in importance weights) makes the estimation via importance sampling ineffective, as it is possible to see in Fig.2.3, where are reported the estimates at each point in the sample. It is highlighted the first 20% of the sample to show clearly the effect of large important weights on the effectiveness of the estimate.

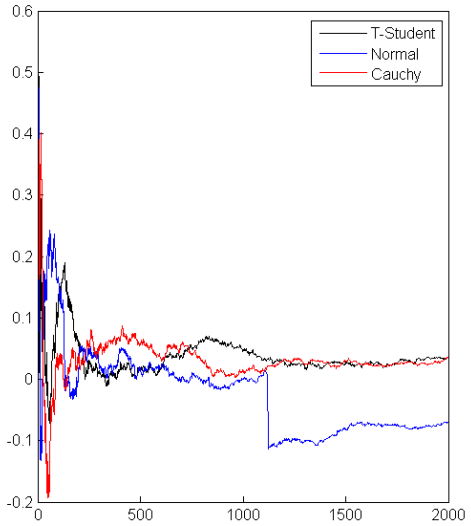

Figure 2.3: *Monte carlo importance sampling estimation for the expected value of a t-student distributed r.v. for different largeness of the sample. In black the Monte Carlo sampling directly from the t-student distribution, in red the importance sampling estimates obtained sampling from a cauchy distribution, in blue the estimates obtained sampling from the gaussian distribution*

To be effective the importance sampling technique requires need the importance function has fatter tails than the original distribution, or equivalently we need to ensure that $|w(y)| < M \quad \forall y \in \text{supp}(Y)$.
Importance sampling will be the core technique for the family of algorithm we will use in our inference (Particle filters), so here it has been shown that particular care has to be paid into the implementation of it.

2.2 Bayesian Inference and MCMC

Markov Chain Monte Carlo (MCMC) is a technique that in the last years had become one of the most used to get inference in time series analysis, in particular when Stochastic Volatility models are involved. The framework of these methods is the Bayesian Statistics.

2.2.1 A short introduction to Bayesian statistics

As stated by Gelman et al [3] bayesian inference is the process of fitting a probability model to a set of data and summarizing the results by a probability distributions on the parameters of the model and on unobserved quantities such as predictions for new observations.

Let us consider a model $\{p(Z|\theta), \theta \in \Theta\}$, a family of possible distribution for the random variable Z indexed by a parameter vector θ, that is the object of our inference. If we consider a subjective definition of probability of an event, that is, for an individual, his personal degree of belief on the event, then bayesian learning is a process which leads from prior knowledge (prior personal degree of belief), formulated before the data on the phenomenon has been collected, to posterior knowledge, including the information acquired by observed data. This knowledge it is reflected in defining probability density for the parameter θ. The fundamental idea of this branch of statistics is that, after seeing the experimental results, we update the the probability distribution we had subjectively specified, and use the Bayes' theorem to combine experimental results (the likelihood) and prior distribution; the result it is called posterior distribution:

$$\pi(\theta|z) = \frac{p(z|\theta)\pi(\theta)}{p(z)} \tag{2.3}$$

Where: $\pi(\theta|z)$ is the posterior density function we want to know (the probability distribution of θ given the knowledge on z, the observed realization of Z), $\pi(\theta)$ is the prior density function (that reflects our prior knowledge about the model and its parameter), $\pi(z|\theta)$ is the likelihood of the observation data z given the choice of parameter θ, and finally $p(z) = \int_\Theta p(z|\theta)\pi(\theta)d\theta$ is the a priori probability distribution of the observation chain.
The most difficult element to estimate in the ratio (2.3) is the a priori probability $p(z)$ since it would require the knowledge of the likelihood function for every point in the state space of the parameter vector, but, since it does not depend on θ, we can consider it like a normalizing constant and limit our study to consider

$$\pi(\theta|z) \propto p(z|\theta)\pi(\theta) \tag{2.4}$$

Moving from the prior $\pi(\theta)$ to the posterior $\pi(\theta|z)$ is what we gained, in knowledge, about the parameter after having taken into account the observation data z. Once acquired this knowledge, usually the posterior information is summarized in point estimates since these are easier to be communicated and handled, the most used point estimate is the mean of the posterior distribution (other popular choices are the median or the mode of the posterior distribution).
Since the priors are completely subjectively defined, it is one of the most delicate point in bayesian estimation to cope with. Usually priors reflects expert information about the model specification, if no reliable prior information is available about the model and we do not want to add information we do not rely on, it is possible to use non informative priors or improper priors. One of the most used improper prior is the uniform one

$$\pi(\theta) \propto k; k \in \mathbb{R}^+$$

another popular choice, called Jeffrey's prior, is to use a prior proportional to the sensitivity of the likelihood to the parameters using the information matrix H:

$$\pi(\theta) \propto \sqrt{\det H(\theta)}$$

If an improper prior is chosen it is to be checked that the posterior distribution is still a probability density; otherwise it is possible to define an uninformative prior simply choosing a distribution with very high variance, two popular choice are normal distribution (eventually truncated) with variance at least ten times greater than the usual scale order of the parameter, or uniform distributions, with support opportunely chosen to be spread enough.
Among Bayesian algorithms Markov Chain Monte Carlo (MCMC) are methods that allow for sampling from the posterior distribution avoiding the computation of $p(z)$ and using Monte Carlo technique to conduct the inference. The main algorithms are Metropolis-Hastings and Gibbs Sampling, even if they can be seen as one the particular case of the other, we present them separately, accordingly to the historical evolution of this subject and to make simpler the presentation of the algorithm in a easily implementable form.

2.3 MCMC

Markov Chain Monte Carlo are techniques which refers to Markov Chain theory to build an algorithm useful to sample from distribution we know up to a normalizing constant. Let consider a kernel density K. A Markov chain[1] is a sequence of a dependent random variables$\{\Theta_t; t \in \mathbb{N}\}$ such that the probability distribution of Θ_t given the past depends just on the previous variable Θ_{t-1}:

$$(\Theta_t|\Theta_0, \Theta_1, \ldots, \Theta_{t-1}) \sim K(\Theta_t|\Theta_{t-1})$$

where $K(\Theta_t|\Theta_{t-1})$ is a transition kernel: an application defined on $\Omega \times \mathcal{B}(\Omega)$ (where Ω is the state space of possible outcomes for Θ_i and $\mathcal{B}$ the Borel algebra) such that

$$\begin{cases} \forall \theta \in \times \;\; K(\theta, \cdot) \text{ is a probability measure} \\ \forall A \in \mathcal{B}(\times) \;\; K(\cdot, A) \text{ is measurable} \end{cases}$$

The starting point for building Metropolis-Hastings (MH) algorithm is the goal we want to achieve: our goal is sampling from the density function $\pi(\Theta|z)$, the posterior distribution of the parameter set given the observed data sequence, but as it has been already said, we don't know it in advance (since we don't know

[1] here we briefly present the theory the discrete time Markov chain, since it is useful to illustrate how the MCMC algorithm work. The references used for this section are [51] (where it is possible to find a thorough discussion about convergence theory for Markov chain) and [52] (where it is presented an implementable version of the main MCMC algorithms).

$p(z)$). For simplicity, we drop the dependence on the observed data sequence Z in the notation.
To sample from $\pi(\theta)$ $[= \pi(\theta|z)]$ we make the hypothesis that this unknown distribution is the stationary distribution probability of an invariant, irreducible and acyclic Markov Chain, where the states are the possible values for the parameters set. The stationarity means:

$$\int_{\Omega} K(\theta_1, \theta_2)\pi(\theta_1)d\theta_1 = \pi(\theta_2)$$

The existence of the stationarity distribution implies the requirement of irreducibility on K. The irreducibility means that for any set $A \in \mathcal{B}(\Omega)$ and any $\theta_0 \in \Omega$ there is a positive probability to move from θ_0 in A in a finite number of steps. A sufficient condition that ensures the irreducibility is that for any θ_0 $K(\theta_0, \cdot) > 0$ everywhere. The simplest way to impose the stationary distribution π is to construct the Markov Chain imposing it satisfies the detailed balance condition:

$$K(\theta_1, \theta_2)\pi(\theta_2) = K(\theta_2, \theta_1)\pi(\theta_1) \tag{2.5}$$

If the chain is irreducible and aperiodic (the property that it does not exists a cycle of states in which the chain could get struck) the Markov chain theory ensure us that, as the time step goes to infinity, the chain will converge towards a limiting distribution p that coincides with the stationary distribution π, independently from the choice of the starting point θ_0 (for almost any choice for θ_0). Hence what we need is to specify in an appropriate way the kernel K. In the Metropolis-Hastings algorithm it is obtained by the product of two factors: a proposal distribution $q(\theta_i|\theta_j) = q_{ij}$ (whom it is easy to sample from) and an acceptance probability α_{ij} that guarantees the condition (2.5); that is:

$$p_{ij} = q_{ij}\alpha_{ij}$$

where:

$$\alpha_{ij} = \min\left[\frac{\pi(\theta_j)}{\pi(\theta_i)}\frac{q_{ji}}{q_{ji}}, 1\right] \tag{2.6}$$

Hence, after a sufficient time $\bar{t}$, we can consider that the stationary distribution is reached, and the chain $\theta^{(\bar{t}+1)}, \ldots, \theta^{(T)}$ can be taken as a sample of the stationary distribution π. Considerations on tests to verify if the chain has reached the limiting distribution are discussed in the section about MCMC diagnostics: Sec 2.3.3. Once we get a sample $\{\Theta_i; i = 1, \ldots N\}$ from the stationary distribution π a law of large numbers for Markov chain ensures that, if $\mathbb{E}_\pi|f| < \infty$, then [51]:

$$\frac{1}{N}\sum_{i=1}^{N} f(\Theta_i) \to \mathbb{E}_\pi[f] \text{ almost surely} \tag{2.7}$$

The sample is not independent drawn but it exists a Central limit theorem if in addition to previous hypothesis, it is satisfied also that $\mathbb{E}_\pi[f^2] < \infty$ and the Markov chain is geometrically ergodic[2][51]:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} f(\Theta_i) - \mathbb{E}_\pi[f] \to \mathcal{N}(0, \tau_f^2) \tag{2.8}$$

where the convergence is in law and $\tau_f^2 = \sigma^2(1 + 2\sum_{k=1}^{\infty} \rho_k)$ and $\rho_k = \gamma_k/\sigma^2$ is the autocorrelation at the k-th lag: $\gamma_k = \mathbb{C}\text{ov}_\pi(f(\theta^n), f(\theta^{n+k}))$, while $\sigma^2 = \gamma_0$

2.3.1 Metropolis-Hastings

Metropolis Hastings is an algorithm to get a sample from the distribution π, building a Markov chain; to ensure that asymptotically the random variable are distributed according to the density π.
Since we have all the elements for the algorithm it is possible to itemize the steps of it. The starting point is a value θ_0 randomly chosen.

Algorithm 1 Metropolis Hastings (MH)

1: **for** t=1:M **do** ▷ M is the length of the chain
2: Sample a candidate $\theta^* \sim q(\Theta|\theta^{(t)})$
3: compute the acceptance ratio $\alpha(\theta^*|\theta^{(t)}) = \frac{\pi(\theta^*)}{\pi(\theta^{(t)})} \frac{q(\theta^*|\theta^{(t)})}{q(\theta^{(t)}|\theta^*)}$
4: accept the new candidate θ^* with probability $\alpha(\theta^*|\theta^{(t)})$ (then $\theta^{(t+1)} = \theta^*$), otherwise reject it and $\theta^{(t+1)} = \theta^{(t)}$
5: **end for**

Hence the Kernel is:

$$K(\theta_1, \theta_2) = q(\theta_2|\theta_1)\alpha(\theta_1, \theta_2) + \delta_{\theta_2}(\theta_1)\left(1 - \int_\Omega \alpha(\theta_1, \theta_2) q(\theta_2|\theta_1)\right)$$

[2] A MC is geometrically ergodic if it exists a constant $\lambda \in [0, 1[$ and an integrable real function $M(\theta)$ such that $||K^n(\theta, \cdot)\pi(\cdot)|| \leq M(\theta)\lambda^n$.

This kernel definition is such that it satisfies the detail balance condition (2.5) and has stationary distribution π:

$$
\begin{aligned}
K(\theta_1,\theta_2)\pi(\theta_1) &= q(\theta_2|\theta_1)\alpha(\theta_1,\theta_2)\pi(\theta_1) + \delta_{\theta_1}(\theta_2)\left(1-\int \alpha(\theta_1,\theta_2)q(\theta_2|\theta_1)d\theta_2\right)\pi(\theta_1) \\
&= q(\theta_2|\theta_1)\min\left\{\frac{\pi(\theta_2)q(\theta_1|\theta_2)}{\pi(\theta_1)q(\theta_2|\theta_1)},1\right\}\pi(\theta_1) + \delta_{\theta_1}(\theta_2)\left(1-\int \alpha(\theta_1,\theta_2)q(\theta_2|\theta_1)d\theta_2\right)\pi(\theta_2) \\
&= \min\{\pi(\theta_2)q(\theta_1|\theta_2), q(\theta_2|\theta_1)\pi(\theta_1)\} + \delta_{\theta_1}(\theta_2)\left(1-\int \alpha(\theta_1,\theta_2)q(\theta_2|\theta_1)d\theta_2\right)\pi(\theta_2) \\
&= q(\theta_1|\theta_2)\min\left\{1,\frac{\pi(\theta_1)q(\theta_2|\theta_1)}{\pi(\theta_2)q(\theta_1|\theta_2)}\right\}\pi(\theta_2) + \delta_{\theta_1}(\theta_2)\left(1-\int \alpha(\theta_1,\theta_2)q(\theta_2|\theta_1)d\theta_2\right)\pi(\theta_2) \\
&= q(\theta_1|\theta_2)\alpha(\theta_2,\theta_1)\pi(\theta_2) + \delta_{\theta_1}(\theta_2)\left(1-\int \alpha(\theta_2,\theta_1)q(\theta_1|\theta_2)d\theta_1\right)\pi(\theta_2) \\
&= K(\theta_2,\theta_1)\pi(\theta_2)
\end{aligned}
$$

Two usual choices for the proposal density $q(\theta_2|\theta_1)$ are:

- independent from the present value of the chain (known as independent Metropolis Hastings): $q(\theta_2|\theta_1) = q(\theta_2)$
- the new proposed value of the chain is sampled from a normal density with mean centered in the present value of it (known as random walk MH): $q(\theta_2|\theta_1) \sim \mathcal{N}(\theta_1,\sigma^2)$[3]

It is worth to notice that now we need to compute $\dfrac{\pi(\theta_2)}{\pi(\theta_1)}$ that is much more easy than computing directly the distribution π (since the difficulty in treating the a priori probability term in (2.3) vanishes because this term can be taken into account as a multiplicative constant that cancel out when the ratio is computed). Hence, after a sufficient long chain $\{\theta^{(t)}\}$ has been generated, we have to get rid of the first part of the chain, to avoid that our choice of θ_0 could influence the results, and perform appropriate diagnostics tests to assess the sanity of our sample. In the section 2.3.3 we deal more deeply with this topics.

2.3.2 Gibbs sampling

Till now, we have not been concerned with multivariate distributions, we treated the univariate case and the multivariate case with the same tools. This approach, in case of low-dimension distribution (can be translated in a few parameter to be

[3]The Random Walk Metropolis Hastings is the algorithm used in our inference in the second part of the thesis, it is worth to be remarked that this choice automatically satisfies all the required hypothesis discussed in the previous section for the convergence of the Markov Chain.

estimated), could be appropriate, but it turns into a very complicate one to be used, in case we want to manage high-dimension distributions.

An alternative tool to the Metropolis-Hastings, in multivariate case, is the Gibbs sampling, its usefulness comes from the possibility to sample sequentially from univariate distributions.
If $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_d)$ the main difference with respect to the previous algorithm is in the distribution used to sample:

$$q(\Theta_i|\Theta) = \pi(\Theta_i|\Theta_{-i}, Z) \tag{2.9}$$

that is, we sample from the full conditionals distribution, the conditional distribution of one component of the parameter random vector given everything else: "$-i$" means $\{j : j \neq i\}$.
Unlike in the MH algorithm, the proposal distribution, here, coincides with the distribution defined by the state-space equations, hence the acceptance ratio is always equal to one (we always accept the new sample element). The advantage of the Gibbs algorithm is to sample from univariate distributions, even if Θ is an r-dimension random vector.
As in the Metropolis-Hastings algorithm the starting point $\theta^{(0)}$ is randomly chosen.

Algorithm 2 Gibbs Sampling Algorithm (GS)

1: **for** t=1:M **do** ▷ M is the length of the chain
2: we sample $\theta_1^{(t)} \sim \pi(\Theta_1|\theta_2^{(t-1)}, \ldots, \theta_d^{(t-1)}, z)$
3: ...
4: we sample $\theta_i^{(t)} \sim \pi(\Theta_2|\theta_1^{(t)}, \ldots, \theta_{i-1}^{(t)}, \theta_{i+1}^{(t-1)}, \ldots, \theta_d^{(t-1)}, z)$
5: ...
6: $\theta_d^{(t)} \sim \pi(\Theta_2|\theta_1^{(t)}, \ldots, \theta_{d-1}^{(t)}, z)$
7: **end for**

One interesting application of the Gibbs sampling is in the case of latent process, in fact it has been used to get inference for models in finance with latent processes. In their work in 1994, Jacquier Polson and Rossi [20] showed that for even simple cases of models with latent processes where the errors are not gaussian or linear (they considered log-Arch models) these algorithms outperform significantly the two main rival algorithms: method of moments and method based on quasi-maximum likelihood (that is Extended Kalman Filter). Both these two algorithms are presented in the last section of the chapter. Since then, the econometrics literature became more and more interested in MCMC algorithms, and when the model is not gaussian and linear (in this case the optimal choice is the Kalman filter), MCMC became standard tools in getting inference for both parameters and latent processes dynamics. In the cases of continuous Stochastic Volatility models (discretized using Euler rule) Eraker Johannes and Polson in

[8] and Eraker in [22] and Yu, Li and Wells in [12], among others, used this algorithm get inference, under both the historical measure and getting an estimate also for the market risk premia using observed data coming from both the spot market and the derivative market.
Let us consider the case in which we want to sample from $\pi(\Theta, X|Z)$ where X is the latent process and Z the chain of the observations. Since $\pi(\Theta, X|Z) = \pi(\Theta|X, Z)\pi(X|\Theta, Z)$, if we know how to sample form the full conditional distributions we can define a Gibbs sampling algorithm to sample "asymptotically" from the required distribution. It is important to notice that whenever we cannot write down the full conditional distribution for a a parameter or it is not simple to sample from it, it is always possible to replace a Gibbs step with a MH step (these kind of algorithms are part of hybrid MCMC algorithm, and it exists a vast literature implementing them successfully in a lot of different cases): In [8], [22] and [12], the Gibbs sampling is applied to Stochastic Volatil-

Algorithm 3 Gibbs Sampling for SV models

for t=1:M **do** ▷ M is the length of the chain
  for j=1:d **do** ▷ d is the dimension of the parameter vector
    sample $\theta_j^{(t)} \sim \pi(\Theta_j|\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots \theta_d^{(t-1)}, x^{(t-1)}, z)$
  end for
  for j=1:d **do** ▷ d is the dimension of the parameter vector
    sample $x_j^{(t)} \sim \pi(X_j|x_1^{(t)}, \ldots, x_{j-1}^{(t)}, x_{j+1}^{(t-1)}, \ldots x_d^{(t-1)}, \theta^{(t-1)}, z)$
  end for
end for

ity models in finance, since we do not know the full conditional for $\theta_j^{(t)} \sim \pi(\Theta_j|\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots \theta_d^{(t-1)}, x^{(t-1)}, z)$ it is implemented a Metropolis-Hastings step to sample for the model parameters, while standard Gibbs sampling steps are used for the stochastic volatility process, since it is possible to write the full conditional distribution, once the continuous dynamics is discretized.

2.3.3 MCMC Diagnostics

It is important to notice that the Gibbs sampling algorithm could be considered as a particular case of the Metropolis-Hastings algorithm (in which the acceptance ratio is 1), hence the considerations about convergency we will make for MH are valid for the Gibbs sampling, too.
The construction of the MCMC procedure are such to ensure that, if the Markov chain is convergent, it converges towards the desired probability density. To prove

the convergence of the algorithm we need to satisfy the two condition of irreducibility and aperiodicity of the chain (typical requests of the Markov chain theory); and it is obvious that they lead to some restrictions for the proposal distribution. Some sufficiency results have been already cited in the introductive part of this section on MCMC.
From these theoretical results we have our working hypothesis for the proposal, to ensure that the theoretical requirements are satisfied[4]. But it is not sufficient to ensure the efficiency of the algorithm. First of all, we specify further what we mean with efficiency: for our purposes the MCMC algorithm is efficient if it reach quickly the stationary distribution and explores its support properly (without, for example, remaining stuck in a mode), saying it differently we need that the Markov chain has good mixing properties.
As before for irreducibility and aperiodicity, we can reach our goal "tuning" the proposal distribution, but differently from before, we have not specific rules to satisfy, but just some rules of thumb to address our choices. It is obvious that using a proposal with a distribution approximately similar to the posterior density would be preferable. The variance can be used as tuning factor to get an "optimal" acceptance ratio[5]. In Gelman, Gilks and Roberts [4], it is discussed the optimal value for acceptance ratio: for an one-dimension problem it is suggested the ratio should be between 40% and 50%, while the dimension increases the suggested ratio value decreases till a value about 23% for high-dimensional problem. A possible strategy to improve the efficiency of the algorithm, it is to run small chain, tuning the proposal density to obtain a good acceptance ratio, then we can run a longer chain to perform our desired analysis.
Once tuned the proposal distribution and got the chain, the essential point is checking if the stationary is reached and its support properly explored. Since the first part of the chain is influenced by the initial choice we made arbitrarily and first states of the chain are not extracted from the limiting distribution, we need to get rid of them (the part that "has memory" of our initial choice): the *burn-in* sequence. The Geweke diagnostics can be useful to determine if our choice for the burn-in was not appropriate (it doesn't prove definitively the effective convergence, but gives an evidence of not convergence in most of the bad cases). The Geweke convergence diagnostics of the chain it is conducted comparing the means of two different batch extracted from the sample. Since the state of the chain are usually autocorrelated it is needed to define the spectral density of the chain $\{\theta^{(t)}; t = 1, \dots M\}$:

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \mathbb{C}\text{ov}(\theta^{(0)}, \theta^{(k)}) \exp\{ik\omega\} \tag{2.10}$$

[4]Since in Metropolis within Gibbs with Random Walk Metropolis Hastings (RWMH) step is the technique we use in the second part of the thesis, we remark again that using RWMH we are ensured to satisfy these requirements.

[5]The ratio that tracks the percentage of accepted proposals

The spectral density (2.10) is related to the asymptotic variance of the chain. Geweke [29] suggests to use this spectral density to build the convergence diagnostics test: the first step is splitting the chain in two non-overlapping batches[6] and computing the means of the two sub-chains:

$$\hat{\mu}_1 = \frac{1}{M_1} \sum_{t=1}^{M_1} \theta^{(t)}$$
$$\hat{\mu}_2 = \frac{1}{M_2} \sum_{t=M-M_2+1}^{M} \theta^{(t)}$$

Then it is needed to estimate the two asymptotic variance, $\frac{\hat{\sigma}_{1,2}}{M_{1,2}}$, by (2.10) opportunely truncated. Geweke [29] proved that the distribution of the statistics

$$\frac{\sqrt{M}(\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\frac{\hat{\sigma}_1}{M_1} + \frac{\hat{\sigma}_2}{M_2}}}$$

converges asymptotically to a standard normal distribution and can be used to check the lack of convergence of the chain or detect the burn-in to be discarded. Possible alternatives to this convergence test is the Kolmogorov−Smirnov test on two batches after thinning the subchain to construct two quasi−iid samples from the two batches coming from the MCMC starting chain[7].
Another issue is to avoid that the chain can be stuck in a mode and we don't recognize the bad behavior (if we perform the tests to check the convergence to the stationary distribution it will not highlight any misbehavior, in this case). A possible way is to run different chain from different starting points $\theta^{(0)}$, and compare the results. Some simple graphs can be used to help in detecting bad mixing property (and can be used in tuning the proposal distribution): a plot of the time series can highlight if the chain remain stuck for an excessive long time in a particular value, an autocorrelation plot can, instead, show the speed of the chain in forgetting the dependence from previous values: a slow-decay function in the autocorrelation graph suggests bad mixing properties.
Supposing we are sure we reached the stationary distribution and we eliminated a sufficiently long burn-in period, then we need to pay attention to how to define the useful estimator using the sample we have generated; in particular, a last remark

[6] Geweke test is the one used to check the lack of convergence in our estimation, conducted in the second part of the thesis, we used for the two batches the first 35% of the chain and the last 50% of it, once removed the burn-in.

[7] Thinning the chain is a largely used technique, according to which the chain it is divided in batches, sufficiently large to consider each mean of the batch zero-correlated with the other batch means, then the batch means are used as a new sample with zero autocorrelation.

is due about the variance estimation for the stationary distribution. While for the mean estimate we can just compute the mean of the values of the chain excluded the burn-in values, using the law of large numbers (2.7), the sample variance estimate can be a biased estimator if we do not take into account the autocorrelations (2.8). This because (as highlighted in [14]) we are not extracting i.i.d. sample from the distribution but we are generating a chain with Markov property, hence the values are correlated among themselves. To solve this problem and to get an unbiased estimator there are two possibilities: using the (2.8) or thinning the chain, dividing the chain in equally spaced batches, computing the standard deviation of the mean values of the batches[8]. The batch standard deviation is a unbiased estimator of the standard deviation of the stationary distribution, under the hypothesis that the batch width is longer than the characteristic mixing time, that is the time needed by the chain to forget the initial value (indications to infer it can be obtained by the autocorrelation graph).

2.4 Particle filter Algorithms

Particle filters have been introduced by Gordon et al. in [44], as a recursive bayesian algorithm to overcome standard algorithm limitations in dealing with non-linear models or non-gaussian noise. The algorithm was the so called bootstrap filter, it was proposed as an alternative to the Extended Kalman filter (EKF) algorithm for non linear system state space model, and they showed how the bootstrap filter, in the considered example in [44], even with a modest number of sampled paths is significantly superior to the EKF algorithm. The reason of the PF advantage is that this family of algorithms does not rely on any local linearisation or functional approximation, which may lead to gross distortion of the true underlying structure.
Hence the particle filters fits perfectly the need to filter latent processes in multifactor dynamics when the we have non linear system state space (for example when we deal with CIR processes) or non-Gaussian noise (as when jump processes are included in dynamics).
A possible alternative in the family of the Monte Carlo methods, to get inference on our model specification, are the Monte Carlo Markov Chain (MCMC) algorithms that we discussed in previous section 2.3, but in Sec.2.4.5 it is shown how to integrate these two Monte Carlo families to get inference on model specifications and latent process.

2.4.1 Application range

Particle filter algorithms are bayesian recursive Monte Carlo techniques that can be applied to state space form models.

[8] the sample obtained by thinning is zero correlated but not necessarily i.i.d.

A general state space model is a discrete time system model, specified via a parameter set $\theta \in \Theta$, where an hidden state vector $\{x_t; t = 0, \dots T\}$ (with $x_t \in \mathcal{X}$) of the model determines the probability density with which an observable variable $\{z_t; t = 1, \dots T\}$ is observed. The dynamic model is represented:

$$\begin{cases} z_t & \sim p(z_t|x_t, z_{1:t-1}, \theta) \\ x_t & \sim p(x_t|x_{0:t-1}, z_{1:t-1}, \theta) \\ x_0 & \sim p(x_0|\theta) \end{cases} \tag{2.11}$$

with $t = 1, \dots, T$. The main components are: an initial distribution $p(x_0|\theta)$, a measurement density $p(z_t|x_t, z_{1:t-1}, \theta)$ and a transition density $p(x_t|x_{0:t-1}, z_{1:t-1}, \theta)$. When the transition density function depends just on the most recent past value, $p(x_t|x_{0:t-1}, z_{1:t-1}, \theta) = p(x_t|x_{t-1}, \theta)$, the model is said first-order Markovian; if the transition density depends only on the most recent past value, and the observation at any time t is independent from the other observed value conditionally to the present value of the hidden state, x_t, then the model is said to be an Hidden Markov Model:

$$\begin{cases} z_t & \sim p(z_t|x_t, \theta) \\ x_t & \sim p(x_t|x_{t-1}, \theta) \\ x_0 & \sim p(x_0|\theta) \end{cases} \tag{2.12}$$

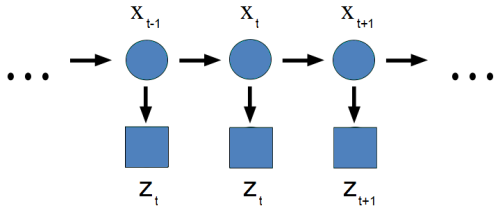

Figure 2.4: *Structure of an Hidden Markov model. With circles are indicated the hidden variables, while the boxes represent the observable variables*

It is worth noting that the models we introduced in the chapter 1 can be represented under the form (2.12), so also ours. Hence, since our estimation goal is restricted to (2.12), our description of the algorithms will follow this system but the algorithms presented can be trivially extended to the more general model (2.11). The first issue to take into account is the estimation of hidden state values, that is the estimation of $\{x_t; t = 0, \dots T\}$, considering θ fixed and known. From the specification of the model indicated in (2.11) at time $t = 1$ we know the density $p(x_0|\theta)$. Then, starting from $t = 1$ and going further on (supposing to know

$p(x_{t-1}|z_{1:t-1}, \theta)$), we can obtain the so called prediction density:

$$p(x_t|z_{1:t-1}, \theta) = \int_{\mathcal{X}} p(x_t|x_{t-1}, z_{1:t-1}, \theta)p(x_{t-1}|z_{1:t-1}, \theta)dx_{t-1} \tag{2.13}$$

Since we move to time t, a new information becomes available, z_t, and by the Bayes theorem, we can update the prediction density:

$$p(x_t|z_{1:t}, \theta) = \frac{p(z_t|x_t, z_{1:t-1}, \theta)p(x_t|z_{1:t-1}, \theta)}{p(z_t|z_{1:t-1}, \theta)} \tag{2.14}$$

(2.14) is called filtering density. The problem in using (2.13) and (2.14) to build an efficient algorithm to make inference on the hidden state process $\{x_t; t = 0, \dots T\}$ is due to the not feasibility of the integral computation involved in (2.13). Hence we need to find a numerical approximations. Among other possibilities to numerically approximate it (extended Kalman filter and Gibbs sampling) we give a presentation of Particle Filter (PF) algorithms. The first one proposed by Gordon et al. in [44] was the bootstrap filter, and it was presented again in a slightly modified version by Doucet Godsill and Andrieu [2] under the name Sequential Importance Sampling Algorithm (SIS). We describe this version and his main variations in the next part of this chapter.

2.4.2 Sequential Importance Sampling algorithm

The main point of this algorithm is that, since it is not possible to sample from the state posterior $p(x_{0:T}|z_{1:T})$ we adopt an approach that sequentially implements the Importance Sampling technique to generate a Monte Carlo sample. Sequentially, at any time step t, as a new observation z_t is available we predict and filter the hidden variable, considering, for the moment, the parameter set θ known. Let us suppose to draw a weighted sample $\{(x_t^i, \omega_t^i); i = 1, \dots, N\}$from the filtering density (2.14) at time t. Each simulated value x_t^i is called a "particle" with associated a probability weight ω_t^i. Alternatively we can say that we can use this "particle sample" as Monte Carlo estimate for the filtering density:

$$\hat{p}(x_t|z_{1:t}, \theta) = \sum_{i=1}^{N} \omega_t^i \delta_{x_t}(x_t^i) \tag{2.15}$$

And we can approximate the prediction density (2.13) and the filtering density(2.14), using of the estimation in (2.15):

$$p(x_{t+1}|z_{1:t}, \theta) \approx \sum_{i=1}^{N} \omega_t^i p(x_{t+1}|x_t^i, \theta) \tag{2.16}$$

$$p(x_{t+1}|z_{1:t+1}, \theta) \approx \sum_{i=1}^{N} \omega_t^i \frac{p(z_{t+1}|x_{t+1}^i, z_{1:t}, \theta)p(x_{t+1}|x_t^i, \theta)}{p(z_{t+1}|z_{1:t}, \theta)} \tag{2.17}$$

the two quantities just defined are called empirical prediction density and empirical filtering density (2.17). To get a new sample from the filtering density we draw a a sample using Importance Sampling technique: we propagate the previous particle set $\{x_t^i; i = 1, \ldots, N\}$ to $\{x_{t+1}^i; i = 1, \ldots, N\}$ through the importance function $q(x_{t+1}|x_t^i, z_{1:t+1}, \theta)$. Then we update the weights ω_t^i to the new weight set $\{\omega_{t+1}^i; I = 1, \ldots, N\}$, using the importance weights defined in section 2.1.1.

$$\omega_{t+1}^i \propto \frac{p(z_{t+1}|x_{t+1}^i, z_{1:t}, \theta)p(x_{t+1}^i|x_t^i, \theta)}{q(x_{t+1}|x_t^i, z_{1:t+1}, \theta)}\omega_t^i \tag{2.18}$$

In the bootstrap filter version of the algorithm $q(x_{t+1}|x_t^i, z_{1:t+1}, \theta) = p(x_{t+1}|x_t^i, \theta)$.

Summarizing up: we want to approximate the smoothing density with the particle set:

$$p(x_{0:T}|z_{1:T}, \theta) \approx \sum_{i=1}^{N} \omega_T^i \delta_{x_{0:T}}(x_{0:T}^i)$$

and by definition the particle weight are the importance weight of a sample draw from a proposal distribution $q(x_{0:T}|z_{1:T}, \theta)$:

$$\omega_T^i = \frac{p(x_{0:T}|z_{1:T}, \theta)}{q(x_{0:T}|z_{1:T}, \theta)} \tag{2.19}$$

It is possible to elaborate the (2.19) using the Bayes formula, to get a recursive definition for this importance weights, for the model (2.12):

$$\begin{aligned}
\omega_t^i &= \frac{p\left(x_{0:t}|z_{1:t}\right)}{q\left(x_{0:t}|z_{1:t}\right)} && (2.20)\\
&= \frac{p(x_{0:t}, z_t|z_{1:t-1})}{q(x_{0:t}|z_{1:t})p(z_t|z_{1:t-1})} && (2.21)\\
&= \frac{p(x_{0:t-1}|z_{1:t-1})p(x_t, z_t|x_{0:t-1}, z_{0:t-1})}{q(x_{0:t}|z_{1:t})p(z_t|z_{1:t-1})} && (2.22)\\
&= \frac{p(x_{0:t-1}|z_{1:t-1})p(z_t|x_{0:t}, z_{0:t-1})}{q(x_{0:t}|z_{1:t})p(z_t|z_{1:t-1})}p(x_t|x_{0:t-1}, z_{0:t-1}) && (2.23)\\
&= \frac{p(x_{0:t-1}|z_{1:t-1})p(z_t|x_t)}{q(x_{0:t}|z_{1:t})p(z_t|z_{1:t-1})}p(x_t|x_{t-1}) && (2.24)\\
&= \frac{p(x_{0:t-1}|z_{1:t-1})}{q(x_{0:t-1}|z_{1:t-1})}\frac{p(x_t|x_{t-1})p(z_t|x_t)}{p(z_t|z_{1:t-1})q(x_t|x_{0:t-1}, z_{1:t-1})} && (2.25)
\end{aligned}$$

Since $p(z_t|x_{1:t-1}, \theta)$ is not dependent on the given particle we can treat it like a normalizing factor, hence, the non-normalized weights in the SIS algorithm come from the recursive formula:

$$\omega_t^i = \omega_{t-1}^i \frac{p(z_t|x_t, \theta)p(x_t|x_{t-1}, \theta)}{q(x_t|x_{t-1}, \theta)} \tag{2.26}$$

It is important to notice that we can recursively compute also the importance weights, this allow us to write down the Sequential Important Sampling in an algorithmic form.

Algorithm 4 Sequantial Importance Sampling Algorithm (SIS)

1: **for** k=1:T **do** ▷ iteration on time
2: **for** j=1:N **do** ▷ each j identify a particle
3: sample

$$x_k^{(j)} \sim q(x_k|x_{0:k-1}^{(j)}, z_{0:k})$$

▷ if $k = 1$ sample from $q(x_0|\theta)$

4: compute the importance weights:

$$\omega_k^{(j)} = \frac{p(z_k|x_k^{(j)})p(x_k^{(j)}|x_{k-1}^{(j)}, z_{1:k-1}, \theta)}{q(x_k|x_{0:k-1}^{(j)}, z_{0:k}, \theta)}\omega_{k-1}^{(j)}$$

5: **end for**
6: normalize the weights: $w_k^{(j)} = \frac{\omega_k^{(j)}}{\sum_{l=1}^{N}\omega_k^{(l)}}$
7: **end for**

In Fig.2.5 it is reported a graphical scheme of the SIR algorithm. At a given step:

- the stating point is the initial particle set $\{(x_t^i, \omega_t^i); i = 1, \dots N\}$ approximating the distribution $p(x_t|z_{1:t}$
- the particles are propagated via the importance function $q(x_{t+1}|z_{t+1}, x_t)$
- finally, the weight are update to have a new sample: $\{(x_{t+1}^i, \omega_{t+1}^i); i = 1, \dots N\}$ approximating the distribution $p(x_{t+1}|z_{1:t+1})$

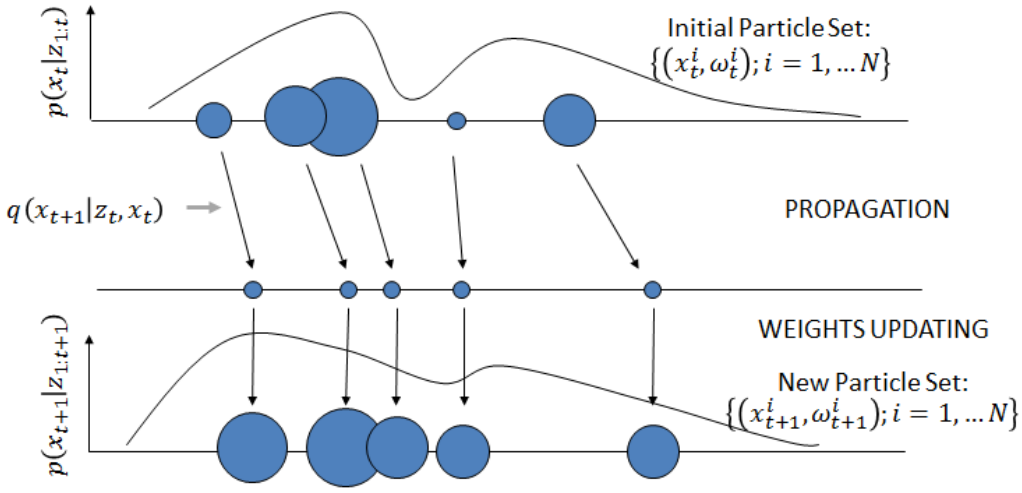

Figure 2.5: *Scheme of a SIS algorithm at step t*

2.4.3 SIR and APF

As discussed in [2] the unconditional variance of the importance weights increases over time, to limit the degeneracy of the algorithm the most natural strategy is to select an importance function which minimizes the variance of the importance weights conditionally on the simulated trajectories $x_{0:t-1}^i$ and the observation chain $z_{0:t}$. As proved in [2] the optimal importance function to minimize this variance is

$$q(x_t|x_{t-1}, z_{1:t}, \theta) = p(x_t|x_{t-1}, z_t, \theta) \tag{2.27}$$

But this requires to sample from $p(x_t|x_{t-1}^i, z_t)$, that is possible if we are able to evaluate

$$p(z_t|x_{t-1}^i) = \int_{\mathcal{X}} p(z_t|x_t)p(x_t|x_{t-1}^i)dx_t \tag{2.28}$$

but this integral have no analytical solution, generally. The second method to prevent the degeneracy of the SIS algorithm is the resampling step technique, that modifies the SIS algorithm, and it is at the basis of the Sequential Importance Resampling (SIR) algorithm and the Auxiliary Particle Filter (APF) algorithm. The resampling step is helpful in eliminating trajectories with small importance weights and focusing on the more probable trajectories; on the contrary, the resampling step has a computational cost in time required to resample, so it can be used a measure of the degeneracy of the algorithm to discriminate if the resample step is needed. A suitable measure, introduced by Liu in 1996, and discussed in [2] and in [7], is the estimated effective sample size $\hat{N}_{\text{eff}}$:

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{l=1}^{N}(\omega_k^{(l)})^2} \tag{2.29}$$

When $\hat{N}_{\text{eff}}$ is below a given threshold a resample step is applied and it is drawn a sample from the particle set $\{(x^i_{0:t}, \omega^i_t); i = 1, \ldots, N\}$: the resampled particle set is characterized by equal weights and we can use it to approximate the smoothing density. The SIR algorithm is:

Algorithm 5 Sequential Importance Resampling (SIR)

1: **for** k=1:T **do** ▷ iteration on time
2: **for** j=1:N **do** ▷ each j identify a particle
3: sample

$$x^{(j)}_k \sim q(x_k | x^{(j)}_{0:k-1}, z_{1:k}, \theta)$$

▷ if $k = 1$ sample from $q(x_0|\theta)$

4: compute the importance weights:

$$\omega^{(j)}_k = \omega^{(j)}_{k-1} \frac{p(z_k|x^{(j)}_k) p(x^{(j)}_k | x^{(j)}_{k-1})}{q(x_k | x^{(j)}_{0:k-1}, z_{0:k})}$$

5: **end for**
6: normalize the weights: $w^{(j)}_k = \dfrac{\omega^{(j)}_k}{\sum_{l=1}^N \omega^{(l)}_k}$
7: **for** j=1:N **do** ▷ RESAMPLE step (if $\hat{N}_{\text{eff}} < N_{\text{threshold}}$)
8: sample $x^{(j)}_{0:k}$ from $\displaystyle\sum_{i=1}^N w^{(i)}_k \delta_{x^{(i)}_{0:k}}$
9: set all the weights equal to $\frac{1}{N}$
10: **end for**
11: **end for**

In Fig.2.6 it is reported a graphical scheme of the SIR algorithm at a given step:

- the stating point is the initial particle set $\{(x^i_t, \omega^i_t); i = 1, \ldots N\}$ approximating the distribution $p(x_t|z_{1:t}$
- we resample from it, getting an equally weighted sample (approximating the same distribution $p(x_t|z_{1:t})$; in this way, we allow only the most probable particle to survive
- the survived particles are propagated via the importance function $q(x_{t+1}|z_{t+1}, x_t)$
- finally, the weight are update to have a new sample: $\{(x^i_{t+1}, \omega^i_{t+1}); i = 1, \ldots N\}$ approximating the distribution $p(x_{t+1}|z_{1:t+1})$

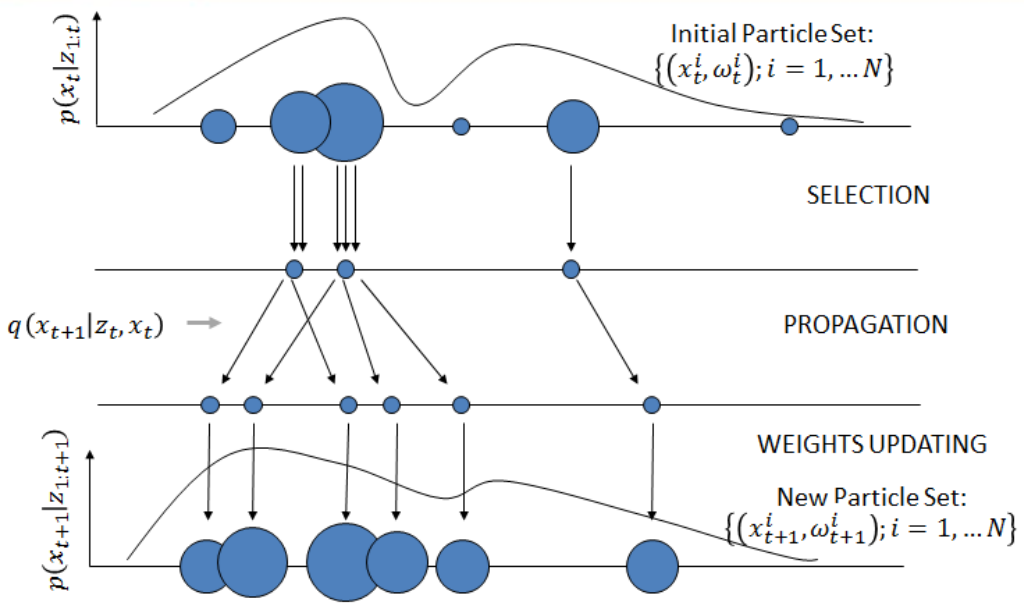

Figure 2.6: *Scheme of a SIR algorithm at step t*

An alternative to the SIR algorithm was proposed by Pitt and Shephard in [49], since in its original version uses auxiliary variables, its name is Auxiliary Particle Filter (APF). Following on, it is presented the version discussed in [7], which differs from the original one since it applies the resampling step just one time instead of two times; in this version, the resample step is performed at each time step.
The APF, with respect to the SIR algorithm, introduces an important feature: the resampling step (removing the low probability trajectory) is performed before the propagation step: in resampling we select the "ancestors" of the particles, and the selection is performed on the basis of an importance function closest as possible to the optimal importance function $p(x_t|x_{t-1}, z_t, \theta)$ (2.27).

Algorithm 6 Auxiliary Particle Filter Algorithm (APF)

1: **for** j=1:N **do** sample $x_0^j \sim q(x_0|\theta)$ and set

$$\omega_0 = \frac{p(x_0|\theta)}{q(x_0|\theta)}$$

2: **end for**
3: **for** k=1:T **do** ▷ iteration on time
4: the sample $\{(x_{0:n-1}^{(i)}, w_{k-1}^{(i)}); i = 1, \dots N\}$ is given
5:
6: **for** j=1:N **do** ▷ each j identify a particle
7: sample

$$\tilde{x}_k^{(j)} \sim q(\tilde{x}_k|x_{0:k-1}^{(j)}, z_{0:k})$$

8: sample ▷ RESAMPLE step

$$j^i \sim q(j|z_{1:k}, \theta) \propto w_{k-1}^{(j)} p(z_k|\tilde{x}_k, \theta) q(\tilde{x}_k|x_{0:k-1}^{(j)}, z_{0:k-1})$$

9: sample

$$x_k^{(j)} \sim q(x_k|x_{0:k-1}^{(j^i)}, z_n)$$

10: compute the importance weights:

$$\omega_{k-1}^{(j)} = \frac{p(z_k|x_k^{(j)}) p(x_k^{(j)}|x_{k-1}^{(j)})}{q(x_k|x_{0:k-1}^{(j)}, z_{0:k-1}) q(\tilde{x}_k^{(j)}|x_{0:k-1}^{(j)}, z_{0:k-1})}$$

11: **end for**
12: normalize the weights: $w_k^{(j)} = \dfrac{\omega_k^{(j)}}{\sum_{l=1}^N \omega_k^{(l)}}$
13: **end for**

An important remark about the APF (Algorithm 6) is that can be viewed as a SIR algorithm with the peculiar choice of the importance function. So, the efficiency of the APF algorithm depends just on the approximation of the density $p(x_t|z_{t+1}, x_{1:t}, \theta)$. In literature, as discussed in [7] the usual choice is to take in the algorithm 6:

$$\begin{cases} \tilde{x}_k & = \mu_k = \mathbb{E}[x_k|x_{k-1}, z_{k-1}, \theta] \\ q(\tilde{x}_k|x_{0:k-1}^{(j)}, z_{0:k}) & \propto p(z_k|\mu_k, \theta) \end{cases} \tag{2.30}$$

Further attempts to address the problem of sample impoverishment and degeneration of the algorithm, particularly relevant in Particle Filter literature are two proposal: resample-move proposed by Gilks and Benzuini in [31] (where they

suggest to add a MCMC step after resampling to reduce the degeneracy of the Particle filter algorithm) and block sampling by Doucet, Briers and Senecal in [1], both techniques propose to rejuvanate the sample by substituting old sampled (and resampled) values by new values for the particles, using a MCMC step or an additional importance sampling step. In Resample Move algorithm [31] it is suggested to use a Gibbs sampler after the resampling, using a transition kernel invariant with respect to the smoothing density distribution, since it is usually not possible to sample directly from this distribution, we can substitute the Gibbs step with a Metropolis-Hastings step with an opportune acceptance ratio, as usual. Since the moved particles are approximately distributed according to the desired distribution (we sampled with Particle filter: we are sampling from the unknown distribution via importance sampling) we do not need any burn-in. In block sampling technique described in [1] it is suggested a variation with respect to the resample-move in [31]: to sample new values for old particles it is used a new importance function, discarding old particle after the "move" step, in this way also the weights associated to the particle change. In the pseudo-code algorithm (Algorithm7) it is presented the Resample-Move technique in a SIR algorithm where, with K_k it is indicated the kernel of the Gibbs sampler:

$$K_k(x'_{1:k}, x_{1:k}) = \delta_{x'_{1:k-L}}(x_{1:k-L}) \prod_{s=k-L+1}^{k} p(x'_s|z_{1:k}, x'_{1:s-1}, x_{s+1:k}) \tag{2.31}$$

2.4.4 Parameter Estimation with Particle filter

Particle filter can be used also to make inference about the parameter set. The starting point is considering the likelihood function of the data:

$$P(z_{1:N}) = \int \omega_N p(x_{0:N}|z_{1:N}) d(x_{0:N}) \tag{2.32}$$

This formulation allows us to estimate the likelihood (as explained in [7]), at least for the SIS algorithm, using of our simulated particle set:

$$\hat{P}(z_{1:N}) = \frac{1}{M} \sum_{j=1}^{M} \omega_N^{(j)} \tag{2.33}$$

But if we resampled like in SIR or APF algorithm this approach cannot be applied. In these case we can decompose:

$$P(z_{1:N}) = p(z_1) \prod_{i=1}^{N} p(z_i|z_{1:i-1}) \tag{2.34}$$

Algorithm 7 Resample Move

for k=1:L **do** ▷ iteration on time
  for j=1:N **do** ▷ each j identify a particle
    sample

$$x_k^{(j)} \sim q(x_k|x_{0:k-1}^{(j)}, z_{1:k}, \theta)$$

▷ if $k = 1$ sample from $q(x_0|\theta)$
    compute the importance weights:

$$\omega_k^{(j)} = \omega_{k-1}^{(j)} \frac{p(z_k|x_k^{(j)})p(x_k^{(j)}|x_{k-1}^{(j)})}{q(x_k|x_{0:k-1}^{(j)}, z_{0:k})}$$

  end for
  normalize the weights: $w_k^{(j)} = \frac{\omega_k^{(j)}}{\sum_{l=1}^{N} \omega_k^{(l)}}$
  for j=1:N **do** ▷ RESAMPLE step
    sample $x_{0:k}^{(j)}$ from $\sum_{i=1}^{N} w_k^{(i)} \delta_{x_{0:k}}(x_{0:k}^{(i)})$
    set all the weights equal to $\frac{1}{N}$
  end for
end for
for k=L+1:T **do** ▷ each j identify a particle

  for j=1:N **do** ▷ each j identify a particle
    sample

$$x_k^{(j)} \sim q(x_k|x_{0:k-1}^{(j)}, z_{1:k}, \theta)$$

▷ if $k = 1$ sample from $q(x_0|\theta)$
    compute the importance weights:

$$\omega_k^{(j)} = \omega_{k-1}^{(j)} \frac{p(z_k|x_k^{(j)})p(x_k^{(j)}|x_{k-1}^{(j)})}{q(x_k|x_{0:k-1}^{(j)}, z_{0:k})}$$

  end for
  normalize the weights: $w_k^{(j)} = \frac{\omega_k^{(j)}}{\sum_{l=1}^{N} \omega_k^{(l)}}$
  for j=1:N **do** ▷ RESAMPLE and MOVE steps
    sample $x_{0:k}^{(j)}$ from $\sum_{i=1}^{N} w_k^{(i)} \delta_{x_{0:k}}(x_{0:k}^{(i)})$
    set all the weights equal to $\frac{1}{N}$
    sample $x_{k-L+1:k}^{\prime(j)} \sim K_k(x_{k-L+1:k}, x_{0:k}^{(j)})$
    set $x_{0:k}^{(j)} = (x_{0:k-L}^{(j)}, x_{k-L+1:k}^{\prime(j)})$
  end for
end for

We know how to estimate $p(z_i|z_{1:i-1})$ with all our PF algorithms, and it is simply:

$$\hat{p}(z_i|z_{1:i-1}) = \sum_{j=1}^{N} p(z_i|x_i^{(j)}) w_i^{(j)} \tag{2.35}$$

The search for the optimal parameter set can be, then, performed using a Maximum Likelihood estimator, using a minimization algorithm to find the maximum of the loglikelihood function. Obviously, the use of a gradient based minimization algorithm brought with himself all the drawbacks typical of this algorithm (the main one: local minimum searching); due to the stochastic nature of the algorithm we should use a stochastic version of the gradient algorithm as explained in [7].
A second possible choice to get the parameter set is using another common algorithm: expectation maximization. Let us consider [5] the marginal loglikelihood $L(\theta)$ of the observation process given the parameter set and we approximate it using the expectation with respect to the hidden data X. Considering the model (2.12):

$$L(\theta) = \mathbb{E}_\theta[\log p_\theta(z_{1:T})] = \sum_{k=1}^{T} \log \int p_\theta(z_k|x_k) p_\theta(x_k|z_{k-1}) dx_k \tag{2.36}$$

We can approximate the loglikelihood defining the function $Q(\theta, \theta')$:

$$Q(\theta, \theta') = \mathbb{E}_{\theta'}[\log p_\theta(z_{1:T})] \tag{2.37}$$

where we computed the expectation using as parameter set a current estimate θ'. In [5] it is proven that if Q increases the Log-Likelihood function increases, then we can use the function Q to find the Maximum Likelihood. The EM algorithm consists in computing the expectation (2.36) and then maximizing $Q(\theta, \theta')$ with respect to θ. Iterating this two step we will reach a stable point θ^* that is our estimate. In [5] it is shown how we can estimate the $Q(\theta, \theta')$ function, making use of elements that we can compute via a PF algorithm:

$$\begin{aligned} \hat{Q}(\theta, \theta') = & \sum_{j=1}^{M} w_1^{(j)} \log p_\theta(x_0^{(j)}) + \\ & + \sum_{t=1}^{N} \sum_{j=1}^{M} w_{t+1}^{(j)} \log p_\theta(x_{t+1}^{(j)}|x_t^{(j)}) + \\ & + \sum_{t=1}^{N} \sum_{j=1}^{M} w_t^{(j)} \log p_\theta(z_t|x_t^{(j)}) \end{aligned} \tag{2.38}$$

where the weights $w^{(j)}$ and the particles $x^{(j)}$ are obtained by PF using the parameter set θ'. To perform the maximization step, maximizing $\hat{Q}(\theta, \theta')$ with respect to

Algorithm 8 EM Algorithm

1: **while** $\theta_k - \theta_{k-1} > V_{\text{thres}}$ **do** ▷ V_{thres} is a setted threshold value
2: Use a PF algorithm using as parameter set θ_{k-1}, obtaining the weights $w^{(j)}$ and the particles $x^{(j)}$
3: compute $\hat{Q}(\theta, \theta_{k-1})$
4: compute $\theta_k = \arg\max_\theta \left(\hat{Q}(\theta, \theta_{k-1})\right)$
5: $k = k + 1$
6: **end while**

θ, we need again an optimization algorithm like the gradient based ones. Starting from a randomly chosen value θ_0 for the parameter set, the algorithm is : Some variants to this algorithm have been introduced by Andrieu and Doucet in [6].
An other possible choice to estimate the parameter is estending the parameter state of the hidden state variables to incorporate also the θ state space $S_n = (X_{0:n}, \theta)$, so we can use directly the Particle Filter algorithms already discussed, introducing some sort of dynamics for the parameter, given the severe issues of degeneracy the use of Resample-Move algorithm is needed. This naive technique has been used, but the algorithm has been proved to be inefficient.

2.4.5 Particle Markov Chain Monte Carlo

All the previous algorithms suffer of well-known inefficiency drawbacks and degeneracy issues: when the largeness of the state space of the hidden state variables object of our estimation increases with respect to the number of particles, the joint density $p_\theta(x_{0:T}|z_{1:T})$ approximation becomes poor and the particle set need to be rejuvenated. PMCMC, a new class of algorithm introduces by Doucet, Andrieu and Holenstein in 2010 [10] overcomes these difficulties since they not try to approximate directly the joint posterior density $p_\theta(x_{0:T}|z_{1:T})$ but to return single sample, approximately distributed according to $p_\theta(x_{0:T}|z_{1:T})$ (nesting a PF algorithm in a MCMC setting)[9]. The core of the algorithm is the estimation of the marginal likelihood $p(\theta|z_{1:n})$ using the posterior density estimate $\hat{p}(x_{0:n}|z_{1:n}, \theta)$ obtained via a PF algorithm, the unbiasedness of the estimated likelihood has been proved by Del Moral in [41]. The version reported the Algorithm9 is the Particle Marginal Metropolis Hasting (PMMH) [10].
According to this algorithm we are using a Metropolis Hastings algorithm where the proposal density is:

$$q((\theta^*, x^*_{0:T})|(\theta, x_{0:T})) = q(\theta^*|\theta)p(x^*_{0:T}|z_{1:T}, \theta^*) \tag{2.39}$$

[9]Moreover, since it relies in two nested bayesian method, to get inference for both the parameter set and the latent processes, it can be considered a more natural solution for double inference than the EM algorithm and its variants.

Algorithm 9 Particle Marginal Metropolis Hastings (PMMH) - SIR

for i=0 **do**
  Set $\theta(0)$ arbitrarily
  Run a PF algorithm to get a MC sample for the posterior distribution $p(x_{0:T}|z_{1:T}, \theta(0))$
  extract a sample of 1 particle from the empirical smoothing distribution

$$X_{0:T}(0) \sim \hat{p}(x_{0:T}|z_{1:T}, \theta(0))$$

  Set the marginal likelihood estimate $\hat{p}(z_{1:T}|\theta(0))$ computed by:

$$\hat{p}(z_t|z_{1:t-1}, \theta(0)) = \left(\sum_{k=1}^{N} \frac{\omega_t^{(k)}}{N} \right)$$

end for
for i=1:N_{MH} **do** ▷ Metropolis Hastings steps
  Sample θ^* from $q(\theta|\theta(i-1))$
  Run a PF algorithm to get a MC sample for the posterior distribution $p(x_{0:T}|z_{1:T}, \theta*)$
  extract a sample of 1 particle from the empirical smoothing distribution

$$X_{0:T}* \sim \hat{p}(x_{0:T}|z_{1:T}, \theta^*)$$

  Set the marginal likelihood estimate $\hat{p}(z_{1:T}|\theta^*)$ considering the estimator:

$$\hat{p}(z_t|z_{1:t-1}, \theta^*) = \left(\sum_{k=1}^{N} \frac{\omega_t^{(k)}}{N} \right)$$

  compute the probability probability

$$\alpha = \min \left\{ 1, \frac{\hat{p}(z_{1:T}|\theta^*)p(\theta^*)}{\hat{p}(z_{1:T}|\theta(i-1))p(\theta(i-1))} \frac{q(\theta(i-1)|\theta^*)}{q(\theta^*|\theta(i-1))} \right\}$$

  with probability α set $(\theta(i), X_{0:T}(i)) = (\theta^*, X^*_{0:T})$ otherwise $(\theta(i), X_{0:T}(i)) = (\theta(i-1), X_{0:T}(i-1))$
end for

Under the usual condition for the proposal density $q(\theta^*|\theta)$ for the Metropolis Hastings algorithm, the authors in [10] proved that the sequence of sample is a sample from a Markov chain whose distributions (at each time step) converge to the limiting distribution $p(X_{0:T}, \theta|z_{1:T})$, if an additional technical requirement on the resampling scheme is satisfied[10].
Pitt and Shephard extended [40] the standard PMMH algorithm 9 (where a SIR algorithm it is nested in the MCMC structure) with a version using an APF algorithm to estimate the marginal likelihood, the pseudo code is in Algorithm10. In the second part of the thesis we used the Algorithm 9 to get our inference.

2.5 Other methods

2.5.1 Method of Moments

One possible alternative to the Monte Carlo methods is to use the method of moments (MM), the idea underlying MM methods is simple: after having computed the theoretical moments of our model, we can get an estimate of the parameter set comparing them with the moments of the empirical distribution of realized observations. Even if the idea is very simple, his application is not simple at all, since there are a lot of practical issues to manage. First of all the choice of the moments. Obviously, the first moments are the most significant, but to which order of moments it is important to include in the algorithm, and how to weight the different moments is still an open issue in literature (for sure the first two moments are more important than the sixth and seventh and we have to take it into account). But these are not the only practical issues to cope with: if our model has a latent process we cannot usually find the moments function analytically.
Duffie and Singleton proposed in [19] a method of moments based on simulation: given a parameter set specification we can simulate via a Monte Carlo the distribution obtained with our model and use this distribution to compute numerically the moments. Numerical methods can, then, be applied to infer the best parameter specification to make the simulated moments the closest[11] as possible to the moments of the empirical distribution. To present this methods we follow the résumé written by Renault [38]. The notation is the same used in the previous chapters, and in general in the whole thesis.
With $\sum_{i=1}^{T} K[z_t]$ we indicate the vector of the moments based on the observed chain $\{z_t\}_{t=0,\dots N}$, and with $\sum_{i=1}^{T} K[z_t^{(h)}(\theta)]$ the moment of the h^{th} simulation,

[10] This requirement is satisfied for the classical SIR algorithm [10] but it is not proved that Resample Move and Block Sampling techniques satisfy it.

[11] To keep the norm of the errors the closest as possible to zero; since we want to construct an "efficient" norm (or said in a different way an efficient cost function) we come back to the practical issues we discussed before (number of moments to include and optimal weight matrix specification).

Algorithm 10 Particle Marginal Metropolis Hastings (PMMH) - APF

for i=0 **do**
  Set $\theta(0)$ arbitrarily
  Run a PF algorithm to get a MC sample for the posterior distribution $p(x_{0:T}|z_{1:T}, \theta(0))$
  extract a sample of 1 particle from the empirical smoothing distribution

$$X_{0:T}(0) \sim \hat{p}(x_{0:T}|z_{1:T}, \theta(0))$$

  computed by:

$$\hat{p}(z_t|z_{1:t-1}, \theta(0)) = \left(\sum_{k=1}^{N} \frac{\omega_t^{(k)}}{N}\right)\left(\sum_{k=1}^{N} \omega_{t-1|t}^{(k)}\right)$$

  Set the marginal likelihood estimate $\hat{p}(z_{1:T}|\theta(0))$
end for
for i=1:N_{MH} **do** ▷ Metropolis Hastings steps
  Sample θ^* from $q(\theta|\theta(i-1))$
  Run a PF algorithm to get a MC sample for the posterior distribution $p(x_{0:T}|z_{1:T}, \theta^*)$
  extract a sample of 1 particle from the empirical smoothing distribution

$$X_{0:T}^* \sim \hat{p}(x_{0:T}|z_{1:T}, \theta^*)$$

  Set the marginal likelihood estimate $\hat{p}(z_{1:T}|\theta^*)$ computed by:

$$\hat{p}(z_t|z_{1:t-1}, \theta^*) = \left(\sum_{k=1}^{N} \frac{\omega_t^{(k)}}{N}\right)\left(\sum_{k=1}^{N} \omega_{t-1|t}^{(k)}\right)$$

  compute the probability ratio

$$\alpha = \min\left\{1, \frac{\hat{p}(z_{1:T}|\theta^*)p(\theta^*)}{\hat{p}(z_{1:T}|\theta(i-1))p(\theta(i-1))}\frac{q(\theta(i-1)|\theta^*)}{q(\theta^*|\theta(i-1))}\right\}$$

  with probability α set $(\theta(i), X_{0:T}(i)) = (\theta^*, X_{0:T}^*)$ otherwise $(\theta(i), X_{0:T}(i)) = (\theta(i-1), X_{0:T}(i-1))$
end for

then our parameter estimate is obtainable searching for the θ which minimize the following norm.

$$\|\frac{1}{T}\sum_{i=1}^{T} Z_t \left\{ K[z_t] - \frac{1}{M}\sum_{h=1}^{M} K[z_t^{(h)}(\theta)] \right\}\| \tag{2.40}$$

Z_t is a matrix of chosen instruments and could be used as weight function to define the importance of the different moments we have taken into account. A slight modification of this technique, allow to construct a new estimator: the efficient method of moments (EMM) estimator. Replacing $\frac{1}{T}\sum_{i=1}^{T} Z_t K[z_t]$ with the solution of a maximization problem:

$$\max_{\beta} \frac{1}{T}\sum_{i=1}^{T} q_t(z(0,t),\beta)$$

where $q_t(z(0,t),\beta)$ is the log-likelihood of an auxiliary model used as approximation of our real model[12]. In this formulation the estimate for θ is obtained by minimizing the norm:

$$\|\frac{1}{T}\sum_{i=1}^{T} \frac{\partial q_t(z(0,t),\beta)}{\partial \beta} - \frac{1}{M}\sum_{h=1}^{M} \frac{\partial q_t(z^{(h)}(0,t),\beta)}{\partial \beta}\| \tag{2.41}$$

It has been proved by Gouriroux, Monfort and Renault in [11] that the estimator is unbiased and efficient, meaning that asymptotically has the same efficiency of the maximum likelihood estimator for the auxiliary model[13]. Since we search for β as the maximizer of $\frac{1}{T}\sum_{i=1}^{T} q_t(z(0,t),\beta)$ the estimator that minimize (2.42) becomes simply:

$$\min_{\theta} \|\frac{1}{TM}\sum_{i=1}^{T}\sum_{h=1}^{M} \frac{\partial q_t(z^{(h)}(0,t),\beta)}{\partial \beta}\| \tag{2.42}$$

2.5.2 Quasi Maximum Likelihood

The quasi Maximum likelihood (QML) method is a family of methods based on an approximation of the likelihood function. Since we don't know analytically the functional form of the likelihood for a process with stochastic volatility, a possible solution is to use an approximated model to estimate the likelihood function, then

[12] Obviously, now we can choose a model for which we know how to compute analytically the likelihood function

[13] The EMM estimator is asymptotically equivalent (also as distribution form) to the MLE estimator for the auxiliary model, and inherits its asymptotic properties

this function is used to build the estimator for our parameter set.The approximated likelihood function is the quasi likelihood function.
The main algorithm of this family, that can be applied to Stochastic Volatility models is the Extended Kalman Filter, a QML version of the well-known Kalman filter. We follow the discussion and the algorithm presented in [37]. Kalman Filtering is a filter technique that has been built to study state-space models where the noises are gaussian and the relation between the states and the observation is linear, but has been extended to include also non linear systems.
The standard Kalman filter applies just to Gaussian linear models. As usual we indicate with X the unobservable state and with Z the observable variable. The requirements for the application of the Kalman filter is that $X \sim \mathcal{N}(m_x, S_{xx})$ and $Z \sim \mathcal{N}(m_z, S_{zz})$. Let us indicate with S_{xz} the covariance between the two random variable. Since both the variable are affected by gaussian noises and their relation in the state-space formulation is linear, via the standard Kalman filter we can infer[37] that the conditional distribution of the latent process X given the observation process Z is normal with mean and variance:

$$\begin{cases} m_{x|z} = m_x + K(z - m_z) \\ K = S_{xz}S_{zz}^{-1} \\ S_{x|z} = S_{xx} - KS_{xz} \end{cases}$$

where K is the Kalman gain.

Even if our model does not fit the Kalman filter requirements, we can use the Kalman filter on the local first order linearization of it, than with the likelihood estimated by the Kalman filter on the linearized model. If we maximize it to search for the estimate of the parameter set we are using a QML estimator.

This method is known as Extended Kalman Filter (EKF). Let our model be:

$$\begin{cases} x_k = f(x_{k-1}, \xi_k) \\ z_k = g(x_k, u_k) \end{cases}$$

with ξ_k and u_k two uncorrelated sequence of i.i.d. gaussian noises; ξ_k and u_k are also uncorrelated with x_{k-1} and x_k.
Let $\xi_k \sim \mathcal{N}(0, Q_k)$ and $u_k \sim \mathcal{N}(0, R_k)$. Then we define two estimate for x_k: a priori estimate $\hat{x}_k^- = E(x_k)$ (estimation taken at time $k-1$) and a posteriori estimate $\hat{x}_k = E(x_k|z_k)$ (estimation taken after we observed the k^{th} observation value). Hence, we can also define two estimation error: $e_k^- = x_k - \hat{x}_k^-$ and $e_k = x_k - \hat{x}_k$, and two error covariance matrices $P_k^- = \left[e_k^- e_k^{-T}\right]$ and $P_k = \left[e_k e_k^T\right]$. Then defining:

$$\begin{cases} A_{ij} = \frac{\partial f_i}{\partial x_j}(\hat{x}_{k-1}, 0) \\ W_{ij} = \frac{\partial f_i}{\partial \xi_j}(\hat{x}_{k-1}, 0) \\ H_{ij} = \frac{\partial g_i}{\partial x_j}(\hat{x}_k^-, 0) \\ U_{ij} = \frac{\partial g_i}{\partial u_j}(\hat{x}_k^-, 0) \end{cases}$$

The estimation at time k is [37]:

$$\begin{cases} \hat{x}_k^- = f(\hat{x}_{k-1}, 0) \\ P_k^- = A_k P_{k-1} A_k^T + W_k Q_{k-1} W_k^T \end{cases}$$

We define a Kalman gain matrix K_k by $P_k = (I - K_k H_k) P_k^-$.
Hence, we can write [37]:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - h\left(\hat{x}_k^-, 0\right)) \tag{2.43}$$

The optimal Kalman gain is the matrix minimizing the mean square error P_k within the class of the linear estimator. This is:

$$K_k = P_k^- H_k^T \left(H_k P_k^- H_k^T + U_k R_k U_k^T\right)^{-1} \tag{2.44}$$

To get our parameter estimator it is used a joint filter which gives us not only the state process estimate but also the parameter estimate: the density we have to take into account is the

$$p(x_{1:T}, \theta | z_{1:T}) = \frac{p(z_{1:T}|x_{1:T}, \theta) p(x_{1:T}|\theta) p(\theta)}{p(z_{1:T})}$$

If we do not now in advance anything on θ (said in an equivalent way, there is no prior information in $p(\theta)$) then the estimation will be obtained by maximizing the likelihood, in formulae

$$(\hat{x}_{1:T}, \hat{\theta}) = \underset{(x_{1:T}, \theta)}{\arg\max}\, p(z_{1:T}|x_{1:T}, \theta) \tag{2.45}$$

This algorithm works well when we are using a model whose linearized version is a good approximation and the noises are gaussian. We have already discussed in section about MCMC algorithm that in different papers like [20] methods based on MCMC showed better performances than the corresponding ones of Methods of Moments and Quasi-Maximum likelihood algorithms. In [37] it is shown that in case of model like Bates, PF algorithms show the best performances, among the alternatives discussed, in getting inference both for the latent processes dynamics and the model parameters.

CHAPTER 3

Historical and Risk-Neutral Estimation in a Commodity Market with a two factor Stochastic Volatility model

3.1 Introduction

Stochastic volatility models are a well-known choice in commodity markets, Geman [28] and Hikspoors [36] discussed some of the most popular models of this kind in commodity finance. The seminal paper, to which they refer is the one by Schwartz [53], where the commodity spot values are modeled by a mean reverting process and the convenience yield is incorporated in the discount factor. While Cortazar and Schwartz [16] proposed an extension of the original Schwartz model adding one factor to describe the long-term interest rate by an Ornstein-Uhlenbeck process, Eydeland and Geman [23] proposed a different extension adding a mean reverting OU-process to describe the instantaneous variance of the diffusion term in the spot dynamics.
As Geman [27] pointed out, the use of mean reversion in the dynamics of spot values is controversial, in particular for crude oil market (the object of our analysis) it is not always needed. In the same work it was suggested a simple test to discriminate if a mean reverting dynamics is preferable to a simple diffusion process, the results shows that for time series starting from 2005 there is no strong evidence for a mean reverting modeling choice.

Further developments were implemented, as in Ribeiro and Hodges [50], to describe the dynamics for the oil sector of commodity market. Moreover in this paper (Ribeiro and Hodges [50]), it is also pointed out that the possibility for convenience yield to assume negative values (to explain the contango effect in commodity market) from a theoretical viewpoint is due to the cost of storage implied in prices dynamics; if this term was separated off the convenience yield, then a Cox-Ingersoll-Ross process for the convenience yield would automatically exclude arbitrage opportunities in the market. In the model proposed in this article, we stand on this remark by Ribeiro and Hodges, modeling the convenience yield (once the cost of storage is separated off) with a CIR dynamics but without imposing any linear dependence of the spot dynamics diffusion term on the convenience yield; on the contrary, we prefer to include an extra volatility factor (with a CIR dynamics, like in the well known Heston model[1]). The possibility of jumps with finite activity (modeled by a compound Poisson process with normally distributed jump size) is allowed in a second version of the model, which has been tested with the same data set used for inference in the model without jumps. A third version of the model substitutes the jump term with a seasonality term (modeled by the usual periodic function, as suggested also by Hikspoors [36]), allowing for a comparison between the effect of jumps and of the seasonality term in the goodness of fit of analyzed data set. Data set used for the estimation come from the WTI index spot values, quoted in NYMEX market, and the futures written on it, more details about data sets are given in a dedicated section. The time window considered spans from $01/02/2007$ to $31/12/2010$.
Inference under both the historical and the risk-neutral measures are quite common in literature for models of this kind, and they could be easily extended to the present one. Some authors resorted to Bayesian analysis to get inference in stochastic volatility framework: since the seminal paper by Jacquier, Polson and Rossi [20] a fast-growing literature is exploring the application of Bayesian estimation techniques, in particular MCMC methods, to get inference for models belonging to this family. A branch of this literature is devoting to study joint estimation under historical and risk neutral measure, using both stock and derivatives prices data. The link between the two measures is provided by a suitable parametric choice of the Radon-Nikodym derivative.
The main references in this section are the two papers by Eraker [22] and by Eraker, Johannes and Polson [8] in which some popular stochastic volatility models are analyzed using a Gibbs sampling algorithm. Further references in which the Gibbs sampling method is used to get inference for stochastic volatility models, eventually including jumps, are the papers by Forbes, Martin and Wright [25] and, more recently, by Yu, Li and Wells [12], who extended the results by Eraker [22] including different jump models in the analysis and compare them; all these

[1]Together with the model we present a closed formula for future prices in Section4.3 and the prices for option written on futures in Section4.9

paper refer to equity markets, analyzing S&P500 or DAX data. Another Bayesian technique which is becoming popular to get inference in a stochastic volatility framework, and when latent factor are present in general, is the particle filter (PF) method. This is a bayesian filter algorithm based on sequential importance sampling for bayesian networks; differently from Kalman filter, it can be used also for non-gaussian and non-linear dynamics. A complete survey on the theoretical background and implementation details on particle filters are the papers by Andrieu and Johansen [7], and Arulampalam et al. [42], while for application to stochastic volatility models Javaheri [37], Johannes, Polson an Stroud [43], Aihara [55]. Since the success of PF techniques, different authors have worked on efficient estimation with Particle filters, and conjunction of MCMC and PF algorithms; this culminated in a paper by Andrieu, Doucet and Holestein [10], where have been introduced the particle Markov Chain Monte Carlo algorithms that implement for inference on parameter set of a model an MCMC where the marginal likelihood is estimated by a nested Auxiliary particle filter. We have adopted an algorithm belonging to this family to carry out inference on the model chosen.

3.2 The Models Proposed

We studied three possible different variants of a basic model (Model B) consisting in a two-factor model, including a volatility process and a convenience yield process into spot price dynamics. Both the spot variance and the convenience yield processes follow a CIR dynamics. Under the historical measure $\mathbb{P}$ the dynamics of the model B is the following:

$$\begin{cases} \dfrac{dS_t}{S_t} & = (\mu + c - \delta_t)dt + \sqrt{V_t}dW_{S_t}^{(\mathbb{P})} \\ d\delta_t & = \alpha(\bar{\delta} - \delta_t)dt + \sigma\sqrt{\delta_t}dW_{\delta_t}^{(\mathbb{P})} \\ dV_t & = \beta(\bar{V} - V_t)dt + \xi\sqrt{V_t}dW_{V_t}^{(\mathbb{P})} \\ dW_{S_t}^{(\mathbb{P})} & dW_{V_t}^{(\mathbb{P})} = \rho dt \end{cases} \tag{3.1}$$

In (3.1) we used the standard notation: we identified the spot value at time t with S_t, with δ_t it is represented the convenience yield process (once we separated it from the cost of storage and insurance, c) and with V_t the variance process. $W_{S_t},\ W_{\delta_t},\ W_{V_t}$ are 3 Brownian motion, with W_{δ_t} independent from the other two process, while W_{S_t} and W_{V_t} are correlated as written in (3.1).
Besides model B, we considered the same model allowing for jumps (Model J) in the spot dynamics (modeled with jump arrivals time poisson distributed and lognormally distributed jump size), and the basic model with a seasonality term in spot dynamics (Model S). The model with seasonality includes in spot dynamics the usual sinusoidal term (like discussed also by Hikspoors [36]) The only

stochastic differential equation changing in the system dynamics moving from one model to the other is the spot equation. In the following system we introduce a general dynamics (from which we can easily obtain our three models):

$$\begin{cases} dx_t & = (\mu + c - \delta_t - \frac{1}{2}V_t)dt + \sqrt{V_t}dW_{x_t}^{(\mathbb{P})} + dJ_{x_t}^{(\mathbb{P})} \\ d\delta_t & = \alpha(\bar{\delta} - \delta_t)dt + \sigma\sqrt{\delta_t}dW_{\delta_t}^{(\mathbb{P})} \\ dV_t & = \beta(\bar{V} - V_t)dt + \xi\sqrt{V_t}dW_{V_t}^{(\mathbb{P})} \\ S_t & = \exp\{g(t_{\text{year}}) + x_t\} \\ g(t_{\text{year}}) & = \zeta_1(sen(2\pi t_{\text{year}} + \omega_1)) + \zeta_2(cos(2\pi t_{\text{year}} + \omega_2)) \\ dW_{x_t}^{(\mathbb{P})} & dW_{V_t}^{(\mathbb{P})} = \rho dt \end{cases} \tag{3.2}$$

with t_{year} the number of days from the first of January of the same year divided by 365. $J_{x_t}^{(\mathbb{P})}$ is a jump process, where the number of jump arrivals on a given time are poisson distributed with λ_j parameter, and the jump size are normally distributed with mean μ_J and variance σ_J^2. We make the usual choice of describing the price dynamics through the log-price process: $x_t = \log[S_t]$. We decided to keep the rates not stochastic and we fix them at the FED levels. As already discussed by Schwartz [53], comparing performances in capturing the futures observed structure for models with and without stochastic interest rate, the performances in using stochastic interest rates do not show any significant improvement, if we exclude the longest maturity futures (about ten-years maturity). Hence, accordingly our longest maturity considered is 5 years futures, we preferred to do not allow stochastic interest rates to make the model more parsimonious.
From (3.2) we can obtain the three models fixing some parameters

- Model B : we fix λ_J, the parameter of Poisson-distributed jump time arrivals, to zero and $\zeta_1 = \zeta_2 = 0$
- Model J : we fix $\zeta_1 = \zeta_2 = 0$
- Model S : we fix $\lambda_J = 0$

In order to describe the dynamics under the risk-neutral measure, we need to define the Radon-Nikodym derivative of this measure with respect to the historical one. The parametric form we choose is the same proposed by Heston[35] and by Pan [45], which preserves the model structure under the measure change. This can be specified by the following relations between the Wiener processes under the two measures, provided by the Girsanov theorem:

$$\begin{cases} dW_{\delta_t}^{(\mathbb{Q})} & = dW_{\delta_t}^{(\mathbb{P})} - \frac{\eta_\delta}{\sigma}\sqrt{\delta_t}dt \\ dW_{V_t}^{(\mathbb{Q})} & = dW_{V_t}^{(\mathbb{P})} - \frac{1}{\sqrt{1-\rho^2}}\left(\rho\eta_{S_t} + \frac{\eta_V}{\xi}\right)\sqrt{V_t}dt \\ dW_{x_t}^{(\mathbb{Q})} & = dW_{x_t}^{(\mathbb{P})} + \eta_{S_t}\sqrt{V_t}dt \end{cases} \tag{3.3}$$

Where we have denoted by η_S the risk premium associated with the Brownian part of the spot dynamics.
Hence, we can write the dynamics just described under the measure $\mathbb{P}$ by the system (3.2), also under the risk-neutral measure $\mathbb{Q}$[2] :

$$\begin{cases} dx_t & = (r + c - \delta_t - \frac{1}{2}V_t - \mu^*)dt + \sqrt{V_t}dW_{x_t}^{(\mathbb{Q})} + dJ_{x_t}^{(\mathbb{Q})} \\ d\delta_t & = \left(\alpha(\bar{\delta} - \delta_t) + \eta_\delta \delta_t\right) dt + \sigma\sqrt{\delta_t}dW_{\delta_t}^{(\mathbb{Q})} \\ dV_t & = \left(\beta(\bar{V} - V_t) + \eta_V V_t\right) dt + \xi\sqrt{V_t}dW_{V_t}^{(\mathbb{Q})} \\ S_t & = \exp\{g(t_{\text{year}}) + x_t\} \\ dW_{S_t}^{(\mathbb{Q})} & dW_{V_t}^{(\mathbb{Q})} = \rho dt, \end{cases} \tag{3.4}$$

where μ^* and $dJ^{(\mathbb{Q})}$ denote respectively the predictable compensator and the compensated jump measure. It is worth remarking that, since the cost of storage is separated off the dynamics of the convenience yield, δ_t this is modeled by a CIR process that prevents it from assuming negative values, automatically excluding arbitrage opportunities in the market, as pointed out in Ribeiro and Hodges [50]. Moreover, since $\delta_t - c$ can assume both positive and negative values, normal and inverted futures market structures are both allowed.
To cope with the inference in this setting we use a Euler discretization method. The discretized model (under the risk neutral measure) can be written as follows:

$$\begin{cases} x_{t+1} & = x_t + (r + c - \delta_t - \frac{1}{2}V_t - \mu^*)\Delta t + \sqrt{1-\rho^2}\sqrt{V_t}\epsilon_{t+1}^{(S)} + \rho\sqrt{V_t}\epsilon_{t+1}^{(V)} + \sum_{i=1}^{N_{t+1}^{(J)}} \epsilon_{i,t+1}^{(J)} \\ \delta_{t+1} & = \delta_t + \left(\alpha(\bar{\delta} - \delta_t) + \eta_\delta \delta_t\right) dt + \sigma\sqrt{\delta_t}\epsilon_{t+1}^{(\delta)} \\ V_{t+1} & = V_t + \left(\beta(\bar{V} - V_t) + \eta_V V_t\right) dt + \xi\sqrt{V_t}\epsilon_{t+1}^{(V)} \end{cases} \tag{3.5}$$

Where each $\epsilon_{t+1}^{(S)}$, $\epsilon_{t+1}^{(V)}$ and $\epsilon_{t+1}^{(\delta)}$ are normally distributed random variables with zero mean and variance Δt. In the discretized jump addend, the $N_{t+1}^{(J)}$ are independent Poisson distributed r.v. with parameter λ_J, while $\epsilon_{i,t}^{(J)}$ are independent normal r.v. with mean $(\mu_J - \eta_J)$ and variance σ_J^2. All the random variables just listed are independent on each other.
Under the historical measure the same discretization holds, provided that the risk premia $\{\eta_\delta,\ \eta_V,\ \eta_J\}$ and the compensator μ^* are set to zero, while the drift coefficient r_f has to be substituted by μ.
The futures price is given by the following expression:

$$F(0, \tau) = \exp\{A_0(\tau) + x_t + A_2(\tau)\delta_t\} \tag{3.6}$$

where $\tau = T - t$ is the futures time to maturity. Details about computations and the specifics of the functions $A_0(\tau)$ and $A_2(\tau)$ are provided in Section4.3. In case

[2]for the basic model we just fix to zero μ^* and $dJ_{x_t}^{(\mathbb{Q})}$, for the model with seasonality we add the seasonality factor ($g(t_{\text{year}})$) to the spot dynamics like in (3.2)

we consider model allowing for seasonality, the future formula slightly changes; it becomes:

$$F(0,\tau) = \exp\{g(T_{\text{year}}) + A_0(\tau) + x_t + A_2(\tau)\delta_t\} \tag{3.7}$$

where T_{year} is the time in years the maturity day differ from the first of January of the same year.
Since the prices of a futures are affected by different kind of noises (possible incomplete specification of the model, market inefficiency, random noise, etc) we modeled the logarithm of the price of the futures making the hypothesis that the market price, $\ln F^{\text{M}}(0,\tau)$, are represented by the theoretical price got by (4.8) plus an error distributed as a white noise $\epsilon_{\text{fut}} \sim \mathcal{N}(0,\sigma_\epsilon^2)$[3]:

$$\ln F^{\text{M}}(0,\tau) = \ln F(0,\tau) + \epsilon_{\text{fut}} \tag{3.8}$$

3.3 Futures Prices

When the underlying follows, under the the dynamics $\mathbb{Q}$ (4.6) the Kolmogorov backward equation for the generic contract value, $f(t,x_t,\delta_t,V_t,J_t)$, at time t, when the underlying follows, under the $\mathcal{Q}$ measure, the dynamics (4.6) is:

$$\begin{aligned}
&\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}(r_f + c - \lambda\mu_J^* - \delta_t - \frac{1}{2}V_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}V_t + \\
&\quad + \frac{\partial f}{\partial \delta}\left[\alpha(\bar{\delta} - \delta_t) - \eta_\delta\delta_t\right] + \frac{1}{2}\frac{\partial^2 f}{\partial \delta^2}\sigma^2\delta_t + \\
&\quad + \frac{\partial f}{\partial V}\left[\beta(\bar{V} - V_t) - \eta_V V_t\right] + \frac{1}{2}\frac{\partial^2 f}{\partial V^2}\xi^2 V_t + \\
&\quad + \frac{\partial^2 f}{\partial V \partial x}\rho\xi V_t + \\
&\quad + \lambda\mathbb{E}\left[f(t, x_t + ln(1+J), \delta_t, V_t) - f(t,x_t,\delta_t,V_t)\right] = 0
\end{aligned} \tag{3.9}$$

The PDE (4.10) has to be solved with the usual terminal condition $f(t=T) = H(x_T,\delta_T,V_T)$ with H payoff at time T maturity, and since we are dealing with futures the final payoff is:

$$H(x_T,\delta_T,V_T) = \exp\{x_T\} \tag{3.10}$$

We make the hypothesis of a solution form:

$$f_t = \exp\{A_0(t) + A_1(t)x_t + A_2(t)\delta_t + A_3(t)V_t\}$$

[3] $\sigma_\epsilon^2 = 0,08$ is consistent with the value observed for the mean of the daily spread of the highest-lowest prices for futures in the considered range

if we try this guess solution into (4.10), we get the follow ODE system:

$$\begin{cases} - & \dfrac{\partial A_0(\tau)}{\partial \tau} + A_1(\tau)(r_f + c) + A_2(\tau)\alpha\bar{\delta} + A_3\beta\bar{V} = 0 \\ - & \dfrac{\partial A_1(\tau)}{\partial \tau} = 0 \\ - & \dfrac{\partial A_2(\tau)}{\partial \tau} - A_1(\tau) - A_2(\tau) + \frac{1}{2}A_2^2(\tau)\sigma^2 = 0 \\ - & \dfrac{\partial A_3(\tau)}{\partial \tau} - \frac{1}{2}A_1(\tau) + \frac{1}{2}A_1^2(\tau) - A_3(\tau)(\beta + \eta_V) + \frac{1}{2}A_3^2(\tau)\xi^2 + A_1(\tau)A_3(\tau)\rho\xi = 0 \end{cases} \tag{3.11}$$

where we changed variable from t to $\tau = T - t$ [4] .
Since the payoff is (4.11) we obtain the initial condition for the ODE system:

$$\begin{cases} A_0(0) & = 0 \\ A_1(0) & = 1 \\ A_2(0) & = 0 \\ A_3(0) & = 0 \end{cases} \tag{3.12}$$

Solving previous system we get :

$$\begin{cases} A_0(\tau) & = (r_f + c)\tau - \dfrac{2\alpha\bar{\delta}}{\sigma^2}\left[B\tau + \log\left(\dfrac{D - B\exp\{C\tau\}}{D - B}\right)\right] \\ A_1(\tau) & = 1 \\ A_2(\tau) & = -\frac{2}{\sigma^2}\dfrac{\exp\{C\tau\} - 1}{\frac{\exp\{C\tau\}}{D} - \frac{1}{B}} \\ A_3(\tau) & = 0 \end{cases} \tag{3.13}$$

with:

$$\begin{cases} C & = \sqrt{(\alpha + \eta_\delta)^2 + 2\sigma^2} \\ D & = \dfrac{(\eta_\delta + \alpha) + C}{2} \\ B & = \dfrac{(\eta_\delta + \alpha) - C}{2} \end{cases} \tag{3.14}$$

3.4 The Data

The data used for the analysis are relative to WTI Cushing Crude Oil spot and futures quotations on NYMEX market from $1/02/2007$ to $28/09/2011$. We used the the first 85% if the data-set to get the inference on our model and reserved the last 15% of dates for out of the sample performances study. The range of

[4] Note that in the Futures price obtained there is no dependence on the jump process parameters in strict analogy with Yan [59].

dates for the estimation goes from $1/02/2007$ to $31/12/2010$, and the spot data are presented in Figure 3.1.
Spot data are collected from the US Energy information administration website where a large collection of energy related time series (among which the WTI FOB spot prices) is provided. Daily closing data are taken into account for any available working day in the interval. A plot of the spot data used in estimating parameter set of the different models analyzed is in figure Fig.3.1

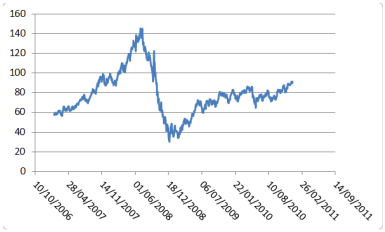

Figure 3.1: *Spot FOB data for WTI crude oil*

Besides spot data, for any working day we recorded a panel of 12 future contract values at opening time. Their maturity day is fixed on the first working day of each month of 2012. So for any trading day we analyzed a spot datum and 12 futures data, and in the range there are 988 dates. Hence the data set consists in 988 spot values and 11856 future contracts.
In addition to this data set we reserved a panel of data to evaluate out of the sample performances. The data set include again a FOB spot datum and a panel of 12 future data (with different maturity one for each month of 2012, the maturities are set on the first trading day), the range of dates goes from $01/01/2011$ to $28/09/2012$. The number of working days in the range is 187.
Future contracts are usually characterized by their behaviour when time to maturity goes to zero. If at a certain date the futures quotation increases when the time to maturity become longer, it usually said the future market is normal, otherwise is said inverted. In figure Fig.4.2 are indicate the dates for which we can observe a normal future market (equivalent to the value $+1$) and the dates with an inverted future market(equivalent to the value -1).

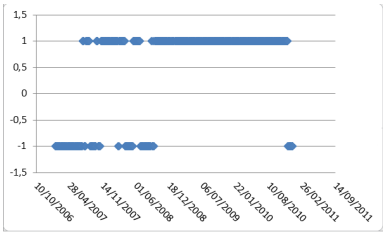

Figure 3.2: *The dates in the data set used for the parameter estimation are divided in normal future market days ($+1$ value) and inverted future market days (-1 value)*

In the analyzed period and in the data set used to evaluate out of the sample performances are present both the future market behaviour. The two behaviour are shown in the figures Fig.3.3 and Fig.3.4 and illustrate, respectively, the future market structure at the date $11/05/2011$ and at $25/08/2011$.

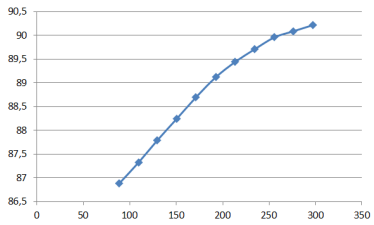

Figure 3.3: *Normal future market: WTI future quotations* 11 *May* 2011*, different maturities. On x-axis are the days to delivery*

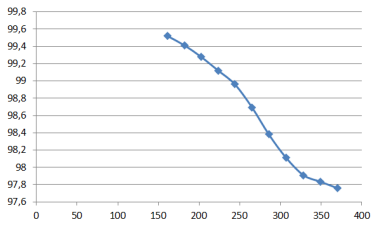

Figure 3.4: *Inverted future market: WTI future quotations* 25 *August* 2011*, different maturities. On x-axis are the days to delivery*

From the available range of dates we extracted a second data-set from $11/12/2010$ to $28/09/2011$; again we used about the first 85% of the dates to get inference and the remaining part to evaluate out of the sample performances evaluation. The data set used for the inference goes from $11/12/2010$ to $30/06/2011$. The spot quotations are reported in the Figure 4.1, in this reduced dataset it is excluded the period from June 2008 to December 2008, when a disastrous fall in FOB quotations was observed in the market (the barrel price went from about $150\$$ to about $30\$$). In Figure 3.6 are reported, as done for the previous dataset, with dates characterized by a normal future market structure and with dates characterized by an inverse future structure. We excluded the model with seasonality from the analyzed models, since the dates range spans just one year and a half, hence any inference on yearly periodic functions would be affected by the too short range considered. The number of working dates in this range is 350 and, as before, we considered for each date one spot FOB quotation and 12 futures for each date, hence our inference data-set consists in 350 spot data and 4200 futures data. The out of the sample performances are evaluated on the data-set consisting in 81 spot data and 972 futures data, these data correspond to the range of dates from $01/07/2011$ to $28/09/2011$.

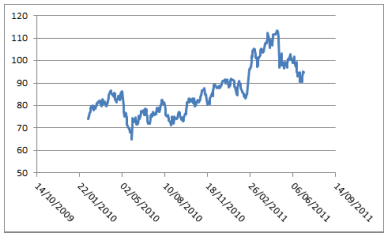

Figure 3.5: *Spot FOB data for WTI crude oil for the reduced dataset*

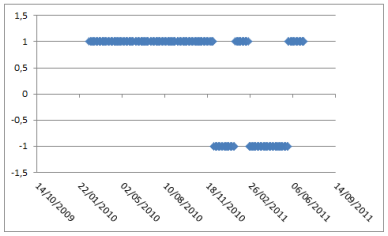

Figure 3.6: *The dates in the data set used for the parameter estimation are divided in normal future market days ($+1$ value) and inverted future market days (-1 value)*

3.5 Inference Algorithm

The reference object of our inference is the vector $\{\Theta, V_{0:T}, \delta_{0:T}, J_{0:T}\}$ where $V_{0:T}$, $\delta_{0:T}$, $J_{0:T}$ are the three latent processes (the variance process, the convenience yield and the jump process) and Θ is the set of parameters, that is our main inference target:

$$\Theta = \{\epsilon, \mu, c, \alpha, \bar{\delta}, \sigma, \eta_\delta, \beta, \bar{V}, \xi, \rho, \lambda, \mu_J, \sigma_J\}$$

[5]. For simplicity we shall indicate with $X_{0:T}$ the vector of the three latent processes $\{V_{0:T}, \delta_{0:T}, J_{0:T}\}$, and by $Z_{1:T}$ the set of observed market data (coming from log return time series, $Y_{1:T}$, and futures price quotations $F_{1:T}$).
To make inference in a so large state space, Particle Markov Chain Monte Carlo methods (from now on PMMC) represents an efficient technique, since it allows to simulate the latent processes in a single simulation block. The PMMC allows to sample from $p(\Theta, X_{0:T}|Z_{1:T})$, that is the joint probability distribution of the parameter set and the latent processes given the observed data. To sample with a Monte Carlo technique from such a probability distribution we make use of a particle filter algorithm[6] to estimate the marginal likelihood $L(z_{1:T}|\Theta)$, this will

[5] in the case of the model with seasonality we replace the parameters referring to jumps($\{\lambda, \mu_J, \sigma_J\}$) with the parameters referring to seasonality: $\{\zeta_1, \zeta_2, \omega_1, \omega_2\}$

[6] Details about the generic Particle Filter algorithm and probability distributions used in the simulation we conducted are in Section3.6.

be used in defining an acceptance ratio probability that will ensure that after a certain amount of time, whatever point in state space we started from, we will sample from the right distribution, getting unbiased estimate for Θ.
The MCMC method used is a variant of the well known Metropolis-Hastings algorithm

- At step (s=0)
  after setting a starting point $\Theta = \theta(0)$ arbitrarily, then a Sampling Important Resampling algorithm is implemented. SIR algorithm allows to simulate the latent processes $x_{1:T}^{(0)}$ from the distribution $p(X_{1:T}|Z_{1:T}, \theta(0))$ and an estimate of the marginal likelihood $p(Z_{1:T}|\theta(0))$

- At step (s from 1 to MC, the length of the Markov chain sequence we want to simulate)
  a new value set, θ^*, is sampled from a proposal (symmetric) distribution $q(\cdot|\theta(s-1))$ and the new sampling value is accepted with a probability:

$$\min\left(1,\ \frac{p(z_{1:T}|\theta^*)\tilde{p}(\theta^*)}{p(z_{1:T}|\theta(s-1))\tilde{p}(\theta(s-1))}\right)$$

  where $\tilde{p}(\cdot)$ is the prior distribution, while $p(z_{1:T}|\theta^*)$ is the likelihood for the observation chain given the parameter set θ^*.
  Again the marginal likelihood and $x_{1:T}^*$ comes form a PF algorithm run.
  If accepted: $\theta(s) = \theta^*$, otherwise: $\theta(s) = \theta(s-1)$.

According to this algorithm we are sampling from a Markov chain whose limiting distribution is:

$$p(\Theta,\ X_{1:T}|\ Z_{1:T}) = p(\theta|z_{1:T})p_\theta(x_{1:T}|\ z_{1:T})$$

We can also discard the information about hidden state process and use the sample to get inference about θ, if we need just the last one.
To reduce the number of rejected proposal, and increase mixing of the chain different techniques are known, as Metropolis within Gibbs variant, that is the one we implemented in our sampling algorithm. The output coming from the PMMH (particle Metropolis Hastings algorithm), as any MCMC output need to be resized, removing the burn in, that is the part of the chain needed by algorithm to get to the stationary distribution. To check that convergence to stationary distribution was reached (for all the parameters) we adopted the Geweke test [29].

3.6 Particle Filter

We implemented the ASIR algorithm proposed by Pitt and Shephard [40] to generalize the standard sampling important resampling algorithms, attempting to

mitigate the sample impoverishment effect, sampling the ancestor of the particle at each time step, adapting the proposed "particle" sample to the new incoming information. The starting point to define the algorithm is the discretized dynamics of the model (4.7) adding the future formula (3.7).
First of all a remark about notations: since the spot data are collected at the end of the day and the future prices at the opening, with V_t, δ_t, J_t we identify respectively the value of the variance and the convenience yield processes at the same time when the future prices are registered, while the J_t identify the jumps occurring within the day till the spot value is observed. The discrepancy between the time of the future observation and the spot one is one other source of error we take into account in the future error with which we observe the futures quotations in the market.
Let us consider the time t a sample for the latent process $\{\delta_{t-1}, V_{t-1}\}$, that is: $\{\delta_{t-1}^{(i)}, V_{t-1}^{(i)}\}$ with associated weights $\pi_{t-1}^{(i)}$, where $i = 1, \dots N_s$ where N_s is the number of particles. Before propagating the particles,that is obtaining a sample for the latent processes at time t we select the ancestor for this new sample. In practise, we define an expected value for the latent process sample at time t: $\{\bar{\delta}_t^{(i)}, \bar{V}_t^{(i)}\}$ and use this estimate to attach a weight function $\omega_{t|t-1}^{(i)}$ to the particles and select the best fitting the new information incoming:

$$\begin{cases} \bar{\delta}_t^{(i)} & = \alpha(\bar{\delta} - \delta_{t-1}^{(i)})\Delta t \\ \bar{V}_t^{(i)} & = \beta(\bar{V} - V_{t-1}^{(i)})\Delta t \\ \omega_{t|t-1}^{(i)} & = \sum_k \phi\left(y_t | m_{y_t}^{(i)} + k\mu_j, \Sigma_{y_t}^{(i)} + k\sigma_j^2\right) \frac{\lambda_j^k e^{-\lambda_j}}{k!} \pi_t^{(i)} \\ m_{y_t}^{(i)} & = (r + c - \bar{\delta}_t^{(i)} - \frac{1}{2}\bar{V}_t^{(i)})\Delta t \\ \Sigma_{y_t}^{(i)} & = \bar{V}_t^{(i)}\Delta t \end{cases} \tag{3.15}$$

Hence we get a N_s long sample $\{\delta_{t-1}^{(s)}, V_{t-1}^{(s)}\}$ from the distribution $\{\delta_{t-1}^{(i)}, V_{t-1}^{(i)}\}$ associated to probability weights $\dfrac{\omega_{t|t-1}^{(i)}}{\sum_i \omega_{t|t-1}^{(i)}}$. This are our selected ancestors, we will propagate using the (4.7). First we simulate the jumps. The number of jumps N_t^j happening in the time passing from $t-1$ to t is distributed according to a Poisson with parameter λ_j:

$$N_t^{(s)} = k \text{ with probability } q(N_t^{j(s)}) = \phi\left(y_t | m_{y_t}^{(s)} + k\mu_j, \Sigma_{y_t}^{(s)} + k\sigma_j^2\right) \frac{\lambda_j^k e^{-\lambda_j}}{k!}$$

For numerical purposes we need to restrict to a finite number of possible jumps, since the time interval consists in one day, we can fix the maximum number of jumps at 1.
Hence, the jump amplitude is simulated from a normal, as suggested in [43] we

incorporate the new return information in our sampling:

$$J_t^{(s)} \sim q(J_t|y_t) = \mathcal{N}\left(N_t^{(s)}\mu_j + N_t^j \frac{\sigma_j^2}{\bar{V}_t^{(s)}\Delta t}\left(y_t - (r + c + (\eta_S - \frac{1}{2})\bar{V}_t - \bar{\delta}_t^{(s)})\right), N_t^{(s)}\sigma_j^2\left(1 - N_t^{(s)}\frac{\sigma_j^2}{\bar{V}_t^{(s)}\Delta t}\right)\right)$$

Once we have a sample for the jumps, we sample the value of the "particles" for the variance and the convenience yield:

$$\begin{cases} \delta_t^{(s)} & \sim p(\delta_t^{(s)}|\delta_{t-1}^{(s)}) = \mathcal{N}\left(\bar{\delta}_t^{(s)}, \sigma^2\delta_{t-1}^{(s)}\Delta t\right) \\ V_t^{(s)} & \sim p(V_t^{(s)}|V_{t-1}^{(s)}, \delta_{t-1}^{(s)}, y_t, J_t) = \\ & = \mathcal{N}\left(\bar{V}_t^{(s)} + (y_t - (r + c - \bar{\delta}_t^{(s)} - \frac{1}{2}\bar{V}_{t-1}^{(s)})\Delta t - J(s)_t)\xi\sqrt{1-\rho^2}, \rho^2\xi^2 V_{t-1}^{(s)}\Delta t\right) \end{cases}$$

Hence we compute the new weights $\pi_t^{(s)}$:

$$\begin{cases} \pi_t^{(s)} & = \dfrac{\pi_{t-1}^{(s)}}{\omega_{t|t-1}^{(s)}} \dfrac{\lambda_j^{N_t^{(s)}} e^{-\lambda_j}}{N_t^{(s)}!} \dfrac{1}{q(N_t^{(s)}|y_t)} \\ & \dfrac{\phi(J_t^{(s)}|\mu_j, \sigma_j^2)}{q(J_t^{(s)}|y_t)} \phi(y_t|(r + c - \delta_t^{(s)} - \frac{1}{2}V_{t-1}^{(s)})\Delta t + J_t^{(s)}, V_{t-1}^{(s)}\Delta t) \end{cases} \tag{3.16}$$

where $\epsilon_t^{(V,s)}$ is the random value got for $\epsilon_t^{(V)}$ and corresponding to the s-th particle An unbiased [40] estimate for the marginal likelihood of the observation given the parameter set is:

$$\begin{cases} \hat{p}(y_{1:T}, F_{1:T}) & = \hat{p}(y_1, F_1)\prod_{t=2}^T \hat{p}(y_t, F_t|y_{t-1}, F_{t-1}) \\ \hat{p}(y_t, F_t|y_{t-1}, F_{t-1}) & = \left(\sum_{s=1}^{N_s} \dfrac{\pi_t^{(s)}}{N_s}\right)\left(\sum_{s=1}^{N_s} \omega_{t|t-1}^{(s)}\right) \end{cases} \tag{3.17}$$

At this stage we can add an other re-sample step from the particles with weights $\pi_t^{(s)}$, conditional on a measure of the impoverishment of the sample, if we resample we have to set all the weights equal to $\dfrac{1}{N_s}$.
To obtain the algorithm for the model without jumps it is sufficient to set all $N_t^{(s)} = 0$ for all t in the algorithm presented in this section.

3.7 Numerical Results

To conduct our inference for each of our models discussed in section 4.2, we simulated Markov chains of 30000 steps (from these we discard about 10000 sampled

points as burn-in of the Markov chain[7]). We used the mean of the posterior distribution $p(\Theta|Z_{0:T})$ as point estimate for each parameter; these results are reported in the Table 4.1. Together with means we reported also the standard deviation of the posterior distributions.

Table 3.1: *Parameter inference: posterior means (and posterior standard deviation) of the model parameter set Θ for three analyzed models. All values are expressed on a daily basis and are scaled by a 100 factor*

Model Parameters	Basic Model	Model with Seasonality	Model with Jumps
η_S	0.0892	0.0859	0.1041
	(0.0069)	(0.0717)	(0.0046)
c	0.2043	0.1972	0.1895
	(0.0061)	(0.0943)	(0.00257)
α	0.0942	0.1242	0.4633
	(0.0415)	(0.0651)	(0.0652)
$\bar{\delta}$	0.1977	0.1965	0.1868
	$(1.10E-3)$	(0.0058)	$(5.25E-4)$
σ	0.0511	0.0694	0.27
	(0.0061)	(0.0226)	(0.0021)
η_δ	0.0293	0.0256	0.0357
	$(6.95E-4)$	(0.0035)	$(3.96E-4)$
β	0.5294	1.4486	0.7909
	(0.2811)	(0.8001)	(0.5878)
$\bar{V}$	0.0429	0.0392	0.0235
	$(7.22E-4)$	(0.0092)	$(5.61E-4)$
ξ	0.2616	0.3405	0.155
	(0.0666)	(0.1159)	(0.015)
ρ	-64.44	-33.06	-55.98
	(2.0302)	(16.60)	(0.64)
ζ_1	$-$	0.6712	$-$
	$(-)$	(0.2534)	$(-)$
ζ_2	$-$	-0.7275	$-$
	$(-)$	(0.1607)	$(-)$
ω_1	$-$	2.6048	$-$
	$(-)$	(2.2499)	$(-)$
ω_2	$-$	3.4644	$-$
	$(-)$	(4.2490)	$(-)$
λ_J	$-$	$-$	5.689
	$(-)$	$(-)$	(0.2001)
μ_J	$-$	$-$	-0.7241
	$(-)$	$(-)$	$(8.91E-3)$
σ_J	$-$	$-$	7.373
	$(-)$	$(-)$	(0.1594)

[7]The lack of convergence has been tested by Geweke test, and at 1% significance we can reject the hypothesis of not convergence of the Markov Chain. For any further detail about test, estimation error and algorithm used to compute them we refer to the manual written by Smith [56].

Since for all the three models the cost of storage (c) estimate is greater than long run convenience yield ($\bar{\delta}$) one, then $c - \delta_t$ assumes often positive values, which reflects in a normal market effect in futures structure, as it is present for most of the dates in the considered period. From Fig.3.7 and Fig.3.8 we can observe that the estimate for δ_t is greater than the estimate for c in the same periods when in the markets are observed an inverted future market, as it is possible to see comparing Fig.4.2 and Fig.3.6 with Fig.4.2.
The inference results show for all the models a low level in long-run volatility, for the basic model the estimated value for $\bar{V}$ is 0.0429 (corresponding to an annualized long-run volatility close to 3.5%), a value that well represent the historical standard deviation we recognize in the return time series analyzed, if we exclude the period from June 2008 to December 2008, in which spot quotation faced a dramatic fall resulting in a sudden increase in volatility level. As expected the model with seasonality show similar values to the basic model for the volatility process parameters, since the seasonality function do not affect the volatility dynamics, just the log-spot one. The main difference among this two models is in the correlation parameter, in model with seasonality we found a lower value for correlation. The posterior for the parameters $\zeta_1,\ \zeta_2,\ \omega_1,\ \omega_2$ show mean values close to zero, and higher variances with respect to the others parameter posteriors. Comparing the results of the inference carried out with the basic model and with the model that allows the spot dynamics to jump we can observe similar point estimates for all the parameters involved in the spot and in the convenience yield SDE, the differences are in the parameter estimates for the volatility process, where we can observe in particular a lower value for the long-run volatility in the case with jumps with respect to the basic model; obviously, this is the effect of jumps that absorb partially the volatility of the spot dynamics. The intensity of jumps has a value significantly different from zero, the point estimate suggest an average of 5.6% days with jumps.
Together with inference about the parameters we got also inference on the path of the latent processes (convenience yield and variance process). In the following figures (Fig.3.7, Fig.3.8 and Fig.4.5) are shown the inference on the latent processes for the different models, the blue line represent the mean of all path, that we can interpret as our estimate for the process. In all the dynamics we can notice how during the period from June 2008 to December 2008, the filtered dynamics behaves like in presence of outliers. This will have an impact also in the diagnostics performances of the inference, but comparing Table 4.1 and Table 3.2(inference obtained using the two different data sets, including and excluding the discussed period) we can see that it does not impact heavily in the parameter estimates.

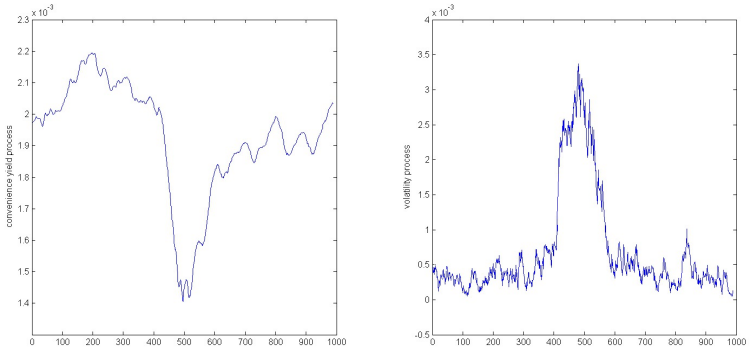

Figure 3.7: *Inference on dynamics for the convenience yield and the volatility process got under the model without seasonality or jumps*

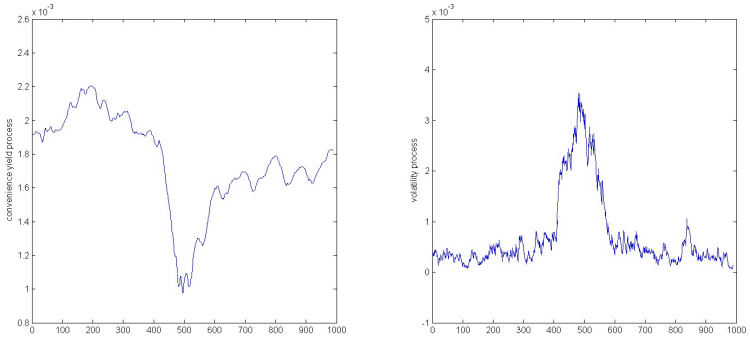

Figure 3.8: *Inference on dynamics for the convenience yield and the volatility process got under the model with seasonality*

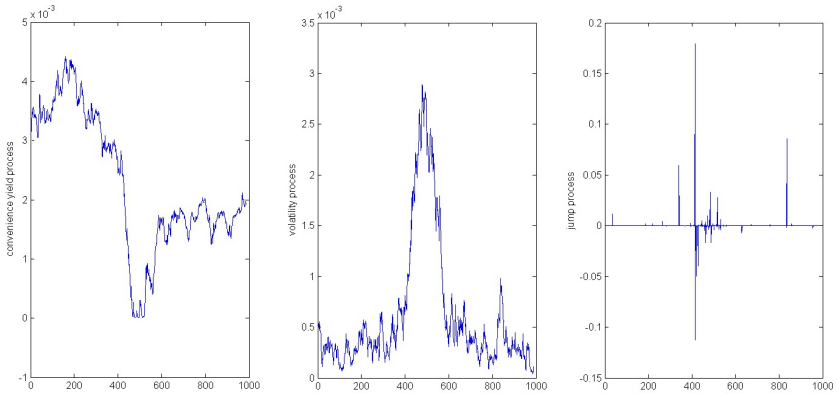

Figure 3.9: *Inference on dynamics for the convenience yield and the volatility process got under the model with jumps*

It is interesting to notice when the filtered process for the convenience yield is above the "threshold" represented by the estimated cost of storage plus the risk free rate. In this cases it is more valuable to keep the commodity sooner, since the yield produced is positive, hence in these cases the future prices are generally higher for shorter maturities than longer (inverted future market). If we compare the filtered dynamics we reported above, with the figure 4.2 (where are indicated the normal future market day, and the inverted ones), the equivalence discussed works quite fine.
As discussed in the section 4.5 we conduct the inference for the basic model and for the model with jumps also for a restrict set of dates, excluding the period with the sudden fall in WTI quotations.
Comparing inferences reported in Table 3.2 with the one reported in Table 4.1 we can notice the slightly lower estimate for the intensity of the jump process, since the new dataset spot dynamic is less erratic than the previous one. Other changes are in the elasticity constant of the mean reversion processes (in the convenience yield and in the volatility process), since now, without the large movements due to the period excluded from dataset, hence higher value for the "mean reverting speed" are preferred. Following we report the filtered latent process for the two models.

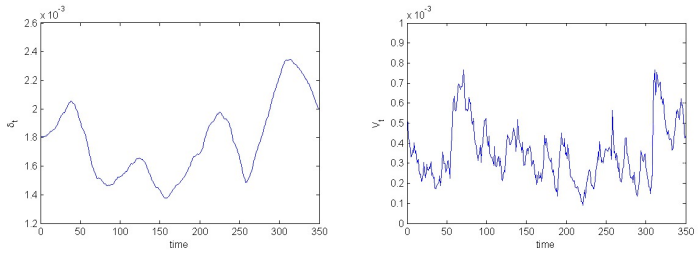

Figure 3.10: *Inference on dynamics for the convenience yield and the volatility process got under the model without seasonality or jumps*

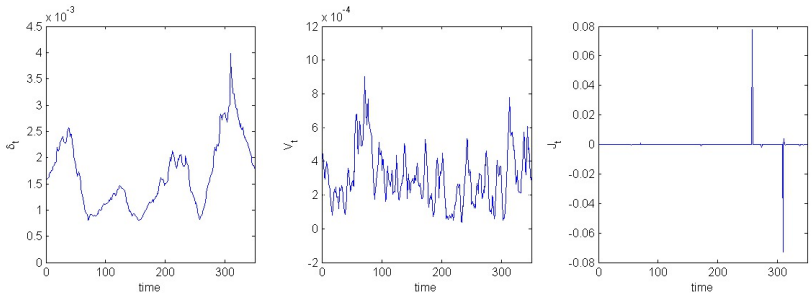

Figure 3.11: *Inference on dynamics for the convenience yield and the volatility process got under the model with jumps*

Table 3.2: *Parameter inference: posterior means (and posterior standard deviation) of the model parameter set Θ for three analyzed models. All values are expressed on a daily basis and are scaled by a* 100 *factor*

Model Parameters	Basic Model	Model with Jumps
η_S	0.1209	0.2012
	(0.0069)	(0.0194)
c	0.1983	0.1715
	(0.0061)	(0.00587)
α	1.7618	3.521
	(0.0415)	(1.38)
$\bar{\delta}$	0.1950	0.1606
	$(1.10E-3)$	(0.0066)
σ	0.0616	0.4581
	(0.0061)	(0.0411)
η_δ	0.0285	0.0318
	$(6.95E-4)$	(0.0013)
β	3.3356	2.559
	(0.2811)	(1.08)
$\bar{V}$	0.0461	0.0283
	$(7.22E-4)$	(0.00194)
ξ	0.2646	0.5267
	(0.0666)	(0.034)
ρ	-66.04	-54.99
	(2.0302)	(0.981)
λ_J	$-$	4.039
	$(-)$	(0.447)
μ_J	$-$	-1.084
	$(-)$	(0.0062)
σ_J	$-$	9.363
	$(-)$	(0.5387)

3.8 In the sample and out of the sample performances

To compare performances by the two model we analyzed both in the sample and out-of the sample results. For in the sample results we, as Yu, Li, Wells [12], analyzed the ϵ residuals to verify if the assumption of normality is satisfied , if it is then the model well describes the dynamics of data we are studying. Out of the sample we ran a particle filter algorithm using the parameter set we got from inference and check the RMSE and MAE for futures and option on futures, to value which perform better perform for risk-management purposes.
To analyze the goodness of fit and to compare the different models, we studied

the residuals $\epsilon_t^{(S)}, \epsilon_t^{(V)}, \epsilon_t^{(\delta)}$ from (4.7):

$$\begin{cases} \epsilon_t^{(S)} &= \dfrac{x_{t+1} - x_t - (\mu + c - \delta_t)\Delta t - J_{t+1}}{\sqrt{V_t \Delta t}} \\ \epsilon_t^{(V)} &= \dfrac{V_{t+1} - V_t - \beta(\bar{V} - V_t)\Delta t}{\xi\sqrt{V_t \Delta t}} \\ \epsilon_t^{(\delta)} &= \dfrac{\delta_{t+1} - \delta_t - \alpha(\bar{\delta} - \delta_t)\Delta t}{\sigma\sqrt{\delta_t \Delta t}} \end{cases} \tag{3.18}$$

Since the model hypothesis is they are distributed according to a standard normal, we use the Kolmogorov-Smirnov test to check this hypothesis and the skewness and kurtosis of the distributions to compare the different models.

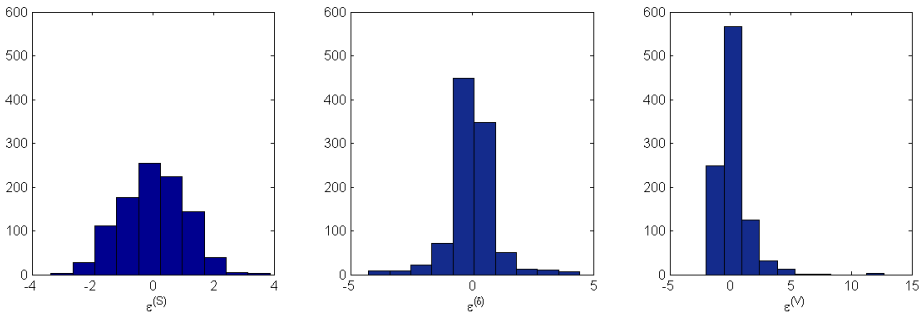

Figure 3.12: *Histogram for the residuals for the basic model*

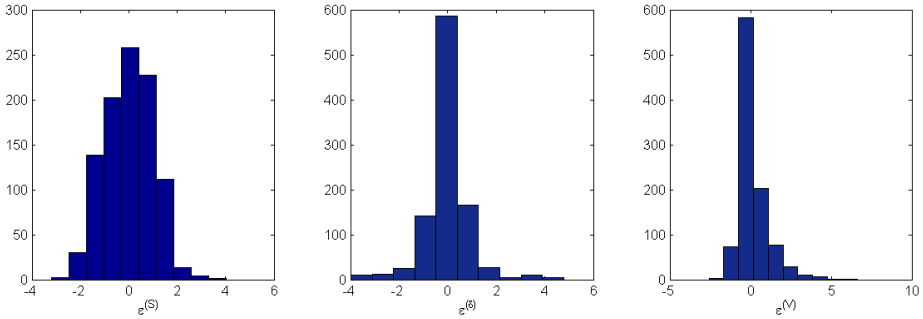

Figure 3.13: *Histogram for the residuals for the model with seasonality term*

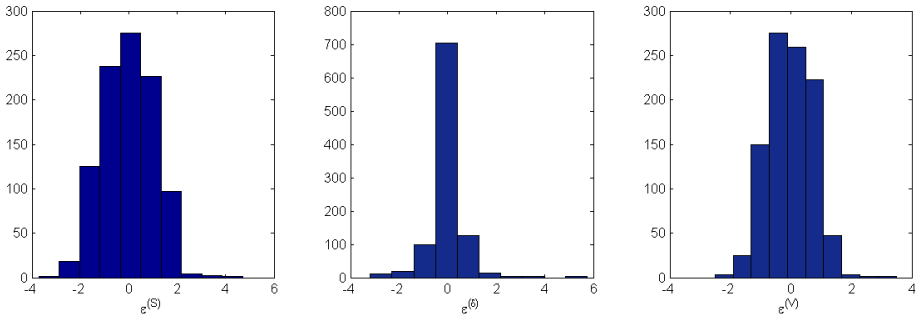

Figure 3.14: *Histogram for the residuals for the model with jumps*

Table 3.3: *Analysis of the residuals under the historical measure for the three model, for each residual is reported the p-value of the Kolmogorov-Smirnov test, the skewness and the kurtosis of the distribution*

Residual		Basic Model	Model with Seasonality	Model with Jumps
	KS test	0.151	0.47	0.1399
ϵ_S	skewness	-0.045	-0.048	0.061
	kurtosis	2.998	2.86	3.152
	KS test	$1.6E-15$	$2.4E-20$	$1.3E-24$
ϵ_δ	skewness	-0.141	0.33	0.389
	kurtosis	9.26	9.77	10.29
	KS test	$3.1E-15$	$3.1E-23$	$1.8E-6$
ϵ_V	skewness	2.95	1.98	0.039
	kurtosis	12.31	8.924	2.697

For the risk neutral dynamics we evaluate the square root of the mean of quadratic errors (RMSE) and the absolute mean error (AME)for both the data sets: the data set used for parameter estimation (in the sample set ITS) and the data set outside the first one (out of the sample OTS). Results are shown in Table **??**.

Futures error		Basic Model	Model with Seasonality	Model with Jumps
ITS	RMSE	1.036	0.956	1.08
	AME	0.525	0.457	0.499
OTS	RMSE	1.033	0.878	0.35
	AME	0.583	0.488	0.21

As it is shown by p-value for the Kolmogorov-Smirnov test we have to reject the normality hypothesis for at least two of the residuals: the convenience yield and the volatility residuals, the presence of the period with the fall of WTI spot

has a great impact in the sampled path, in particular we notice in the histograms (Fig. 4.6, 4.7, 4.8) of the residuals $\epsilon_\delta,\ \epsilon_V$ a skewed distribution , and with some values far form the mode values that brought a much more large kurtosis than for a normal distribution. Model with jumps seems to mitigate the effect of leptokurtic distribution, explaining a part of the outliers variance value with an increased jump activity. The model with seasonality does not improve the performances of the model regarding $\epsilon_\delta,\ \epsilon_V$, but the p-value of the KS test for the spot residuals is much higher than the correspondent value for the other two models. The errors associated with the futures, both in the sample and out of the sample, suggest there is not a real improvement using seasonality, for jumps out of sample errors shows a better performances, due to the greater flexibility given by using three latent process than two in our particle filter approach. This improvement has a cost in required time to run the single PF algorithm: spanning observed data coming from 998 dates and using 2000 "particles", for the basic model we need 6.45 seconds, for the model including the seasonality term the time is equal (the number of latent process simulated is the same), the model including jumps requires 9.03 seconds. All the algorithms written in Matlab have been running on a i$3(3,07$ GHz$)$ pc.
Using the reduced sample set we confirm the positive impact in performances of including jumps in model dynamics. Performances are in Table 3.4 and in Table 3.5. Using this data-set the model with jumps in spot well-score in the sample performances, revealing itself a good model, succeeding in capturing the dynamics for WTI quotation and futures in the analyzed date interval.

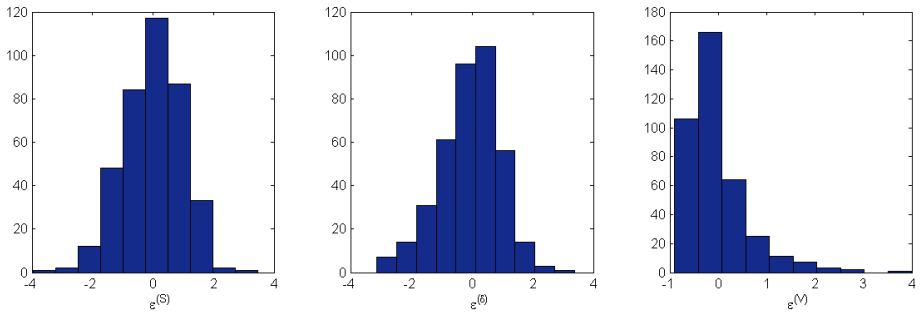

Figure 3.15: *Histogram for the residuals for the basic model using the second data-set*

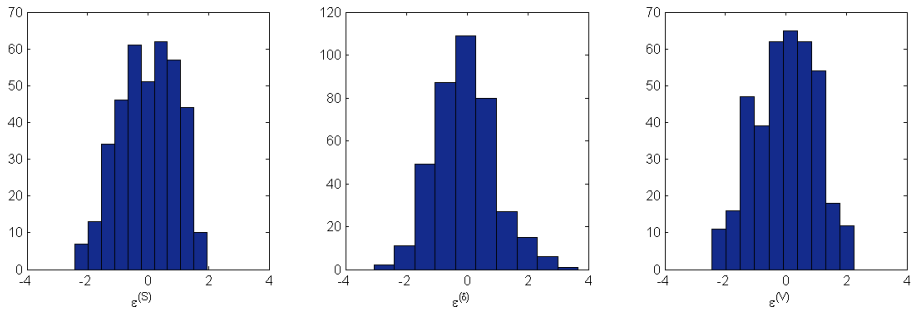

Figure 3.16: *Histogram for the residuals for the model with jumps using the second data-set*

Table 3.4: *Analysis of the residuals under the historical measure for the three model, for each residual is reported the p-value of the Kolmogorov-Smirnov test, the skewness and the kurtosis of the distribution. We refer to the second dataset*

Residual		Basic Model	Model with Jumps
ϵ_S	KS test	0.436	0.241
	skewness	-0.319	-0.175
	kurtosis	3.517	2.181
ϵ_δ	KS test	0.220	0.1085
	skewness	-0.311	0.360
	kurtosis	3.340	3.480
ϵ_V	KS test	$1.7E-14$	0.303
	skewness	2.131	-0.174
	kurtosis	9.701	2.327

Table 3.5: *Second dataset futures performances*

Futures error		Basic Model	Model with Jumps
ITS	RMSE	1.933	0.997
	AME	1.417	0.455
OTS	RMSE	0.735	0.41
	AME	0.574	0.33

3.9 Conclusions

In this paper we analyzed three models in capturing the dynamics of WTI spot characteristics and the future market structure. Two data sets have been used to

get inference, the first one includes a period characterized by a sudden fall in spot quote dynamics, while the second one excludes it. We found that model with jumps has the best performances in both cases and succeeds in properly describing the second set of observed data, even if no one of the three models succeeds in capture the dynamics of the observed data for the first set. Hence we analyzed the performances of the basic model and the model with jumps with a smaller dataset, where we excluded the period with the abrupt movement and we found that the while the basic model still show some problem in well fitting the observed data, the jump including jumps in spot dynamics succeed in it. We reported also the performances in capturing future market dynamics for a set of data outside the set used to get the inference for the parameter set of the model. Again the Jump model shows the best performances but it requires a longer time to run the algorithm with respect to both basic and seasonality model. It is worth remarking that the analysis of our data shows that sample correlation between convenience yield process and spot logarithm process is significant. A model including correlation between these two processes will be the subject of future investigation.

3.10 Appendix: Future Options prices

To evaluate, under the $\mathcal{Q}$ measure, a general contingent G claim, whose payoff is determined by an underlying whose dynamics follows (4.6), with the seasonality function fixed at zero $g(t_{\text{year}}) = 0$, we refer to the PDE:

$$
\begin{aligned}
&\frac{\partial H}{\partial t} + \frac{\partial H}{\partial x}(r_f + c - \lambda\mu_J^* - \delta_t - \frac{1}{2}V_t) + \frac{1}{2}\frac{\partial^2 H}{\partial x^2}V_t + \\
&\quad + \frac{\partial H}{\partial \delta}\left[\alpha(\bar{\delta} - \delta_t) - \eta_\delta \delta_t\right] + \frac{1}{2}\frac{\partial^2 H}{\partial \delta^2}\sigma^2\delta_t + \\
1 &\quad + \frac{\partial H}{\partial V}\left[\beta(\bar{V} - V_t) - \eta_V V_t\right] + \frac{1}{2}\frac{\partial^2 H}{\partial V^2}\xi^2 V_t + \\
&\quad + \frac{\partial^2 H}{\partial V \partial x}\rho\xi V_t + \\
&\quad - r_f H + \lambda\mathbb{E}\left[H(t, x_t + ln(1+J), \delta_t, V_t) - H(t, x_t, \delta_t, V_t)\right] = 0
\end{aligned}
\tag{3.19}
$$

Following Heston [35] and Bakshi, Madan [9] we can decompose the value C at time t of an European option that gives the right at maturity time T to acquire a futures contract, whose deliver is fixed at $T^{'}$, at a fixed price K:

$$C(t, \tau, \tau^{'} - \tau) = G(t, \tau)\Pi(t, \tau) - KB(t, \tau)\Pi(t, \tau)$$

where $\tau = T - t$ and $\tau^{'} = T^{'} - t$; $B(t, \tau)$ is the price at time t of a discount bond and $G(t, \tau)$ is the price at time t of a forward which deliver at time $t + \tau$ the asset

$S(t+\tau)$.
The decomposition has been proved by Yan [59]. Bakshi and Madan [9] proved that:

$$\begin{aligned} B(t,\tau) &= f(t,\tau;0) \\ G(t,\tau) &= \frac{1}{i}\frac{\partial f(t,\tau;0)}{\partial \phi} \\ f_1(t,\tau;\phi) &= \frac{1}{iG(t,\tau)}\frac{\partial f(t,\tau;\phi)}{\partial \phi} \\ f_2(t,\tau;\phi) &= \frac{1}{G(t,\tau)} f(t,\tau;\phi) \end{aligned} \tag{3.20}$$

where $f_1(t,\tau;\phi)$ and $f_2(t,\tau;\phi)$ are respectively the fourier transform of the two probability terms $\Pi_1(t,\tau)$ and $\Pi_2(t,\tau)$. Hence each term can be recovered once computed $f(t,\tau;\phi)$, that is the characteristic function of the underlying of the contingent claim (in this case the characteristic function of the logarithm of the future prices). Since $f(t,\tau;\phi)$ is the price of a contingent claim paying $\exp\{i\phi H(t+\tau,\tilde{\tau}-\tau)\}$ at maturity $t+\tau$, its value can be found solving (4.17) with the terminal condition:

$$f(t+\tau;0) = \exp\{i\phi H(t+\tau,\tilde{\tau}-\tau)\} \tag{3.21}$$

If we use the trial solution:

$$f(t,\tau;\phi) = \exp\{\theta_0(\tau) + \theta_\delta(\tau)\delta_t + \theta_V(\tau)V_t + i\phi[x + A_0(t) + A_2(t)\delta_t]\}$$

in (4.17), we get the following ODE system.

$$\begin{cases} - & \dfrac{\partial \phi_0(\tau)}{\partial \tau} + i\phi(r_f + c - \lambda\mu_J^*) + \phi_\delta \alpha\bar{\delta} + \theta_V \beta \bar{V} = 0 \\ - & \dfrac{\partial A_1(\tau)}{\partial \tau} = 0 \\ - & \dfrac{\partial \phi_\delta(\tau)}{\partial \tau} + i\phi - \phi_\delta(\alpha + \eta_\delta) + \frac{1}{2}\phi_\delta^2(\tau)\sigma^2 = 0 \\ - & \dfrac{\partial \theta_V(\tau)}{\partial \tau} - \frac{1}{2}i\phi - \frac{1}{2}\phi^2 - \theta_V(\beta + \eta_V) + \frac{1}{2}\xi^2\theta_3^2 + i\phi\theta_V\rho\xi = 0 \end{cases} \tag{3.22}$$

where $\phi_\delta(\tau) = \theta_\delta(\tau) + i\phi A_2(\tilde{\tau}-\tau)$ and $\phi_0(\tau) = \theta_0(\tau) + i\phi A_0(\tilde{\tau}-\tau)$ with terminal condition:

$$\begin{cases} \phi_0(0) & = i\phi\beta_0(\tilde{\tau}) \\ \phi_\delta(0) & = i\phi\beta_\delta(\tilde{\tau}) \\ \theta_V(0) & = 0 \end{cases} \tag{3.23}$$

Solving the ODE system we get:

$$\theta_V = -\frac{2}{\xi^2}\left(\frac{\exp\{C_V\tau\}-1}{\frac{\exp\{C_V\tau\}}{A_V}-\frac{1}{B_V}}\right)$$

with

$$\begin{cases} C_V &= \sqrt{(\eta_V - i\phi\rho\xi+\beta)^2+\xi^2(\phi^2+i\phi)} \\ A_V &= \frac{1}{2}[(\beta+\eta_V-i\phi\rho\xi)+C_V] \\ B_V &= \frac{1}{2}[(\beta-+\eta_V-i\phi\rho\xi)-C_V] \end{cases}$$

And:

$$\phi_\delta = -\frac{2}{\sigma^2}\frac{A_\delta\exp\{C_\delta\tau\}+B_\delta\bar{C}}{\exp\{C_\delta\tau\}+\bar{C}}$$

where:

$$\begin{cases} C_\delta &= \sqrt{(\alpha+\eta_\delta)^2+2\sigma^2 i\phi} \\ A_\delta &= \frac{1}{2}[(\eta_\delta+\alpha)+C_\delta] \\ B_\delta &= \frac{1}{2}[(\eta_\delta+\alpha)-C_\delta] \\ \tilde{C}_\delta &= \sqrt{(\alpha+\eta_\delta)^2+2\sigma^2} \\ \tilde{A}_\delta &= \frac{1}{2}[(\eta_\delta+\alpha)+\tilde{C}_\delta] \\ \tilde{B}_\delta &= \frac{1}{2}[(\eta_\delta+\alpha)-\tilde{C}_\delta] \\ \bar{C} &= \dfrac{i\phi\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}-1}{\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}}{\tilde{A}_\delta}-\frac{1}{\tilde{B}_\delta}}-A_\delta}{B_\delta - i\phi\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}-1}{\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}}{\tilde{A}_\delta}-\frac{1}{\tilde{B}_\delta}}} \end{cases}$$

Finally we solve for ϕ_0:

$$\begin{aligned}\phi_0 =& i\phi(rf+c-\lambda\mu_J^*)\tau - \frac{2\beta\bar{V}}{\xi^2}\left((B_V+C_V)\tau+\log\frac{A_V\exp\{-C_V\tau\}-B_V}{C_V}\right)- \\ &-\frac{2\alpha\bar{\delta}}{\sigma^2}\left((B_d+C_d)\tau+\log\frac{1+\exp\{-C_d\tau\}\tilde{C}}{1+\tilde{C}}\right)+ \\ &+i\phi\left((rf+c)\tilde{\tau}-\frac{2\alpha\bar{\delta}}{\sigma^2}\left((\tilde{B}_\delta+\tilde{C}_\delta)\tilde{\tau}+\right.\right. \\ &\left.\left.+\log\frac{\tilde{A}_\delta\exp\{-\tilde{C}_\delta\tilde{\tau}\}-\tilde{B}_\delta}{\tilde{A}_\delta-\tilde{B}_\delta}\right)\right)+\lambda_J\left(\exp\left\{i\phi\mu_J-\frac{1}{2}\phi^2\sigma_J^2\right\}-1\right)\end{aligned}$$

with:

$$\mu_J^* = \exp\left\{\mu_J+\frac{1}{2}\sigma_J^2\right\}-1$$

CHAPTER 4

A double correlated 3 factor model for Crude Oil Market

4.1 Introduction

One of the key elements in modeling crude oil markets is the convenience yield process. Since the seminal work by Schwartz [53], where the convenience yield has been introduced as a discount factor in a reduced form model for commodities, and later by Cortazar and Schwartz [16], where a detailed discussion on these models is presented, several different studies have been developed, in which the convenience yield have been related to the inventory level, giving a macroeconomic interpretation to the convenience yield. A long list of articles discussed this topic. Among the others, that by Liu and Tang [39] analyzed both the modeling problem in commodity market and the connection with the theory of inventory, showing its important role in explaining the commodity spot dynamics. From a pure financial point of view the convenience yield level reflects in the structure of the future markets both in the contango and backwardation regimes. Hence, even if it has been described as a discount yield, the dynamics usually proposed was Vasicek-type in order to allow both positive and negative values. Liu and Tang [39] and Ribeiro and Hodges [50] proposed two different models with convenience yield valuated up to a constant as a CIR dynamics. Even if the reason leading to their choices are quite different, the models exhibit strong similarities:

while Ribeiro and Hodges start from separating the cost of storage and insurance from the possible positive revenues from physically owning the commodity, Liu and Tang motivate their choice with the aim to explain the econometric evidence they found of heteroskedasticity in the market analyzed. Both the models agree in considering the Wiener process driving the spot dynamics correlated with the convenience yield process and the square root of the same process multiplies up to a scale factor the Wiener diffusion process driving the spot dynamics. The purpose of the present paper is to provide a further contribution on the subject by performing a detailed analysis of three different, but related models of crude oil market based on some inference results obtained via a recently developed estimation method, usually called "Particle Filtering". In our analysis we started from the model proposed by Liu and Tang [39] and studied its performances in catching spot and futures time series for WTI in the period $2010 - 2011$, comparing this model with two different variants including a new factor: volatility process. In order to get an affine process we need to add a second Wiener process not correlated with the first one. An attempt in a similar direction has already been done by Yan [59], who proposed the extension of the simple Schwartz model by adding a volatility factor to explain the skew and smile phenomenon observable in the option on the futures market. Since the data set used is not so huge and the maturities of futures taken into account are long up to 3 years, we didn't allow interest rates to have a stochastic dynamics since, as it was already discussed in the original work by Schwartz, the interest rates have a positive impact in describing the dynamics just when we consider longer maturities futures. It was also shown by Liu and Tang that correlation of interest rates dynamics with log-spot dynamics is not significant. Because of the similarities between the present models and the stochastic volatility models for equities, in order to obtain estimates we decided to adopt techniques analogous to those discussed in Eraker [22] and Yu, Li and Wells [12], where a Markov Chain Monte Carlo method is used to get inference on $S\&P500$ index analyzing both the spot market and the option market data. Our algorithm, belonging to Markov Chain Monte Carlo family, has been introduced by Doucet, Andrieu and Holenstein [10]. Pitt et al. [40] further developed the algorithm proposed, by suggesting to use particle filters (particularly suitable for latent process filtering) to estimate the marginal likelihood of the parameters. The optimality of the Particle Filter algorithm in applying to Stochastic volatility models is the topic discussed by Johannes, Polson and Stroud [43]. The structure of the paper is the following: in Section 2 we present the models we are going to study, while in Section 3 we compute in close form for the models examined the futures prices we need in order to apply our inference procedure; in Section 4 we present our estimation algorithm, while in Section 5 we illustrate the data set on which base our inference; in Section 6 we provide and discuss the main results of our study, while in Section 7 we analyze the in and out of the sample performances of our estimation method. Section 8 contains some final remarks and

suggests possible developments of the present work. The explicit computation of futures options prices is postponed to the appendix.

4.2 The Models

We studied three different models, the first one being our benchmark, and this is the model originally presented by Liu and Tang [39]. The other two extend the first one by including two stochastic processes into the dynamics: the first one introduces volatility process and the second one includes also a jump process, of the compund Poisson type, with exponentially distributed arrival jump times and lognormal jump size density. The benchmark model has been introduced by Liu and Tang, who describe the stochastic dynamics under the historical measure $\mathbb{P}$.
We decided to keep the rates not stochastic and we keep it fixed at the FED levels. As already discussed by Schwartz [53], by comparing performances in capturing the futures prices structure for models with and without stochastic interest rate, the performances in using stochastic interest rates do not show any significant improvement, if we exclude the longest maturity futures (about ten-years maturity). Hence, since our longest maturity considered is 5 years futures, we preferred to do not allow stochastic interest rates to make the model more parsimonious.
Liu Tang Model is a two-factor model with a stochastic convenience yield following a CIR dynamics[1]:

$$\begin{cases} dx_t = \left[\mu + c - \delta_t - \frac{1}{2}(\sigma_\delta^2 \delta_t + V_0 + V_\delta \delta_t)\right] dt + \sigma_\delta \sqrt{\delta_t} dW_1^{(\mathbb{P})} + \sqrt{V_0 + V_\delta \delta_t} dW_2^{(\mathbb{P})} \\ d\delta_t = \alpha(\bar{\delta} - \delta_t)dt + \sigma\sqrt{\delta_t} dW_1^{(\mathbb{P})} \end{cases} \tag{4.1}$$

We made the usual choice of describing the price dynamics through the log-price process: $x_t = \log[S_t]$. The convenience yield process is given by $\delta_t - c$, and it can assume both positive and negative values. W_1 and W_2 are two independent Brownian motions, the superscript $\mathbb{P}$ being a reminder that this description holds under the historical measure.
In order to explain both the time series of spot prices and futures prices, we need to specify the risk premia to link the historical and risk neutral measures. We followed for this model the same choice made by Liu and Tang, with risk premia linearly proportional to the square root of the state variables, maintaining the

[1]The model in its structure is similar to the one proposed by Ribeiro and Hodges [50], but has some relevant differences, in particular in the correlation between the spot process and convenience yield process.

same model structure under both measures.

$$\begin{cases} dW_{\delta_t}^{(\mathbb{Q})} = dW_{\delta_t}^{(\mathbb{P})} + \lambda(Y_t)dt \\ \lambda(Y_t) = \sqrt{Y_t}\Theta \\ \Theta = \left[\dfrac{\eta_\delta}{\sigma}, \eta_x\right] \end{cases}$$

Hence we recognize that in (4.1):

$$\mu = r_f + \eta_\delta \frac{\sigma_\delta}{\sigma}\delta_t + \eta_x \left(V_0 + V_\delta \delta_t\right)$$

The dynamical model expressed in (4.1) under the historical measure, under the risk neutral measure maintain the same structure:

$$\begin{cases} dx_t = \left[r_f + c - \delta_t - \dfrac{1}{2}(\sigma_\delta^2 \delta_t + V_0 + V_\delta \delta_t)\right] dt + \sigma_\delta \sqrt{\delta_t} dW_1^{(\mathbb{Q})} + \sqrt{V_0 + V_\delta \delta_t} dW_2^{(\mathbb{Q})} \\ d\delta_t = [\alpha(\bar{\delta} - \delta_t) - \eta_\delta \delta_t]dt + \sigma\sqrt{\delta_t} dW_1^{(\mathbb{Q})} \end{cases} \tag{4.2}$$

The other two models considered introduce the volatility factor V_t, both of them are a three factor model where the variance process and the convenience yield process are modeled via a CIR dynamics, the second one including a jump process in the log-spot dynamics. The first of these models (the one without jumps):

$$\begin{cases} dx_t = (\mu + c - \delta_t)dt + \sqrt{V_t} dW_{S_t}^{(\mathbb{P})} \\ d\delta_t = \alpha(\bar{\delta} - \delta_t)dt + \sigma\sqrt{\delta_t} dW_{\delta_t}^{(\mathbb{P})} \\ dV_t = \beta(\bar{V} - V_t)dt + \xi\sqrt{V_t} dW_{V_t}^{(\mathbb{P})} \\ dW_{S_t}^{(\mathbb{P})} dW_{V_t}^{(\mathbb{P})} = \rho dt \end{cases} \tag{4.3}$$

The last model taken into account has the same structure of the previous one, allowing the spot dynamics to jump. The jump part is modeled via a compound Poisson process: $J_{x_t}^{(\mathbb{P})}$ is the jump process, where the jump arrival times are exponentially distributed with λ_j parameter, and the jump sizes (in the log-price) are normally distributed with mean μ_J and variance σ_J^2.

$$\begin{cases} dx_t = (\mu + c - \delta_t)dt + \sqrt{V_t} dW_{S_t}^{(\mathbb{P})} + dJ_{x_t}^{(\mathbb{P})} \\ d\delta_t = \alpha(\bar{\delta} - \delta_t)dt + \sigma\sqrt{\delta_t} dW_{\delta_t}^{(\mathbb{P})} \\ dV_t = \beta(\bar{V} - V_t)dt + \xi\sqrt{V_t} dW_{V_t}^{(\mathbb{P})} \\ dW_{S_t}^{(\mathbb{P})} dW_{V_t}^{(\mathbb{P})} = \rho dt \end{cases} \tag{4.4}$$

Since the (4.3) model can be considered a particular case of (4.4), setting the parameter $\lambda_J = 0$ we discussed just the second one, recalling that the same conclusions hold also for the first one, just by setting to zero the parameter λ_J.
In order to describe the dynamics under the risk-neutral measure, we need again

to define the Radon-Nikodym derivative of this measure with respect to the historical one. The parametric form we choose follows the one proposed by Heston [35] and discussed by Dai and Singleton [18] and by Pan [45], and it has been used by Yu, Li and Wells [12] in their analysis of Bates model under both historical and risk neutral measures. This choice preserves the model structure under the measure change. This can be specified by the following relations between the Wiener processes under the two measures, provided by the Girsanov theorem:

$$\begin{cases} dW_{\delta_t}^{(\mathbb{Q})} = dW_{\delta_t}^{(\mathbb{P})} - \dfrac{1}{\sqrt{1-\rho_\delta^2}}\left(\rho_\delta \eta_x - \dfrac{\eta_\delta}{\sigma}\right)\sqrt{\delta_t}dt \\ dW_{V_t}^{(\mathbb{Q})} = dW_{V_t}^{(\mathbb{P})} - \dfrac{1}{\sqrt{1-\rho_V^2}}\left(\rho_V \eta_x - \dfrac{\eta_V}{\xi}\right)\sqrt{V_t}dt \\ dW_1^{(\mathbb{Q})} = dW_1^{(\mathbb{P})} + \eta_x \sigma_\delta \sqrt{\delta_t}dt \\ dW_2^{(\mathbb{Q})} = dW_2^{(\mathbb{P})} + \eta_x \sqrt{V_t}dt \end{cases} \tag{4.5}$$

Where we have denoted by η_x the risk premium associated with the two Wiener processes in the spot dynamics. We chose to associate the same η_x to both the processes since both the processes affect the spot dynamics in the an analogous way. In addition to this, if we defined two different risk premia one for each process, it could be possible to redefine an unique risk premia as a weighted average of these two.
Hence, we can write the dynamics just described under the measure $\mathbb{P}$ by the system (4.4), also under the risk-neutral measure $\mathbb{Q}$[2] :

$$\begin{cases} dx_t = (r + c - \delta_t - \dfrac{1}{2}V_t - \mu^*)dt + \sqrt{V_t}dW_{x_t}^{(\mathbb{Q})} + dJ_{x_t}^{(\mathbb{Q})} \\ d\delta_t = \left(\alpha(\bar{\delta} - \delta_t) + \eta_\delta \delta_t\right)dt + \sigma\sqrt{\delta_t}dW_{\delta_t}^{(\mathbb{Q})} \\ dV_t = \left(\beta(\bar{V} - V_t) + \eta_V V_t\right)dt + \xi\sqrt{V_t}dW_{V_t}^{(\mathbb{Q})} \\ S_t = \exp\{g(t_{\text{year}}) + x_t\} \\ dW_{S_t}^{(\mathbb{Q})}dW_{V_t}^{(\mathbb{Q})} = \rho dt, \end{cases} \tag{4.6}$$

where μ^* and $dJ^{(\mathbb{Q})}$ denote respectively the predictable compensator and the compensated jump measure. It is worth remarking that, since the cost of storage is separated off the dynamics of the convenience yield, δ_t this is modeled by a CIR process that prevents it from assuming negative values, automatically excluding arbitrage opportunities in the market, as pointed out in Ribeiro and Hodges [50]. It is worth to remark that again the convenience yield is expressed by $\delta_t - c$ and it can assume both positive and negative values to catch both the normal future market structure and the inverted future structure.
To cope with the inference in this setting we use a Euler discretization method.

[2] for the basic model we just fix to zero μ^* and $dJ_{x_t}^{(\mathbb{Q})}$, for the model with seasonality we add the seasonality factor ($g(t_{\text{year}})$) to the spot dynamics like in (4.4).

The discretized model (under the historical measure) can be written as follows:

$$\begin{cases} x_{t+1} = x_t + \left[\mu + c - \left(\frac{\sigma^2}{2} + 1\right)\delta_t - \frac{1}{2}V_t\right]\Delta t + \sqrt{V_t}\epsilon_{t+1}^{(1)} + \sqrt{\delta_t}\epsilon_{t+1}^{(2)} + \sum_{i=1}^{N_{t+1}^{(J)}} \epsilon_{i,t+1}^{(J)} \\ \delta_{t+1} = \delta_t + \alpha\left(\bar{\delta} - \delta_t\right)dt + \sigma\sqrt{\delta_t}\epsilon_{t+1}^{(\delta)} \\ V_{t+1} = V_t + \beta\left(\bar{V} - V_t\right)dt + \xi\sqrt{V_t}\epsilon_{t+1}^{(V)} \end{cases} \tag{4.7}$$

Where each $\epsilon_{t+1}^{(1)}$ is correlated with $\epsilon_t^{(\delta)}$ and $\epsilon_{t+1}^{(2)}$ is correlated with $\epsilon_{t+1}^{(V)}$ and are normally distributed random variables with zero mean and variance Δt. In the discretized jump addend, the $N_{t+1}^{(J)}$ are independent Poisson distributed r.v. with parameter λ_J, while $\epsilon_{i,t}^{(J)}$ are independent normal r.v.
The futures price is given by the following expression:

$$F(0,\tau) = \exp\{A_0(\tau) + x_t + A_2(\tau)\delta_t\} \tag{4.8}$$

where $\tau = T - t$ is the futures time to maturity. Details about computations and the specifics of the functions $A_0(\tau)$ and $A_2(\tau)$ are provided in the following section. Details for these two function for the model proposed by Liu and Tang, can be found in their paper [39].
Since the futures prices are affected by different kind of noises (possible incomplete specification of the model, market inefficiency, random noise, etc) we modeled the logarithm of their market price, $\ln F^{\mathrm{M}}(0,\tau)$, by assuming it is represented by the theoretical price given by (4.8) plus an error distributed as a white noise $\epsilon_{\text{fut}} \sim \mathcal{N}(0,\sigma_\epsilon^2)$[3]:

$$\ln F^{\mathrm{M}}(0,\tau) = \ln F(0,\tau) + \epsilon_{\text{fut}} \tag{4.9}$$

4.3 Futures Prices

When the underlying follows, under the the dynamics $\mathbb{Q}$ (4.6) the Kolmogorov backward equation for the generic contract value, $f(t, x_t, \delta_t, V_t, J_t)$, at time t, when the underlying follows, under the $\mathcal{Q}$ measure, the dynamics (4.6) is:

$$\begin{aligned} &\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}(r_f + c - \lambda\mu_J^* - (1 + \frac{\sigma_\delta}{2})\delta_t - \frac{1}{2}V_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(V_t + \sigma_\delta^2\delta_t) + \\ &\quad + \frac{\partial f}{\partial \delta}\left[\alpha(\bar{\delta} - \delta_t) - \eta_\delta\delta_t\right] + \frac{1}{2}\frac{\partial^2 f}{\partial \delta^2}\sigma^2\delta_t + \\ &\quad + \frac{\partial f}{\partial V}\left[\beta(\bar{V} - V_t) - \eta_V V_t\right] + \frac{1}{2}\frac{\partial^2 f}{\partial V^2}\xi^2 V_t + \\ &\quad + \frac{\partial^2 f}{\partial V \partial x}\rho\xi V_t + \frac{\partial^2 f}{\partial V \partial \delta}\rho_\delta\sigma\sigma_\delta\delta_t + \\ &\quad + \lambda\mathbb{E}\left[f(t, x_t + ln(1+J), \delta_t, V_t) - f(t, x_t, \delta_t, V_t)\right] = 0 \end{aligned} \tag{4.10}$$

[3]Our choice to fix $\sigma_\epsilon^2 = 0,08$ is consistent with the value observed for the mean of the daily spread of the highest-lowest prices for futures in the considered range

The PDE (4.10) has to be solved with the usual terminal condition $f(t=T) = H(x_T, \delta_T, V_T)$ with H payoff at time T maturity, and since we are dealing with futures the final payoff is:

$$H(x_T, \delta_T, V_T) = \exp\{x_T\} \tag{4.11}$$

We make the hypothesis of a solution form:

$$f_t = \exp\{A_0(t) + A_1(t)x_t + A_2(t)\delta_t + A_3(t)V_t\}$$

if we try this guess solution into (4.10), we get the follow ODE system:

$$\begin{cases} - & \dfrac{\partial A_0(\tau)}{\partial \tau} + A_1(\tau)(r_f + c) + A_2(\tau)\alpha\bar{\delta} + A_3\beta\bar{V} = 0 \\ - & \dfrac{\partial A_1(\tau)}{\partial \tau} = 0 \\ - & \dfrac{\partial A_2(\tau)}{\partial \tau} - A_1(\tau) - A_2(\tau) + \frac{1}{2}A_2^2(\tau)\sigma^2 + A_1(\tau)A_2(\tau)\rho_\delta\sigma_\delta\sigma = 0 \\ - & \dfrac{\partial A_3(\tau)}{\partial \tau} - \frac{1}{2}A_1(\tau) + \frac{1}{2}A_1^2(\tau) - A_3(\tau)(\beta + \eta_V) + \frac{1}{2}A_3^2(\tau)\xi^2 + A_1(\tau)A_3(\tau)\rho\xi = 0 \end{cases} \tag{4.12}$$

where we changed variable from t to $\tau = T - t$ [4] .
Since the payoff is (4.11) we obtain the initial condition for the ODE system:

$$\begin{cases} A_0(0) &= 0 \\ A_1(0) &= 1 \\ A_2(0) &= 0 \\ A_3(0) &= 0 \end{cases} \tag{4.13}$$

Solving previous system we get :

$$\begin{cases} A_0(\tau) &= (r_f + c)\tau - \dfrac{2\alpha\bar{\delta}}{\sigma^2}\left[B\tau + \log\left(\dfrac{D - B\exp\{C\tau\}}{D - B}\right)\right] \\ A_1(\tau) &= 1 \\ A_2(\tau) &= -\frac{2}{\sigma^2}\dfrac{\exp\{C\tau\} - 1}{\frac{\exp\{C\tau\}}{D} - \frac{1}{B}} \\ A_3(\tau) &= 0 \end{cases} \tag{4.14}$$

with:

$$\begin{cases} C &= \sqrt{(\alpha + \eta_\delta - \rho_\delta\sigma\sigma_\delta)^2 + 2\sigma^2} \\ D &= \dfrac{(\eta_\delta + \alpha - \rho_\delta\sigma\sigma_\delta) + C}{2} \\ B &= \dfrac{(\eta_\delta + \alpha - \rho_\delta\sigma\sigma_\delta) - C}{2} \end{cases} \tag{4.15}$$

[4] Note that in the Futures price obtained there is no dependence on the jump process parameters in strict analogy with Yan [59].

4.4 Inference Algorithm

The reference object of our inference is the vector $\{\Theta, V_{0:T}, \delta_{0:T}, J_{0:T}\}$ where $V_{0:T}$, $\delta_{0:T}$, $J_{0:T}$ are the three latent processes (the variance process, the convenience yield and the jump process) and Θ is the set of parameters: $\Theta = \{\epsilon, \mu, c, \alpha, \bar{\delta}, \sigma, \eta_\delta, \beta, \bar{V}, \xi, \rho, \lambda, \mu_J, \sigma_J\}$. For simplicity we indicate with $X_{0:T}$ the vector of the three latent processes $\{V_{0:T}, \delta_{0:T}, J_{0:T}\}$, and by $Z_{1:T}$ the set of observed market data (coming from both asset price and futures price quotations).

To make inference in a so large state space, Particle Markov Chain Monte Carlo methods (from now on PMMC) represents an efficient technique. The technique was presented by Andrieu et al. [10], here it also shown trough application to one model of Stochastic Volatility family, the algorithm efficiency in sampling from the distribution of Θ and $X_{0:T}$ given the observed data sequence $Z_{1:T}$. The algorithm belongs to the families of MCMC and Particle filter, and this reflects in the good capability to deal with non linear model. The implemented version is the one presented and discussed by Pitt et al. [40]. The main feature of the algorithm described there is that it is possible to build a MCMC to sample from the distribution $p(\Theta|Z_{1:T})$ (since Θ is multidimensional, with large dimension, a Metropolis within Gibbs technique has been used to improve the mixing of the chain) running the MCMC algorithm with an estimate for the marginal likelihood $p(Z_{1:T}|\Theta)$ has been used (the likelihood is marginal with respect to the parameter set, since the latent processes have been integrated out) and it has been got by using a Auxiliary Particle Filter algorithm. Details about the algorithm, and proof of convergence of the algorithm can be found in Andrieu et al. [10] and in Pitt et al [40]. To design the Auxiliary Particle Filter we extended the algorithm discussed by Johannes, Polson and Stroud [43] and implemented in a version without data augmentation, since as it has been shown in the same paper the the benefits of data augmentation for such a kind of models when using daily data is negligible, moreover in a regime of low volatility of volatility, as it is our analyzed cases.

4.5 The Data

The data used for the analysis are relative to WTI Cushing Crude Oil spot and futures quotations on NYMEX market from $11/01/2010$ to $28/09/2011$. We used the the first 85% if the data-set to get the inference on our model and reserved the last 15% of dates for out of the sample performances study. The range of dates for the estimation goes from $11/01/2010$ to $30/06/2011$, and the spot data are presented in Figure 4.1.
Spot data are collected from the US Energy information administration website where are collected teh closing day spot values registered in the Cushing Oklahoma market and communicated to the managers of the U.S. government website. Daily data are taken into account for any available working day in the interval.

A plot of the spot data used in estimating parameter set of the different models analyzed is in figure Fig.4.1

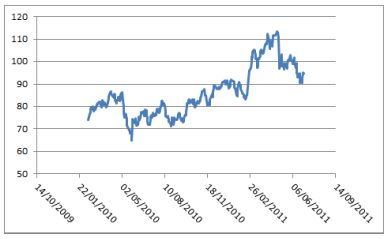

Figure 4.1: *Spot FOB data for WTI crude oil*

Besides spot data, for any working day we recorded a panel of 12 future contract values. Their maturity day is fixed on the first working day of each month of 2012. So for any trading day we analyzed a spot datum and 12 futures data, and in the range there are 350 dates. Hence the data set consists in 350 spot values and 4200 future contracts.
In addition to this data set we reserved a panel of data to evaluate out of the sample performances. The data set include again a FOB spot datum and a panel of 12 future data (with different maturity one for each month of 2012, the maturities are set on the first trading day), the range of dates goes from $01/07/2011$ to $28/09/2011$. The number of working days in the range is 81.
Futures contracts are usually characterized by their behavior at different maturities, when the futures prices are higher than spot value, increasing over longer maturities, hence the market is normal, otherwise inverted. If at a certain date the futures quotation increases when the time to maturity become longer, it usually said the future market is normal, otherwise it is said inverted. In the analyzed data set the usual structure is normal, ad it switch to inverted structure for a few days in the closing part of the time series. In figure Fig.4.2 are indicate the dates for which we can observe a normal future market (equivalent to the value $+1$) and the dates with an inverted future market(equivalent to the value -1).

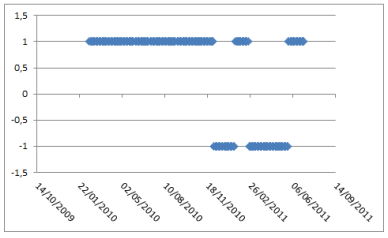

Figure 4.2: *The dates in the data set used for the parameter estimation are divided in normal future market days ($+1$ value) and inverted future market days (-1 value)*

4.6 Numerical Results

To conduct our inference for each of our models discussed in section 4.2, we simulated Markov chains of 35000 steps (from these we discard about 10000 sampled points as burn-in of the Markov chain[5]). We used the mean of the posterior distribution $p(\Theta|Z_{0:T})$ as point estimate for each parameter; these results are reported in the Table 4.1. Together with means we reported also the standard deviation of the posterior distributions.

Table 4.1: *Parameter inference: posterior means (and posterior standard deviation) of the model parameter set Θ for three analyzed models. All values are expressed on a daily basis*

Model Parameters	Liu-Tang	Model with Vol	Model with Vol & Jumps
η_S	$5.44e-4$	$1.31e-3$	$1.91e-3$
	$(6.48e-4)$	$(6.40e-5)$	$(8.05e-5)$
c	$1.32e-3$	$1.25e-3$	$1.30e-3$
	$(7.69e-5)$	$(1.11e-4)$	$(3.35e-5)$
α	$5.10e-3$	$1.74e-4$	$2.82e-4$
	$(1.44e-3)$	$(1.44e-4)$	$(5.672e-4)$
$\bar{\delta}$	$1.24e-3$	$1.20e-3$	$1.21e-3$
	$(3.36e-5)$	$(3.45e-5)$	$(4.34e-5)$
σ	$7.57e-4$	$4.79e-4$	$9.11e-4$
	$(1.63e-4)$	$(1.46e-4)$	$(1.34e-4)$
η_δ	$1.932e-4$	$1.66e-4$	$3.01e-4$
	$(2.24e-5)$	$(1.33e-5)$	$(1.11e-5)$
β	$-$	$1.64e-1$	$1.41e-2$
	$(-)$	$(1.47e-2)$	$(9.59e-3)$
$\bar{V}$	$-$	$3.99e-4$	$5.69e-4$
	$(-)$	$(3.44e-5)$	$(2.73e-5)$
ξ	$-$	$5.96e-3$	$1.03e-2$
	$(-)$	$(2.61e-3)$	$(2.92e-3)$
ρ	$-$	-26.64%	-53.31%
	$(-)$	$(1.10e-1)$	$(1.01e-1)$
λ_J	$-$	$-$	$2.33e-2$
	$(-)$	$(-)$	$(3.41e-3)$
μ_J	$-$	$-$	$-2.89e-3$
	$(-)$	$(-)$	$(4.38e-4)$
σ_J	$-$	$-$	$1.10e-1$
	$(-)$	$(-)$	$(3.74e-3)$
σ_δ	$4.01e-1$	$1.06e-2$	$9.50e-3$
	$(3.27e-4)$	$(2.01e-4)$	$(1.34e-4)$
ρ_δ	$-$	67.38%	72.87%
	$(-)$	$(1.65e-2)$	$(9.17e-1)$
V	$1.83e-5$	$-$	$-$
	$(1.51e-5)$	$(-)$	$(-)$
V_δ	$2.55e0$	$-$	$-$
	$(6.91e-1)$	$(-)$	$(-)$

[5]The lack of convergence has been tested by Geweke test, and at 1% significance we can reject the hypothesis of not convergence of the Markov Chain. For any further detail about test, estimation error and algorithm used to compute them we refer to the manual written by Smith [56].

Since for all the three models the cost of storage (c) estimate is greater than long run convenience yield ($\bar{\delta}$) one, then $c - \delta_t$ assumes often positive values, which reflects in a normal market effect in futures structure, as it is present for most of the dates in the considered period. From Fig.4.3 and Fig.4.4 and Fig.4.5 we can observe that the estimate for δ_t have higher values in similar periods, that correspond to inverted future market structure. The intensity of jumps has a value significantly different from zero, the point estimate suggest an average of 2.3% days with jumps.
Together with inference about the parameters we got also inference on the path of the latent processes (convenience yield and variance process). In the following figures (Fig.4.3, Fig.4.4 and Fig.4.5) are shown the inference on the latent processes for the different models, got as the mean of all path.

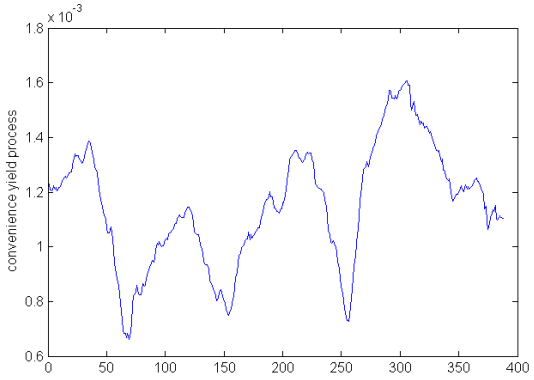

Figure 4.3: *Inference on dynamics for the convenience yield under the Liu & Tang model*

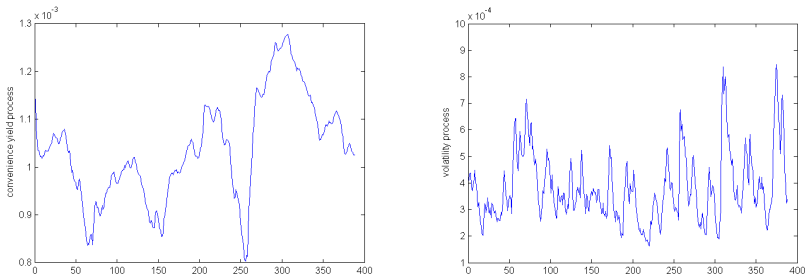

Figure 4.4: *Inference on dynamics for the convenience yield and the volatility process got under the two-factor model*

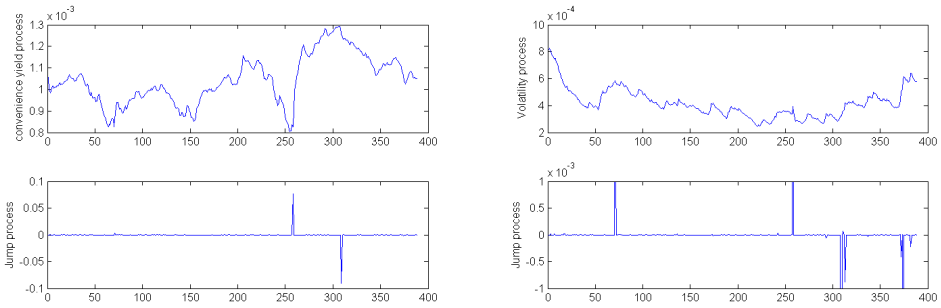

Figure 4.5: *Inference on dynamics for the convenience yield and the volatility process got under the two-factor model with jumps*

In Liu Tang model the estimate of the constant volatility V_0 is low (about $2e-5$ which corresponds to 0.5% yearly basis variance, uncorrelated with convenience yield) this reflect in a higher volatility scale factor for the convenience yield σ_δ and higher estimated values for the process δ_t with respect to the other two model. The long run variance for the two model proposed is between $4e-4$ and $5.6e-4$ that is much more close to the variance of the observed sample spot set ($3.7e-4$).
It is interesting to notice when the filtered process for the convenience yield is above the "threshold" represented by the estimated cost of storage plus the interest rate (in the period considered the rate quoted by FED was at 0.25% on yearly basis, hence on daily basis is slight smaller than $1e-5$): in this cases it is more valuable to keep the commodity sooner, since the yield produced is positive, hence in these cases the future prices are generally higher for shorter maturities than longer (inverted future market). It is possible to notice that both for the two proposed model the net convenience yield is greater than $r_f + c$ just in a few dates at the end of the analyzed period, when effectively the market showed an inverted structure, in the other dates, when the market was characterized by a normal future structure, in our filtered net convenience yield dynamics we found $\delta_t < r_f + c$. The benchmark model generally shows higher values for the convenience yield $\delta_t - c$ with respect to the two proposed model, since the volatility of the spot process has to be explained with the convenience yield process. About the volatility process the model without jumps shows a more erratic process for the volatility with respect to the model with jumps where the jumps activity partially absorbs the volatility of the spot dynamics.

4.7 In the sample and out of the sample performances

To compare performances by the two model we analyzed both in the sample and out-of the sample results. For in the sample results we, as Yu, Li, Wells [12], analyzed the ϵ residuals to verify if the assumption of normality is satisfied , if

it is then the model well describes the dynamics of data we are studying. Out of the sample we ran a particle filter algorithm using the parameter set we got from inference and check the RMSE and MAE for futures and option on futures, to value which perform better perform for risk-management purposes.
To analyze the goodness of fit and to compare the different models, we studied the residuals $\epsilon_t^{(S)}, \epsilon_t^{(V)}, \epsilon_t^{(\delta)}$ from (4.7):

$$\begin{cases} \epsilon_t^{(S)} &= \dfrac{x_{t+1} - x_t - (\mu + c - \delta_t)\Delta t - J_{t+1}}{\sqrt{[V_t + \sigma_\delta^2 \delta_t]\Delta t}} \\ \epsilon_t^{(V)} &= \dfrac{V_{t+1} - V_t - \beta(\bar{V} - V_t)\Delta t}{\xi\sqrt{V_t \Delta t}} \\ \epsilon_t^{(\delta)} &= \dfrac{\delta_{t+1} - \delta_t - \alpha(\bar{\delta} - \delta_t)\Delta t}{\sigma\sqrt{\delta_t \Delta t}} \end{cases} \tag{4.16}$$

Since the model hypothesis is they are distributed according to a standard normal, we use the Kolmogorov-Smirnov test to check this hypothesis and the skewness and kurtosis of the distributions to compare the different models.

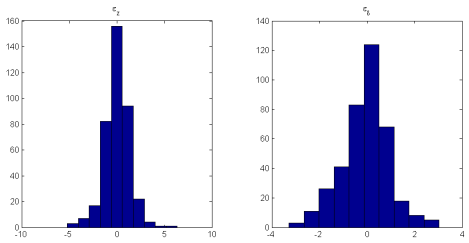

Figure 4.6: *Histogram for the residuals (only for the log-return process and the convenience yield process) using the Liu & Tang model*

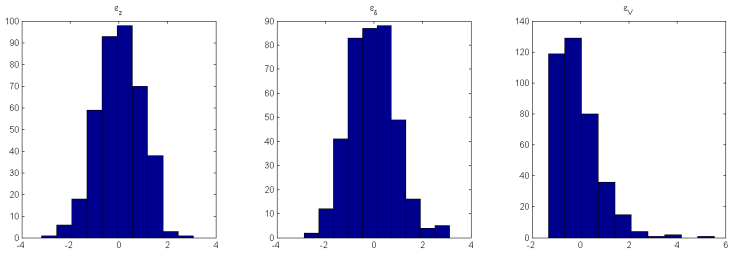

Figure 4.7: *Histogram for the residuals for the two-factor model*

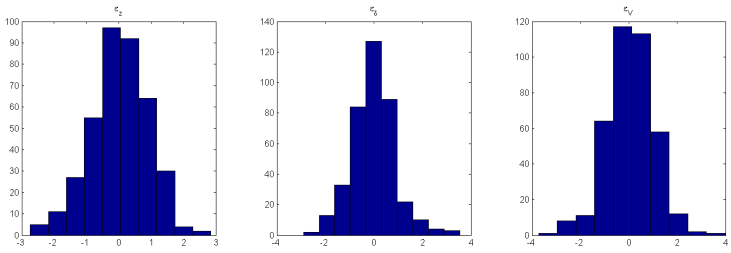

Figure 4.8: *Histogram for the residuals for the two factor model with jumps*

Table 4.2: *Analysis of the residuals under the historical measure for the three model, for each residual is reported the p-value of the Kolmogorov-Smirnov test, the skewness and the kurtosis of the distribution*

Residual		Liu-Tang	Model with Vol	Model with Vol & Jumps
ϵ_S	KS test	0.012	0.086	0.068
	skewness	-0.009	-0.174	-0.230
	kurtosis	5.499	3.03	3.241
ϵ_δ	KS test	0.069	0.049	0.047
	skewness	0.342	0.297	0.336
	kurtosis	3.671	3.313	3.763
ϵ_V	KS test	–	$6.31E-7$	0.0781
	skewness	–	1.708	-0.147
	kurtosis	–	8.497	3.973

For the risk neutral dynamics we evaluate the square root of the mean of quadratic errors (RMSE) and the absolute mean error (AME) for both the data sets: the data set used for parameter estimation (in the sample set ITS) and the data set outside the first one (out of the sample OTS). Results are shown in Table **??**.

Futures error		Liu-Tang	Model with Vol	Model with Vol & Jumps
ITS	RMSE	1.193	1.018	0.551
	AME	0.818	0.606	0.329
OTS	RMSE	0.561	1.045	0.508
	AME	0.418	0.648	0.327

As it is shown by p-value for the Kolmogorov-Smirnov test we have to reject the normality hypothesis for at least two of the residuals: the convenience yield and the volatility residuals, the presence of the period with the fall of WTI spot has a great impact in the sampled path, in particular we notice in the histograms (Fig. 4.6, 4.7, 4.8) of the residuals $\epsilon_\delta,\ \epsilon_V$ a skewed distribution , and with some

values far form the mode values that brought a much more large kurtosis than for a normal distribution. Model with jumps seems to mitigate the effect of leptokurtic distribution, explaining a part of the outliers variance value with an increased jump activity. The model with seasonality does not improve the performances of the model regarding $\epsilon_\delta,\ \epsilon_V$, but the p-value of the KS test for the spot residuals is much higher than the correspondent value for the other two models. The errors associated with the futures, both in the sample and out of the sample, suggest there is not a real improvement using seasonality, for jumps out of sample errors shows a better performances, due to the greater flexibility given by using three latent process than two in our particle filter approach. This improvement has a cost in required time to run the single PF algorithm: spanning observed data coming from 998 dates and using 2000 "particles", for the basic model we need 6.45 seconds, for the model including the seasonality term the time is equal (the number of latent process simulated is the same), the model including jumps requires 9.03 seconds. All the algorithms written in Matlab have been running on a i$3(3,07$ GHz$)$ pc.

4.8 Concluding Remarks

In this paper we analyzed three models in the attempt to capture the dynamics of WTI spot characteristics and the futures market structure. The first model is a model recently proposed by Liu and Tang (2011), which we used as a benchmark model, while the two proposed models are both variants of the previous one including as a new factor the volatility process and possibly jumps. We compared the models analyzing the residual to see what model best fits the observed data set, and we found evidence that, for our parameter set (in the sample analysis) and for a subsequent data set involving a panel of 50 dates for 12 different maturity time futures, in both cases the model with jumps performs much better than the other two model, while in out of the sample performances the benchmark model exhibited better performances than the proposed model without jumps. The inference procedure presented in this paper does not estimate the risk premium associated to volatility process (η_V), since the futures prices are not directly affected by this process. In the Appendix an explicit calculation of futures options price is provided for our models, but we did not use these prices in our estimation procedure. Both the length of the time series and the time required to get the price of each option contribute to make the algorithm quite time consuming, preventing to keep the required time to run the inference procedure including these prices within acceptable limits. In order to allow the algorithm to get an estimate also for the volatility risk premium, a possible choice could be that of taking into account random sampled subsets of all option on futures available quotes. This will be anyway the subject of future work.

4.9 Appendix: Future Options prices

To evaluate, under the $\mathcal{Q}$ measure, a general contingent G claim, whose payoff is determined by an underlying whose dynamics follows (4.6), with the seasonality function fixed at zero $g(t_{\text{year}}) = 0$, we refer to the PDE:

$$\begin{aligned}
&\frac{\partial H}{\partial t} + \frac{\partial H}{\partial x}(r_f + c - \lambda\mu_J^* - (1 + \frac{\sigma_\delta}{2})\delta_t - \frac{1}{2}V_t) + \frac{1}{2}\frac{\partial^2 H}{\partial x^2}(V_t + \sigma_\delta^2\delta_t) + \\
&\quad + \frac{\partial H}{\partial \delta}\left[\alpha(\bar{\delta} - \delta_t) - \eta_\delta\delta_t\right] + \frac{1}{2}\frac{\partial^2 H}{\partial \delta^2}\sigma^2\delta_t + \\
&\quad + \frac{\partial H}{\partial V}\left[\beta(\bar{V} - V_t) - \eta_V V_t\right] + \frac{1}{2}\frac{\partial^2 H}{\partial V^2}\xi^2 V_t + \\
&\quad + \frac{\partial^2 H}{\partial V \partial x}\rho\xi V_t + \frac{\partial^2 H}{\partial \delta \partial x}\rho_\delta\sigma\sigma_\delta\delta_t \\
&\quad - r_f H + \lambda\mathbb{E}\left[H(t, x_t + ln(1 + J), \delta_t, V_t) - H(t, x_t, \delta_t, V_t)\right] = 0
\end{aligned} \tag{4.17}$$

Following Heston [35] and Bakshi, Madan [9] we can decompose the value C at time t of an European option that gives the right at maturity time T to acquire a futures contract, whose deliver is fixed at $T^{'}$, at a fixed price K:

$$C(t, \tau, \tau^{'} - \tau) = G(t, \tau)\Pi(t, \tau) - KB(t, \tau)\Pi(t, \tau)$$

where $\tau = T - t$ and $\tau^{'} = T^{'} - t$; $B(t, \tau)$ is the price at time t of a discount bond and $G(t, \tau)$ is the price at time t of a forward which deliver at time $t + \tau$ the asset $S(t + \tau)$.
The decomposition has been proved by Yan [59]. Bakshi and Madan [9] proved that:

$$\begin{aligned}
B(t, \tau) &= f(t, \tau; 0) \\
G(t, \tau) &= \frac{1}{i}\frac{\partial f(t, \tau; 0)}{\partial \phi} \\
f_1(t, \tau; \phi) &= \frac{1}{iG(t, \tau)}\frac{\partial f(t, \tau; \phi)}{\partial \phi} \\
f_2(t, \tau; \phi) &= \frac{1}{G(t, \tau)}f(t, \tau; \phi)
\end{aligned} \tag{4.18}$$

where $f_1(t, \tau; \phi)$ and $f_2(t, \tau; \phi)$ are respectively the fourier transform of the two probability terms $\Pi_1(t, \tau)$ and $\Pi_2(t, \tau)$. Hence each term can be recovered once computed $f(t, \tau; \phi)$, that is the characteristic function of the underlying of the contingent claim (in this case the characteristic function of the logarithm of the future prices). Since $f(t, \tau; \phi)$ is the price of a contingent claim paying $\exp\{i\phi H(t + \tau, \tilde{\tau} - \tau)\}$ at maturity $t + \tau$, its value can be found solving (4.17)

with the terminal condition:

$$f(t+\tau;0) = \exp\{i\phi H(t+\tau, \tilde{\tau}-\tau)\} \tag{4.19}$$

If we use the trial solution:

$$f(t,\tau;\phi) = \exp\{\theta_0(\tau) + \theta_\delta(\tau)\delta_t + \theta_V(\tau)V_t + i\phi[x + A_0(t) + A_2(t)\delta_t]\}$$

in (4.17), we get the following ODE system.

$$\begin{cases} - & \dfrac{\partial\phi_0(\tau)}{\partial\tau} + i\phi(r_f + c - \lambda\mu_J^*) + \phi_\delta\alpha\bar{\delta} + \theta_V\beta\bar{V} = 0 \\ - & \dfrac{\partial A_1(\tau)}{\partial\tau} = 0 \\ - & \dfrac{\partial\phi_\delta(\tau)}{\partial\tau} - i\phi(1+\dfrac{\sigma_\delta^2}{2}) - \dfrac{\sigma_\delta^2}{2}\phi^2 - \phi_\delta(\alpha+\eta_\delta) + \frac{1}{2}\phi_\delta^2(\tau)\sigma^2 + i\phi\phi_\delta\rho_\delta\sigma_\delta\sigma = 0 \\ - & \dfrac{\partial\theta_V(\tau)}{\partial\tau} - \frac{1}{2}i\phi - \frac{1}{2}\phi^2 - \theta_V(\beta-\eta_V) + \frac{1}{2}\xi^2\theta_3^2 + i\phi\theta_V\rho\xi = 0 \end{cases} \tag{4.20}$$

where $\phi_\delta(\tau) = \theta_\delta(\tau) + i\phi A_2(\tilde{\tau}-\tau)$ and $\phi_0(\tau) = \theta_0(\tau) + i\phi A_0(\tilde{\tau}-\tau)$ with terminal condition:

$$\begin{cases} \phi_0(0) & = i\phi\beta_0(\tilde{\tau}) \\ \phi_\delta(0) & = i\phi\beta_\delta(\tilde{\tau}) \\ \theta_V(0) & = 0 \end{cases} \tag{4.21}$$

Solving the ODE system we get:

$$\theta_V = -\frac{2}{\xi^2}\left(\frac{\exp\{C_V\tau\}-1}{\frac{\exp\{C_V\tau\}}{A_V} - \frac{1}{B_V}}\right)$$

with

$$\begin{cases} C_V & = \sqrt{(\eta_V - i\phi\rho\xi + \beta)^2 + \xi^2(\phi^2 + i\phi)} \\ A_V & = \frac{1}{2}[(\beta + \eta_V - i\phi\rho\xi) + C_V] \\ B_V & = \frac{1}{2}[(\beta + \eta_V - i\phi\rho\xi) - C_V] \end{cases}$$

And:

$$\phi_\delta = -\frac{2}{\sigma^2}\frac{A_\delta\exp\{C_\delta\tau\} + B_\delta\bar{C}}{\exp\{C_\delta\tau\} + \bar{C}}$$

where:

$$\begin{cases} C_\delta &= \sqrt{(\alpha+\eta_\delta-i\phi\rho_\delta\sigma\sigma_\delta)^2+2\sigma^2 i\phi(1+\frac{\sigma_\delta^2}{2}-i\phi\frac{\sigma_\delta^2}{2})} \\ A_\delta &= \frac{1}{2}[(\eta_\delta+\alpha-i\phi\rho_\delta\sigma\sigma_\delta)+C_\delta] \\ B_\delta &= \frac{1}{2}[(\eta_\delta+\alpha-i\phi\rho_\delta\sigma\sigma_\delta)-C_\delta] \\ \tilde{C}_\delta &= \sqrt{(\alpha+\eta_\delta-\rho_\delta\sigma\sigma_\delta)^2+2\sigma^2} \\ \tilde{A}_\delta &= \frac{1}{2}[(\eta_\delta+\alpha-\rho_\delta\sigma\sigma_\delta)+\tilde{C}_\delta] \\ \tilde{B}_\delta &= \frac{1}{2}[(\eta_\delta+\alpha-\rho_\delta\sigma\sigma_\delta)-\tilde{C}_\delta] \\ \bar{C} &= \frac{i\phi\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}-1}{\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}}{\tilde{A}_\delta}-\frac{1}{\tilde{B}_\delta}}-A_\delta}{B_\delta-i\phi\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}-1}{\frac{\exp\{\tilde{C}_\delta\tilde{\tau}\}}{\tilde{A}_\delta}-\frac{1}{\tilde{B}_\delta}}} \end{cases}$$

Finally we solve for ϕ_0:

$$\begin{aligned} \phi_0 =& i\phi(rf+c-\lambda\mu_J^*)\tau-\frac{2\beta\bar{V}}{\xi^2}\left((B_V+C_V)\tau+\log\frac{A_V\exp\{-C_V\tau\}-B_V}{C_V}\right)- \\ &-\frac{2\alpha\bar{\delta}}{\sigma^2}\left((B_d+C_d)\tau+\log\frac{1+\exp\{-C_\delta\tau\}\tilde{C}}{1+\tilde{C}}\right)+ \\ &+i\phi\left((rf+c)\tilde{\tau}-\frac{2\alpha\bar{\delta}}{\sigma^2}\left((\tilde{B}_\delta+\tilde{C}_\delta)\tilde{\tau}+\right.\right. \\ &\left.\left.+\log\frac{\tilde{A}_\delta\exp\{-\tilde{C}_\delta\tilde{\tau}\}-\tilde{B}_\delta}{\tilde{A}_\delta-\tilde{B}_\delta}\right)\right)+\lambda_J\left(\exp\left\{i\phi\mu_J-\frac{1}{2}\phi^2\sigma_J^2\right\}-1\right) \end{aligned}$$

with:

$$\mu_J^*=\exp\left\{\mu_J+\frac{1}{2}\sigma_J^2\right\}-1$$

CHAPTER 5

Concluding remarks and future perspectives

In this thesis several affine models have been presented and tested with data observed in WTI spot market and WTI futures markets.
In particular, in Chapter 3 three models including volatility process have been proposed: a 3 factor model with two latent CIR process (volatility and convenience yield), proposed together with two variants of it, one model including a seasonality term and the other including jumps.
It resulted that adding a seasonality term does not affect significantly the model performances. In Chapter 4, two more models have been introduced as extended versions of models introduced in Chapter 3, allowing the spot process to be correlated both with volatility and with convenience yield process. These proposed models have been compared with the Liu and Tang model [39]. In both cases (double correlated in Chapter 4 and single correlated in Chapter 3) the models including a volatility process factor display better performances when the spot dynamics is allowed to include jump activity, showing better results than all the other discussed models, included the benchmark model proposed by Liu and Tang in [39].
Performances are compared both analyzing data fitting properties (evaluated by the residuals) and the errors in catching and reproducing the futures term structure. Resuming the results got for the time series analyzed, the models with volatility factor performed better than Liu and Tang model in describing the spot

dynamics, as it is possible to notice analyzing the spot residuals; moreover, jump activity, absorbing part of the variance of the spot dynamics, has a positive impact in the whole model performances, thus resulted in normally distributed residuals. As discussed by Yan [59] the volatility process has a positive impact also in describing the option on futures market. Hence, analysis including implied volatilities of options on futures will be the subject of our future research. The affine models allow us to find closed formulae for the characteristic of the option price function, by the method described by Bakshi and Madan [9] which further develops the method firstly proposed by Carr and Madan [13]. The closed form solutions, got following this theory, are reported in Chapter 3 and in Chapter 4. Since we wish to include implied volatility surfaces, the first step is to get the corresponding Black Scholes prices; hence, we can define a market error in analogy with what it has been done with the futures prices (modeling the observed prices as affected by a white noise error with respect to the theoretical prices) and provide the likelihood definition required in the inference process, as explained in Chapter 2 and implemented for futures prices in Chapter 3 and Chapter 4.
Some preliminary attempts have already been conducted, using the whole set of data and using the fast Fourier transform technique (FFT) to get model prices for options. The optimized algorithm running the FFT requires between 1 and 2 seconds to compute the prices at a given date for different strikes, but since the algorithm has to run for each date and for different maturities, the time required grows rapidly when the time series of analyzed data increases. To reduce further the time needed, it is our intention to restrict our analysis to "at the money" options (this allows us to avoid to interpolate in order to get the prices at different strike levels, as it is explained in [26]). To further reduce the time required it is possible to sample a subset of dates in which the option data are taken into account in the inference algorithm, otherwise (if the date does not belong to the selected subset) just the futures term structure and the spot value will be considered. At each run the subset of dates, in which the options are taken into account, is sampled again. This technique has already been implemented by Eraker in [22] to reduce the time required by the Gibbs sampling algorithm used to get inference for Bates model using $S\&P500$ spot and options values. Reducing the analyzed data, in the way just described, it is possible to keep the running time within the order of a few days, even using a single core pc. Other possible research lines could explore the code parallelizing attempt; some efforts in the general technique has already been done by Pitt et al. [40].

Bibliography

[1] M. Briers A. Doucet and S. Senecal. Efficient block sampling strategies for sequential monte carlo methods. *Journal of Computational and Graphical Statistics*, 15(3):693–711, 2006.

[2] S. Godsill A. Doucet and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

[3] H.S. Stern A. Gelman, J.B. Carlin and D.B. Rubin. *Bayesian data analysis*. Chapman & Hall CRC, 2009.

[4] W.R. Gilks A. Gelman and G.O. Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Probability*, 7, N. 1:110–120, 1997.

[5] T.B. Schon A. Wills and B. Ninness. Parameter estimation for discrete-time nonlinear systems using em. In *Proc. 17th IFAC World Congress*, 2008.

[6] C. Andrieu and A. Doucet. Online expectation-maximization type algorithms for parameter estimation in general state space models. In *JProc. IEEE Conf. ICASSP*, 2003.

[7] C. Andrieu and A.M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. In D. Crisan et B. Rozovsky, editor, *Handbook of Nonlinear Filtering*. Oxford University Press, 2008.

[8] M. S. Johannes B. Eraker and N. Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, LVIII No. 3, 2003.

[9] G. Bakshi and D. Madan. Spanning and derivative-security valuation. *Journal of Financial Economics*, 55:205–238, 2000.

[10] A. Doucet C. Andrieu and R. Holenstein. Particle markov chain monte carlo methods. *Journal of Royal Statistical Society B*, J 72, Part 3:269–342, 2010.

[11] A. Monfort C. Gourieroux and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S):85–118, 1993.

[12] H. Li C. Yu and M. T. Wells. Mcmc estimation of lévy jump models using stock and option prices. *Mathematical Finance*, 21, No. 3:383–422, 2011.

[13] Peter Carr, Dilip B. Madan, and Robert H Smith. Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2:61–73, 1999.

[14] E.J. Chen and W.D. Kelton. A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality. *Simulation*, 83 N.10:683–694, 2007.

[15] M. Chernov and E. Ghysels. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56:407–458, 2000.

[16] G. Cortazar and E.S. Schwart. Implementing a stochastic model for oil futures prices. *Energy Economics*, 25, Issue 3:215–238, 2003.

[17] J. Pan D. Duffie and K.J. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68 No.6:1343–1376, 2000.

[18] Q. Dai and K. Singleton. Specification analysis of affine term structure model. *Journal of Finance*, 55:1943–1978, 2000.

[19] D. Duffie and K.J. Singleton. Simulated moments estimation of markov models of asset prices. *Econometrica*, 61 No.4:929–952, 1993.

[20] N.G. Polson E. Jacquier and P.E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 12, Issue 4:371–89, 1994.

[21] N.G. Polson E. Jacquier and P.E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122:185–212, 2004.

[22] B. Eraker. Do stock pices and volatility jumps? reconciling evidence from spot and option prices. *The Journal of Finance*, LIX No. 3, 2004.

[23] A. Eydeland and H. Geman. Pricing power derivatives. *RISK*, 11:71–73, 1998.

[24] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management*. Wiley Finance, 2003.

[25] C.S. Forbes, G.M. Martin, and J. Wright. Bayesian estimation of a stochastic volatility model using option and spot prices. Monash Econometrics and Business Statistics Working Papers 2/02, Monash University, Department of Econometrics and Business Statistics, 2002.

[26] G. Fusai and A. Roncoroni. *Implementing Models in Quantitative Finance: Methods and Cases*. Springer Finance, 2008.

[27] H. Geman. Scarsity and price volatility in oil markets. Trading technical report, EDF, 2000.

[28] H. Geman. *Commodities and Commodity derivatives*. Wiley Finance, 2005.

[29] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *IN BAYESIAN STATISTICS*, pages 169–193. University Press, 1992.

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

[31] W.R. Gilks and C. Berzuini. Following a moving target-monte carlo inference for dynamic bayesian models. *Journal of Royal Statistical Society B*, 63, Part 1:127–146, 2001.

[32] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

[33] G.B. Gorton, F. hayashi, and K.G. Rouwenhorst. The fundamentals of commodity futures returns. Yale ICF Working Papers 07/08, Yale University, International Center for Finance, 2002.

[34] L.A. Grzelak and C.W. Oosterlee. On cross-currency models with stochastic volatility and correlated interest rates. *Applied Mathematical Finance*, 19 No.1:1–35, 2012.

[35] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6 (2):327–343, 1993.

[36] S. Hikspoors. *Multi-Factor Energy Price Models and Exotic Derivatives Pricing*. PhD thesis, University of Toronto, 2008. Ph.D. Thesis (advisor: Prof. S.Jaimungal).

[37] A. Javaheri. *Inside volatility arbitrage*. Wiley Finance, 2005.

[38] Borus Jungbacker and SiemJan Koopman. Parameter estimation and practical aspects of modeling stochastic volatility. In Thomas Mikosch, Jens-Peter Kreil, Richard A. Davis, and Torben Gustav Andersen, editors, *Handbook of Financial Time Series*, pages 313–344. Springer Berlin Heidelberg, 2009.

[39] P. Liu and K. Tang. The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield. *Journal of Empirical Finance*, 18(2):211–224, 2011.

[40] P. Giordani M.K. Pitt, R. Silva and R. Kohn. Auxiliary particle filtering within adaptive metropolis-hastings sampling. http://arxiv.org/abs/1006.1914v1, 2010.

[41] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[42] N. Gordon M.S. Arulampalam, S. Maskell and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. In *IEEE Transaction on signal processing*, 2002. VOL. 50, NO. 2.

[43] N. Polson M.S. Johannes and J.R. Stroud. Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies*, 22:2559–2599, 2009.

[44] D.J. Salmond N.J. Gordon and A.F.M. Smith. Novel approach to nonlinear/non−gaussian bayesian state estimation. In *IEE-Proceedings-F, 140*, pages 107–113, 1993.

[45] J. Pan. The jump-risk premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics*, 63:3–50, 2002.

[46] R.S. Pindyck. Inventories and the short-run dynamics of commodity prices. *The RAND Journal of Economics*, 25, No. 1, 1994.

[47] R.S. Pindyck. The dynamiocs of commodity spot and futures markets: a primer. *The Energy Journal*, 22, No. 3, 2001.

[48] R.S. Pindyck. Volatility in natural gas and oil markets. *The Journal of Energy and Development*, 30, No. 1, 2004.

[49] M.K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94, No. 446:590–599, 1999.

[50] D. Ribeiro and S.D. Hodges. A two-factor model for commodity prices and futures valuation. In *EFMA 2004 Basel Meetings Paper*, 2004.

[51] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[52] P.R. Robert and G. Casella. *Introducing Monte Carlo Mehods with R*. Springer, 2010.

[53] E.S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52, Part 3:923–973, 1997.

[54] S.E. Shreeve. *Volume II: Continuous-Time Models*. Springer-Verlag, 2004.

[55] A. Bagchi S.I. Aihara and S. Saha. Estimating volatility and model parameters of stochastic volatility models with jumps using particle filter. In *Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea*, 2008.

[56] B.J. Smith. boa: An r package for mcmc output convergence assessment and posterior inference. *Journal of Statistical Software*, 21, Issue 11, 2007.

[57] H. Working. The theory of price of storage. *The American Economic Review*, 6:1254–1262, 1949.

[58] B.D. Wright and J.C. Williams. A theory of negative prices for storage. *The Journal of Futures Markets*, 20:59 – 71, 2000.

[59] X.S. Yan. Valuation of commodity derivatives in a new multi-factor model. *Review of Derivatives Research*, 5:251–271, 2007.