Calibration of Stochastic Convenience Yield Models For Crude Oil Using the Kalman Filter

A.Krul
Adriaan.Krul@INGbank.com
tw1249266
Literature report

10-12-07 20-2-2008

Contents

Chapter 1

Introduction

1.1 Introduction

A future contract is an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price. Nowadays, traders calibrate the so-called convenience yield (CY), δ_t, via market data every two days, using the future contract. Convenience yield is the benefit or premium associated with holding an underlying product or physical good, rather than the contract or derivative product. Sometimes, due to irregular market movements such as an inverted market (this is when the short-term contract prices are higher than the long-term contracts), the holding of an underlying good or security may become more profitable than owning the contract or derivative instrument, due to its relative scarcity versus high demand.

An example would be purchasing physical bales of oil rather than future contracts. Should their be a sudden shock-situation wherein the demand for oil increases, the difference between the first purchase price of the oil versus the price after the shock would be the convenience yield. The market shows however that the CY behaves stochastically and has a mean-reverting property, see e.g. [5]. The CY may therefore be modelled by an Ornstein-Uhlenbeck (OU) process. The disadvantage of this process is that it allows for negative CY which can result in cost and carry arbitrage possibilities. To prevent this from happening we assume that the CY follows a Cox-Ingersoll-Ross (CIR) model. For both processes, the spot price follows a geometrical brownian motion.

For commodities, the spot prices and the CY can not be observed directly from the market. To model both stochastic processes (OU and CIR), the non-observability of these *state variables* is difficult. Since the future prices of commodities are widely observered and traded on the market we use a method that links these actual observations with the latent state variables. This method is known as the *Kalman Filter*, (KF). The main idea of the (KF) is to use observable (futures) variables to reconstitute the value of the non-obsvervable (spot prices and CY). We use futures of light crude oil ranging for the period from 01-02-2002 until 25-01-2008.

For applying the Kalman Filter, we assume an affine form for the closed form solution of the future prices for both the OU and the CIR process. By assuming this, solving the stochastic differential equation of both processes is a lot easier and having the closed form solution of the future prices will make the use of the KF considerably easier.

The aim of this thesis is to implement the KF for both OU and CIR and compare the results with the market data. Other commodities are numerically tested and we will try to price options on commodities using this method. The outline of this report is as follows; in Chapter 2 and 3 analytical results are given for the OU and CIR processes (resp.). In Chapter 4, the KF is explained in detail and it starts with an introductory example. Chapter 5 applies the KF to the OU process and discusses the numerical results. In Chapter 6 the KF is set up for the CIR process, but not yet tested numerically. The report ends with a conclusion and further research.

Chapter 2

Stochastic convenience yield model following an Ornstein-Uhlenbeck process

In this chapter analytical results are given for the stochastic CY when it follows an Ornstein-Uhlenbeck process. The assumption is that the spot price S_t follows a geometrical Brownian motion, i.e.,

$$dS_t = (\mu - \delta_t)S_t dt + \eta S_t dW_t, \tag{2.1}$$

where μ is the drift term, η is the volatility term, and W_t a standard Brownian motion. The form of the convenience yield (δ_t) is a result of the studies of Gibson and Schwartz (see [5]) where they find empirical evidence that the convenience yield will have a mean-reverting property, i.e.,

$$d\delta_t = k(\alpha - \delta_t)dt + \sigma dZ_t. \tag{2.2}$$

In (2.2), α is the long range mean to which δ_t tends to revert, k the speed of adjustment, σ the volatility term and Z_t a standard Brownian motion. Here $dW_t dZ_t = \rho dt$. The solution of (2.1) is given by

$$S_T = S_t \exp\left\{(\mu - \frac{1}{2}\eta^2)(T-t) + \eta \int_t^T dW_s\right\}, \tag{2.3}$$

and the solution of (2.2) is

$$\delta_T = \theta\delta_t + (1-\theta)\alpha + \sigma e^{-kT}\int_t^T e^{ks} dZ_s, \tag{2.4}$$

where $\theta = e^{-k(T-t)}$.

2.1 Mean and variance of the Ornstein-Uhlenbeck process

Apply Ito's lemma on $f(\delta_T, t) = \delta_T e^{kt}$ which results in

$$\begin{aligned} df(\delta_t, t) &= k\delta_t e^{kt} dt + e^{kt} d\delta_t \\ &= e^{kt} k\alpha dt + \sigma e^{kt} dZ_t. \end{aligned}$$

This implies

$$\delta_t e^{kt} = \delta_0 + k\alpha \int_0^t e^{ks} ds + \sigma \int_0^t e^{ks} dZ_s.$$

Rearranging and taking expectation yields

$$\mathbb{E}[\delta_t] = \delta_0 e^{-kt} + \alpha(1 - e^{-kt}) = \delta_0\theta + \alpha(1-\theta),$$

where $\theta = e^{-kt}$. For the variance we calculate the covariance matrix by applying the Ito Isometry

$$\begin{aligned}\mathrm{Cov}(\delta_s, \delta_t) &= \mathbb{E}[(\delta_s - \mathbb{E}[\delta_s])(\delta_t - \mathbb{E}[\delta_t])] \\ &= \mathbb{E}[\sigma \int_0^s e^{-k(u-s)} dZ_u \sigma \int_0^t e^{-k(v-t)} dZ_v] \\ &= \sigma^2 e^{-k(s+t)} \mathbb{E}[\int_0^s e^{ku} dZ_u \int_0^t e^{kv} dZ_v] \\ &= \frac{\sigma^2}{2k} e^{-k(s+t)} (e^{2k(\min(s,t))} - 1).\end{aligned}$$

For $s = t$ this gives $\mathbb{V}\mathrm{ar}[\delta_t] = \frac{\sigma^2}{2k}(1-\theta^2)$. We conclude that δ_t is $\mathcal{N}\left(\theta\delta_t + (1-\theta)\alpha, \frac{\sigma^2}{2k}(1-\theta^2)\right)$.

2.2 Market price of risk

For the interest rate we do not have to build in a stochastic model so we keep r deterministic (constant) [4]. Since the CY is non-traded, the convenience yield risk cannot be hedged and it will have a market price or risk, λ_δ associated with it. To see how this is inserted in (2.1) and (2.2), consider the standard geometrical brownian motion

$$\frac{dS_t}{S_t} = \mu dt + \eta dW_t,$$

and taking at both sides the expectation and rearranging to get

$$\mu = \frac{1}{dt}\mathbb{E}[\frac{dS_t}{S_t}].$$

We see that μ is the expected return per unit time of a stock S_t in an economy where the only source of risk is the Brownian motion. We can write

$$\mu = r - \delta_t + b,$$

where r denotes the risk-free rate and b the risk premium. Since we assume that η is non-zero we define

$$\lambda = \frac{b}{\eta}$$

to be the risk premium expected by investors per unit of volatility. It is usually refered to as the market price of equity risk per unit of η. We now have

$$\mu = r - \delta_t + \lambda\eta \quad \text{and so} \quad \lambda = \frac{\mu + \delta_t - r}{\eta}.$$

Inserting this in (2.1) we get

$$\frac{dS_t}{S_t} = \mu dt + \eta(dW_t - (\frac{\mu + \delta_t - r}{\eta})dt) \quad \Rightarrow \quad \frac{dS_t}{S_t} = (r - \delta_t)dt + \eta d\tilde{W}_t, \tag{2.5}$$

where $\tilde{W}_t$ is the standard Brownian motion under the risk-neutral measure. For (2.2) it is easily shown that

$$d\delta_t = k((\alpha - \frac{\sigma\lambda}{k}) - \delta_t)dt + \sigma d\tilde{Z}_t,$$

holds. Again, $d\tilde{Z}_t$ is the standard brownian motion under the risk-neutral measure and $d\tilde{Z}_t d\tilde{W}_t = \rho dt$ as before. Now we have the system of two joint stochastic processes

$$\begin{cases} \dfrac{dS_t}{S_t} = (r - \delta_t)dt + \eta d\tilde{W}_t, \\ d\delta_t = k((\alpha - \frac{\sigma\lambda}{k}) - \delta_t)dt + \sigma d\tilde{Z}_t. \end{cases} \tag{2.6}$$

Note that this derivation is based on the Girsanov theorem.

2.3 The value of a future delivery

Define $Y_T = Y(T, S_T, \delta_T)$ to be the payoff at time T. The discounted expected value of Y_T at time t is given by

$$V_t[Y_T] := e^{-r(T-t)}\mathbb{E}[Y_T]. \tag{2.7}$$

From (2.5) we get

$$\int_t^T dW_s = \int_t^T d\tilde{W}_s - \frac{\mu - r}{\eta}(T-t) - \frac{1}{\eta}(X_T - X_t), \tag{2.8}$$

where $X_t = \int_0^t \delta_s ds$.
From (2.2) we get

$$\int_t^T d\delta_s = \delta_T - \delta_t = k\alpha t - k\int_t^T \delta_s ds + \sigma\int_t^T dZ_s = k\alpha(T-t) - k(X_T - X_t) + \sigma\int_t^T dZ_s. \tag{2.9}$$

Now insert (2.4) to get

$$X_T = X_t + (1-\theta)\frac{1}{k}(\delta_t - \alpha) + \alpha(T-t) + \frac{\sigma}{k}\int_t^T dZ_s - \frac{\sigma}{k}e^{-kT}\int_t^T e^{ks}dZ_s. \tag{2.10}$$

For the future price we consider the case $Y_T \equiv S_T$ and we compute $V_t[S_T]$. The current value of a claim on a future delivery of the commodity on the future date t is

$$\begin{aligned} V_t[S_T] \quad = \quad & S_t \exp\Big\{ \Big[-\alpha + \frac{1}{k}(\sigma\lambda - \sigma\eta\rho) + \frac{1}{2}(\frac{1}{k})^2\sigma^2\Big](T-t) \\ & - \frac{1}{k}\Big[\delta_t - \alpha + \frac{1}{k}(\sigma\lambda - \sigma\eta\rho) + (\frac{1}{k})^2\sigma^2\Big](1-\theta) \\ & + \frac{1}{2}(\frac{1}{k})^2\frac{\sigma^2}{2k}(1-\theta^2)\Big\}. \end{aligned} \tag{2.11}$$

Proof. From (2.7) and (2.3) we get

$$e^{-r(T-t)}S_T = S_t \exp\Big\{(\mu - r - \frac{1}{2}\eta^2)(T-t) + \eta\int_t^T dW_s\Big\}. \tag{2.12}$$

Multiplying (2.8) with η, it follows together with (2.10)

$$\begin{aligned} \eta\int_t^T W_s \quad = \quad & -(T-t)(\mu - r + \alpha - \frac{1}{k}\sigma\lambda) + \frac{1}{k}(\alpha - \delta_t + \frac{1}{k}\sigma\lambda)(1-\theta) + \eta\int_t^T d\tilde{W}_s \\ & - \frac{1}{k}\sigma\int_t^T dZ_s + \frac{1}{k}\sigma e^{-kT}\int_t^T e^{ks}dZ_s. \end{aligned} \tag{2.13}$$

Now we can write (2.12) as follows

$$\begin{aligned} e^{-r(T-t)}S_T \quad = \quad & S_t \exp\Big\{-(\frac{1}{2}\eta^2 + \alpha - \frac{1}{k}\sigma\lambda)(T-t) + \frac{1}{k}(\alpha - \delta_t \\ & - \frac{1}{k}\sigma\lambda)(1-\theta) + \eta\int_t^T d\tilde{W}_s - \frac{1}{k}\sigma\int_t^T d\tilde{Z}_s + \frac{1}{k}\sigma e^{-kT}\int_t^T d\tilde{Z}_s\Big\} \\ = \quad & S_t e^z. \end{aligned} \tag{2.14}$$

From basic stochastic calculus it follows

$$\begin{aligned} \tilde{\mu} \quad = \quad & \mathbb{E}[z] \\ = \quad & -(\frac{1}{2}\eta^2 + \alpha - \frac{1}{k}\sigma\lambda)(T-t) + \frac{1}{k}(\alpha - \delta_t - \frac{1}{k}\sigma\lambda)(1-\theta). \end{aligned} \tag{2.15}$$

For calculating $\mathbb{V}\text{ar}[z]$ we notice that for a general process $g \in L^2[t,T]$ it follows

$$\mathbb{E}\left[\left(\int_t^T g(s)dW_s\right)^2\right] = \int_t^T \mathbb{E}[g(s)^2]ds.$$

Using this we can calculate the following expectations

$$\mathbb{E}\left[\eta^2\left(\int_t^T d\tilde{W}_s \int_t^T d\tilde{W}_s\right)\right] = \eta^2 \int_t^T \mathbb{E}[1^2]dt = \eta^2(T-t) \tag{2.16}$$

$$\mathbb{E}\left[\left(\frac{1}{k}\sigma \int_t^T d\tilde{Z}_s\right)^2\right] = \frac{\sigma^2}{k^2}\int_t^T \mathbb{E}[1^2] = \frac{\sigma^2}{k^2}(T-t) \tag{2.17}$$

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{k}e^{-kT}\int_t^T e^{ks}d\tilde{Z}_s\right)^2\right] &= \left[\frac{\sigma^2}{k^2}e^{-2kT}\int_t^T \mathbb{E}[e^{2ks}]ds\right] \\ &= \frac{\sigma^2}{k^2}e^{-2kT}[\frac{1}{2k}e^{2kT} - \frac{1}{2k}e^{2kt}] \\ &= \frac{\sigma^2}{k^2}\frac{1}{2k}[1-\theta^2]\end{aligned} \tag{2.18}$$

$$\mathbb{E}\left[\left(\eta\int_t^T d\tilde{W}_s\right)\left(\frac{1}{k}\sigma\int_t^T d\tilde{Z}_s\right)\right] = \eta\frac{\sigma}{k}\rho(T-t) \tag{2.19}$$

$$\begin{aligned}\mathbb{E}\left[\left(\eta\int_t^T d\tilde{W}_s\right)\left(\frac{1}{k}\sigma e^{-kT}\int_t^T e^{ks}d\tilde{Z}_s\right)\right] &= e^{-kT}\frac{\eta\sigma}{k}\rho\int_t^T \mathbb{E}[e^{ks}]ds \\ &= e^{-kT}\frac{\eta\sigma\rho}{k}\frac{1}{k}[e^{kT}-e^{kt}] \\ &= \frac{\eta\sigma\rho}{k^2}(1-\theta)\end{aligned} \tag{2.20}$$

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{k}\sigma\int_t^T d\tilde{Z}_s\right)\left(\frac{1}{k}\sigma e^{-kT}\int_t^T e^{ks}d\tilde{Z}_s\right)\right] &= \frac{1}{k^2}\sigma^2 e^{-kT}\rho\int_t^T \mathbb{E}[e^{ks}]ds \\ &= \frac{1}{k^2}(\sigma^2 e^{-kT})\rho\frac{1}{k}(e^{kT}-e^{kt}) \\ &= \frac{\sigma^2\rho}{k^3}(1-\theta).\end{aligned} \tag{2.21}$$

Finally we get

$$\begin{aligned}\tilde{\sigma}^2 &= \mathbb{E}[(z^2)] - (\mathbb{E}[z])^2 \\ &= (\eta^2 - 2\frac{1}{k}\sigma\eta\rho + \frac{1}{k^2}\sigma^2)(T-t) \\ &+ 2(\frac{1}{k^2}\sigma\eta\rho - \frac{1}{k^3}\sigma^2)(1-\theta) \\ &+ \frac{1}{k^2}\frac{\sigma^2}{2k}(1-\theta^2),\end{aligned} \tag{2.22}$$

where $\theta = e^{-k(T-t)}$ as before. From the standard formula of the expected value of a lognormal random variable we get

$$V_t[S_T] = S_t \exp\left\{\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right\}. \tag{2.23}$$

□

The future price of a contract on a commodity with time to maturity $\tau = T - t$ is given by

$$F(S,\delta,\tau) := e^{r\tau} V_t[S_T]. \tag{2.24}$$

This follows from the absence of risk-free arbitrage opportunities, i.e. we must have

$$V_t[S_T - F] = 0. \tag{2.25}$$

In order to find the PDE to which $F(S,\delta,\tau)$ satisfies we use the Feynman-kač theorem.

Theorem 2.1. *Feynman-Kač*
Suppose the underlying processes $y_1(t), y_2(t), .., y_n(t)$ follow the stochastic differential equation

$$dy_i = \mu_i(y_1, y_2, ..y_n, t)dt + \sigma_i(y_1, y_2, .., y_n, t)dW_i.$$

Then the function

$$g(y_1, y_2, .., y_n, t) = \mathbb{E}_{y_1, y_2, .., y_n, t}[f(y_1(T), .., y_n(T))]$$

is given by the solution of the partial differential equation

$$\frac{\partial g}{\partial t} + \sum_{i=1}^{n} \mu_i \frac{\partial g}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^{n} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2 g}{\partial y_i \partial y_j} = 0,$$

subject to

$$g(y_1, y_2, .., y_n, T) = f(y_1, y_2, .., y_n),$$

where $\rho_{ij} = Cov(dW_i, dW_j)/dt$.

$F(S,\delta,\tau)$ satisfies the PDE

$$\begin{cases} \frac{1}{2}\eta^2 S^2 F_{SS} + \rho\eta\sigma S F_{S\delta} + \frac{1}{2}\sigma^2 F_{\delta\delta} + (r-\delta)SF_S + (k(\alpha-\delta)-\lambda)F_\delta - F_\tau = 0, \\ F(S,\delta,0) = S_0. \end{cases} \tag{2.26}$$

A similar result exists for $V(S,\delta,t) := V_t[Y_T]$. The $V(S,\delta,t)$ satisfies the PDE

$$\begin{cases} \frac{1}{2}\eta^2 S^2 V_{SS} + \rho\eta\sigma S V_{S\delta} + \frac{1}{2}\sigma^2 V_{\delta\delta} + (r-\delta)SV_S + (k(\alpha-\delta)-\lambda)V_\delta - rV_\tau = 0, \\ V(S,\delta,T) = Y(S,\delta). \end{cases} \tag{2.27}$$

2.4 Price valuation of a European call option

The value of a European call option at time t is given by

$$V_t[(S_T - K)^+] = V_t[S_T]N[d_1] - e^{-r(T-t)}KN[d_2], \tag{2.28}$$

where $N[\cdot]$ is the standard cumulative distribution function and we define

$$\begin{aligned} d_1 &= \frac{\ln(V_t[S_T]/K) + r(T-t) + \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}, \\ d_2 &= \frac{\ln(V_t[S_T]/K) + r(T-t) - \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}, \end{aligned}$$

where $\tilde{\sigma}$ is given by (2.22).

Proof. Apply (2.7) to (2.28) to obtain

$$V_t[(S_T - K)^+] = e^{-r(T-t)}\mathbb{E}_t[(S_T - K)^+]$$

and by using (2.14) we get

$$V_t[(S_T - K)^+] = \mathbb{E}[(S_T e^z - e^{-r(T-t)}K)^+].$$

Since z is $\mathcal{N}(\tilde{\mu}, \tilde{\sigma})$ we get

$$\begin{aligned} V_t[(S_T - K)^+] &= \exp\left\{\tilde{\mu} + \frac{1}{2}\tilde{\sigma}\right\} S_t N \left[\frac{\ln(S_T/K) + r(T_t) + \tilde{\mu} + \tilde{\sigma}^2}{\tilde{\sigma}}\right] \\ &- \exp\left\{-r(T-t)\right\} N \left[\frac{\ln(S_T/K) + r(T_t) + \tilde{\mu}}{\tilde{\sigma}}\right]. \end{aligned}$$

Together with (2.23) the result immediately follows. □

Chapter 3

Stochastic convenience yield model following a Cox-Ingersoll-Ross process

In this chapter, instead of (2.2) we assume that the convenience yield follows a Cox-Ingersoll-Ross (CIR) process and where the spot price follows a geometrical brownian motion with a time-varying volatility, which is proportional to the square root of the instantaneous convenience yield level, i.e.,

$$\begin{cases} \dfrac{dS_t}{S_t} = (\mu(\cdot) - \delta_t)dt + \eta\sqrt{\delta_t}dW_t, \\ d\delta_t = k(\alpha - \delta_t)dt + \sigma\sqrt{\delta_t}dZ_t, \end{cases} \tag{3.1}$$

In (3.1), S is the price of the underlying, δ is the CY, μ is the drift term, η is the volatility term of $\frac{dS_t}{S_t}$, W_t a standard Brownian motion, α is the long range mean to which δ_t tends to revert, k the speed of adjustment, σ the volatility term of $d\delta_t$, and Z_t a standard Brownian motion. Here $dW_t dZ_t = \rho dt$. The reason why we assume that the CY follows a CIR process is the nonnegativity. A negative CY would make the forward prices go up at more than the interest rate and provide some kind of cash and carry arbitrage through buying the spot commodity and selling a forward. The CIR process excludes negative CY.

3.1 Mean and variance of the CIR process

The integral equation for δ_t is given by

$$\delta_t = \delta_0 + k\int_0^t (\alpha - \delta_u)du + \sigma\int_0^t \sqrt{\delta_u}dZ_u.$$

Taking expectations on both sides and differentiating yields

$$\frac{d}{dt}\mathbb{E}[\delta_t] = k(\alpha - \mathbb{E}[\delta_t]) \quad \Rightarrow \quad \frac{d}{dt}e^{kt}\mathbb{E}[\delta_t] = e^{kt}[k\mathbb{E}[\delta_t] + \frac{d}{dt}\mathbb{E}[\delta_t]] = e^{kt}k\alpha.$$

This leads to

$$e^{kt}\mathbb{E}[\delta_t] - \delta_0 = k\alpha\int_0^t e^{ku}du = \alpha(e^{kt} - 1) \quad \Rightarrow \quad \mathbb{E}[\delta_t] = \alpha + e^{-kt}(\delta_0 - \alpha) = e^{-kt}\delta_0 + (1 - e^{-kt})\alpha.$$

Remark. We see that if $\delta_0 = \alpha$ then $\mathbb{E}[\delta_t] = \alpha\ \forall t$. If $\delta_0 \neq \alpha$, then δ_t exhibits mean reversion, i.e.

$$\lim_{t\to\infty}\mathbb{E}[\delta_t] = \alpha.$$

For the variance we calculate first $d\delta_t^2$ via Ito. Define $f(x) = x^2$. We have

$$\begin{aligned} d\delta_t^2 &= df(\delta_t) \\ &= f'(\delta_t)d\delta_t + \frac{1}{2}f''(\delta_t)d\delta_t d\delta_t \\ &= 2\delta_t[k(\alpha-\delta_t)dt + \sigma\sqrt{\delta_t}dZ_t] + [k(\alpha-\delta_t)dt + \sigma\sqrt{\delta_t}dZ_t]^2 \\ &= 2\alpha k\delta_t dt - 2k\delta_t^2 dt + 2\sigma\delta_t^{3/2}dZ_t + \sigma^2\delta_t dt \\ &= (2k\alpha+\sigma^2)\delta_t dt - 2k\delta_t^2 dt + 2\sigma\delta_t^{3/2}dZ_t. \end{aligned}$$

This leads to

$$\delta_t^2 = \delta_0^2 + (2k\alpha+\sigma^2)\int_0^t \delta_u du - 2k\int_0^t \delta_u^2 du + 2\sigma\int_0^t \delta_u^{3/2}dZ_u.$$

Taking expectation on both sides and differentiating with respect to t yields

$$\frac{d}{dr}\mathbb{E}[\delta_t^2] = (2k\alpha+\sigma^2)\mathbb{E}[\delta_t] - 2k\mathbb{E}[\delta_t^2].$$

From this we get

$$\begin{aligned} \frac{d}{dt}e^{2kt}\mathbb{E}[\delta_t^2] &= e^{2kt}[2k\mathbb{E}[\delta_t^2] + \frac{d}{dt}\mathbb{E}[\delta_t^2]] \\ &= e^{2kt}(2k\alpha+\sigma^2)\mathbb{E}[\delta_t]. \end{aligned}$$

It follows that

$$\mathbb{E}[\delta_t^2] = \frac{\alpha\sigma^2}{2k} + \alpha^2 + (\delta_0-\alpha)(\frac{\sigma^2}{k}+2\alpha)e^{-kt} + (\delta_0-\alpha)^2\frac{\sigma^2}{k}e^{-2kt} + \frac{\sigma^2}{k}(\frac{\alpha}{2}\delta_0)e^{-2kt}.$$

We finally have

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\delta_t] &= \mathbb{E}[\delta_t^2] - (\mathbb{E}[\delta_t])^2 \\ &= \alpha\frac{\sigma^2}{2k}(1-e^{-kt})^2 + \delta_0\frac{\sigma^2}{k}(e^{-kt}-e^{-2kt}). \end{aligned} \tag{3.2}$$

3.2 The logprice and inserting the market price of risk

By defining $x = \ln S_t$ and via Ito, the log price can be derived.

$$\begin{aligned} dx &= \left((\mu(\cdot)-\delta_t)S_t\frac{1}{S_t} + 0 + \frac{1}{2}\eta^2\delta_t S_t^2(-\frac{1}{S_t^2})\right)dt + \eta\sqrt{\delta_t}S_t\frac{1}{S_t}dW_t \\ &= \left((\mu(\cdot)-\delta_t) - \frac{1}{2}\eta^2\delta_t\right)dt + \eta\sqrt{\delta_t}dW_t. \end{aligned} \tag{3.3}$$

Again via the same explanation given in chapter 1 we insert the market price of risk into (3.3). This leads to the two joint stochastic process

$$\begin{cases} \dfrac{dS_t}{S_t} = (r-\delta_t)dt + \eta\sqrt{\delta_t}d\tilde{W}_t, \\ d\delta_t = (k(\alpha-\delta_t)-\lambda_{\delta_t})dt + \sigma\sqrt{\delta_t}d\tilde{Z}_t, \end{cases} \tag{3.4}$$

The log price process is given by

$$dx = \left(r - \left(1+\frac{1}{2}\eta^2\right)\delta_t\right)dt + \eta\sqrt{\delta_t}d\tilde{W}_t. \tag{3.5}$$

3.3 Partial differential equation for the future prices

Following the approach as discussed in Section 2.5, we have the following PDE for the future price $F(S,\delta,\tau)$

$$\begin{cases} \frac{1}{2}\eta^2\delta S^2 F_{SS} + \rho\eta\sigma\delta S F_{S\delta} + \frac{1}{2}\sigma^2\delta F_{\delta\delta} + (r-\delta)SF_S + (k(\alpha-\delta)-\lambda)F_\delta - F_\tau = 0, \\ F(S,\delta,0) = S_0. \end{cases} \tag{3.6}$$

Assuming (see e.g. [2]) this PDE has an affine form solution

$$\begin{cases} F(S,\delta,\tau) = Se^{A(\tau)-B(\tau)\delta}, \\ A(0)=0, \quad B(0)=0. \end{cases} \tag{3.7}$$

To find $B(\tau)$ and $A(\tau)$ we notice that (by substituting (3.7) into (3.6))

$$\begin{aligned} \frac{1}{2}\sigma^2 B^2 + (k-\rho\eta\sigma)B - 1 + B_\tau &= 0, \\ r + (\lambda - k\alpha)B - A_\tau &= 0. \end{aligned}$$

For simplicity first write $a_1 = \frac{1}{2}\sigma^2$ and $a_2 = k - \rho\eta\sigma$ which leads to $f(B) = B_\tau = -a_1B^2 - a_2B + 1$. Now we want to factorize the function $f(B)$.

$$\frac{f(B)}{-a_1} = B^2 + \frac{a_2}{a_1}B - \frac{1}{a_1} = (B+\epsilon)(B-\gamma).$$

From this we get that $(\epsilon-\gamma) = \frac{a_2}{a_1}$ and $\epsilon\gamma = \frac{1}{a_1}$. From these two relations it follows that we have to solve

$$a_1\gamma^2 + a_2\gamma - 1 = 0.$$

This gives

$$\begin{aligned} \gamma_{1,2} &= -\frac{a_2}{2a_1} \pm \frac{\sqrt{a_2^2+4a_1}}{2a_1} \\ &= \frac{-k_2 \pm k_1}{2a_1}, \end{aligned} \tag{3.8}$$

where k_1 and k_2 are given by (3.15). From this point it is assumable that both γ's should work. For convenience of the reader we check this below. We have

$$\begin{aligned} B_\tau = -a_1(B+\epsilon)(B-\gamma) &= 0 \\ \frac{1}{(B+\epsilon)(B-\gamma)}dB + a_1 d\tau &= 0 \\ \frac{1}{\gamma+\epsilon}\left[\frac{-1}{B+\epsilon} + \frac{1}{B-\gamma}\right]dB + a_1 d\tau &= 0 \\ \frac{1}{\gamma+\epsilon}\left[-\ln|B+\epsilon| + \ln|B-\gamma|\right] + \frac{1}{a_1}\tau &= c \\ \ln\left(\frac{|B-\gamma|}{|B+\epsilon|}\right) &= -\frac{\gamma+\epsilon}{a_1}\tau + (\gamma+\epsilon)c \\ B(\tau) &= \frac{(\tilde{c}\epsilon e^{-\frac{\gamma+\epsilon}{a_1}\tau} + \gamma)}{1 - \tilde{c}e^{-\frac{\gamma+\epsilon}{a_1}\tau}}, \end{aligned}$$

with initial condition $B(0) = 0$ we get that $\tilde{c} = -\frac{\gamma}{\epsilon}$. Substituting $\tilde{c}$ and calculating for γ_1 (which refers to the plus sign) gives

$$\begin{aligned} B(\tau) &= \frac{\gamma(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{1 + \frac{\gamma}{\epsilon}e^{-\frac{\gamma+\epsilon}{a_1}\tau}} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{\frac{\epsilon+\gamma e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{\epsilon\gamma}} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{k_2 + a_1(1 + e^{-\frac{\gamma+\epsilon}{a_1}\tau})\gamma} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{k_2 + a_1(\frac{-k_2+k_1}{2a_1}(1 + e^{-\frac{\gamma+\epsilon}{a_1}\tau}))} \\ &= \frac{2(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{2k_2 - k_2 + k_1 - k_2 e^{-\frac{\gamma+\epsilon}{a_1}\tau} + k_1 e^{-\frac{\gamma+\epsilon}{a_1}\tau}} \\ &= \frac{2(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{k_2 + k_1 + (k_1 - k_2)e^{-\frac{\gamma+\epsilon}{a_1}\tau}}. \end{aligned} \tag{3.9}$$

Now calculate

$$\begin{aligned} e^{-\frac{\gamma+\epsilon}{a_1}\tau} &= e^{-\frac{(\frac{-k_2+k_1}{2a_1}+\frac{k_2}{a_1}+\frac{-k_2+k_1}{2a_1})\tau}{a_1}} \\ &= e^{-k_1\tau}. \end{aligned} \tag{3.10}$$

Substituting (3.10) into (3.9) results in

$$B(\tau) = \frac{2(1 - e^{-k_1\tau})}{k_1 + k_2 + (k_1 - k_2)e^{-k_1\tau}} \tag{3.11}$$

Now substituting $\tilde{c}$ and calculating for γ_2 (which refers to the plus sign) gives

$$\begin{aligned} B(\tau) &= \frac{\gamma(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{1 + \frac{\gamma}{\epsilon}e^{-\frac{\gamma+\epsilon}{a_1}\tau}} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{\frac{\epsilon+\gamma e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{\epsilon\gamma}} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{k_2 + a_1(1 + e^{-\frac{\gamma+\epsilon}{a_1}\tau})\gamma} \\ &= \frac{1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau}}{k_2 + a_1(\frac{-k_2-k_1}{2a_1}(1 + e^{-\frac{\gamma+\epsilon}{a_1}\tau}))} \\ &= \frac{2(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{2k_2 - k_2 - k_1 - k_2 e^{-\frac{\gamma+\epsilon}{a_1}\tau} - k_1 e^{-\frac{\gamma+\epsilon}{a_1}\tau}} \\ &= \frac{2(1 - e^{-\frac{\gamma+\epsilon}{a_1}\tau})}{k_2 - k_1 + (-k_1 - k_2)e^{-\frac{\gamma+\epsilon}{a_1}\tau}}. \end{aligned}$$

Now calculate

$$\begin{aligned} e^{-\frac{\gamma+\epsilon}{a_1}\tau} &= e^{-\frac{(\frac{-k_2-k_1}{2a_1}+\frac{k_2}{a_1}+\frac{-k_2-k_1}{2a_1})\tau}{a_1}} \\ &= e^{k_1\tau}. \end{aligned} \tag{3.12}$$

Substituting (3.12) into (3.9) results in

$$\begin{aligned} B(\tau) &= \frac{2(1-e^{k_1\tau})}{-k_1+k_2+(-k_1-k_2)e^{-k_1\tau}}\frac{-e^{-k_1\tau}}{-e^{-k_1\tau}} \\ &= \frac{2(1-e^{-k_1\tau})}{k_1+k_2+(k_1-k_2)e^{-k_1\tau}}, \end{aligned} \tag{3.13}$$

where it is shown that both γ's satisfy. Accordingly

$$A(\tau) = r\tau + (\lambda - k\alpha)\int_t^T B_q dq, \tag{3.14}$$

where

$$\int_t^T B_q dq = \frac{2}{k_1(k_1+k_2)}\ln\left[\frac{(k_1+k_2)e^{k_1\tau}+k_1-k_2}{2k_1}\right] \tag{3.15}$$

$$+ \frac{2}{k_1(k_1-k_2)}\ln\left[\frac{k_1+k_2+(k_1-k_2)e^{-k_1\tau}}{2k_1}\right], \tag{3.16}$$

where

$$\begin{aligned} k_1 &= \sqrt{k_2^2+2\sigma^2}, \\ k_2 &= (k-\rho\eta\sigma). \end{aligned}$$

Chapter 4

Kalman Filter

4.1 Introduction

In 1960 R.E. Kalman published the Kalman Filter (KF). This algorithm makes optimal use of imprecise data on a (quasi-) linear system with Gaussian errors (white noise) to continuously update the best estimate of the systems current state. The power of KF is; to compute these updates it is only necessary to consider the estimates from the previous time step and the new measurement and not all the previous data. The main idea is that we want to estimate the current state and its uncertainty, but we can not directly observe these states. Instead we observe noisy measurements.

4.2 Kalman Algorithm

In general the Kalman Filter tries to estimate a state $x \in \Re^n$ where x_t is given by the stochastic differential equation

$$x_t = A_t x_{t-1} + B_t u_{t-1} + w_{t-1}, \tag{4.1}$$

with a measurement $z \in \Re^n$ given by

$$z_t = H_t x_t + v_t, \tag{4.2}$$

where

- A_t is a $n \times n$ matrix,
- B_t is a $n \times l$ matrix,
- H_t is a $m \times n$ matrix,
- w_t and v_t are the process and measurement noise (resp.) with mean zero and covariance matrices Q and R (resp.).

Now define

$$\mathbb{E}[x_t] = \hat{x}_t, \tag{4.3}$$
$$\mathbb{E}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)'] = P_t. \tag{4.4}$$

The KF algorithm is a predictor-corrector algorithm and uses the time updates for the prediction and the measurement updates as the corrector. The KF algorithm is given by

Predictor (Time updates)
$\hat{x}_t^- = A_t \hat{x}_{t-1} + B_t u_{t-1}$
$P_t^- = A_t P_{t-1} A_t' + Q_t$

Table 4.1:

Corrector (Measurement updates)
$K_t = P_k^- H_t' (H_t P_t^- H_t' + R_t)^{-1}$
$\hat{x}_t = \hat{x}_t^- + K_t (z_t - H_t \hat{x}_t^-)$
$P_t = (I - K_t H_t) P_t^-$

Table 4.2:

In Chapter 5 we discuss the Kalman filter for the CIR process. But first we introduce a simple example following [6].

4.3 Introductory example

In [6] they use the KF to estimate a random constant. We assume all matrices in Table 4.1 and 4.2 to be constant. By setting $A = 1$, $B = 0$ we obtain $\hat{x}_k^- = \hat{x}_{k-1}$, i.e., we skip the updating step. By setting $H = 1$, we get $z_k = x_k + v_k$, i.e., the measurements comes directly from the state x_k. We set $Q = 1E-05$. Now we have to choose an intial state to begin with. Since a random variable is normally distributed with mean zero, we take $x_0 = 0$ to be the intial state. Accordingly we must start with an intial state for P_k, P_0. It turns out that we can arbitraly choose $P_0 \neq 0$ and the filter will eventually converge. Take $P_0 = 1$, by taking P_0 large enough, the choice of x_0 does not influence the kalman filter [7]. The crosses in the following figures are generated by the matlab function

```
y=-0.37727+normrnd2(0,0.025^2,samples,1);
```

4.3.1 Simulation results

By randomly choosing $R = 0.0238$ we get,

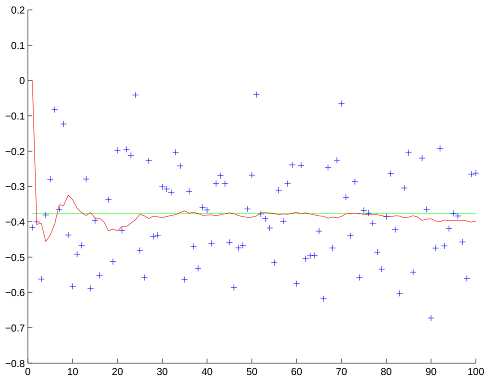

Figure 4.1: Simulation of a random constant, with $R = 0.0238$. The true value of $x = -0.37727$ is given by the green line, the Kalman Filter by the red line and the crosses are the noisy measurements.

Changing the choice of R is made clear by figure 4.2 and 4.3. In figure 4.2 we take $R = 1$, this will cause a much slower converging behaviour than in figure 4.1. This is because the filter responds slower to the noisy measurements. However the filter will eventually converge to the true value of x. In figure 4.3 we take $R = 0.0001$, as we can see the filter quickly responds to the measurements and tries to fit it. With the choice of this R we can expect that it will take very long for the filter to converge to the green line.

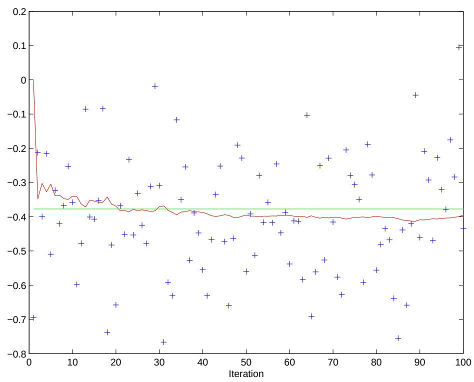

Figure 4.2: Simulation of a random constant, with $R = 1$. The true value of $x = -0.37727$ is given by the green line, the Kalman Filter by the red line and the crosses are the noisy measurements.

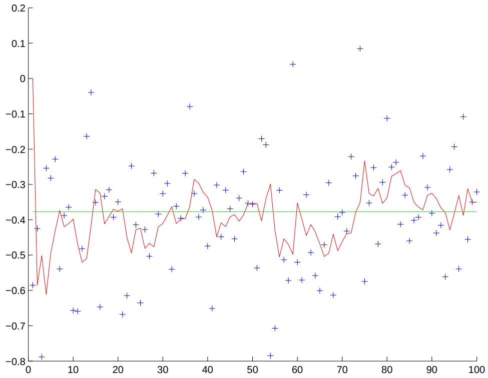

Figure 4.3: Simulation of a random constant, with $R = 0.0001$. The true value of $x = -0.37727$ is given by the green line, the Kalman Filter by the red line and the crosses are the noisy measurements.

4.3.2 How to choose R

Consider for simplicity $Q = 0$, so (6.2) can be written as $z_t = H_t x_t$ or $I_t = z_t - H_t x_t$, where I_t is usually refered to as the innovations. If we plot I_t in figure 4.4 we see that the crosses are distributed around zero. This is expected because R_{I_t} is statistically equal to the variance of the error term of v_t, R. We assumed v_t is $\mathcal{N}(0, R)$. But from the literature [7] it is also known that $R_{I_k} = H_t P_t^- H_t' + R$. We first choose R randomnly, say $R = 0.10$. If we plot both the statistically innovations and the theoretical ones we should have that 68 per cent of the crosses lie inside the theoretical boundary. In figure 4.4 we can see that only 54 per cent of the crosses lie inside the boundary.

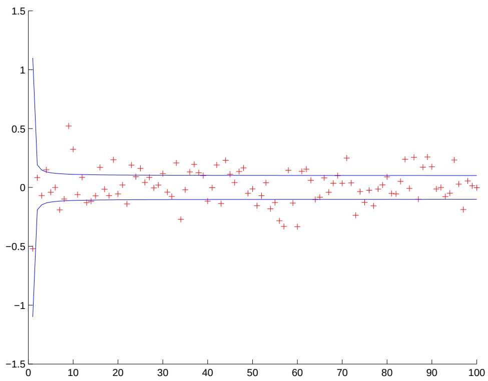

Figure 4.4: Comparison between the theoretical innovations (blue lines) and the statistical innovations (red crosses). The number of red crosses between the blue lines is 54 per cent. $R = 0.10$.

It is therefore necessary to change the value of R, say $R = 0.16$. We see that 71 per cent of the red crosses lie between the blue lines. If we now plot the Kalman Filter using this R we see that this is one of the best choices for R, what we expected.

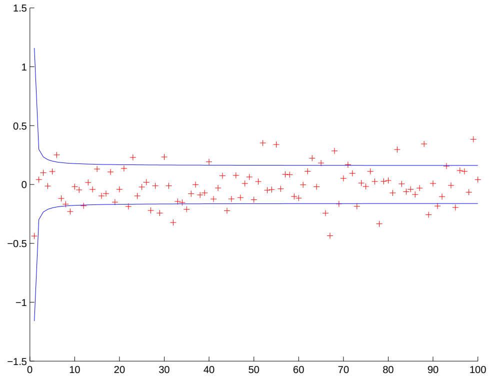

Figure 4.5: Comparison between the theoretical innovations (blue lines) and the statistical innovations (red crosses). The number of red crosses between the blue lines is 71 per cent. $R = 0.16$.

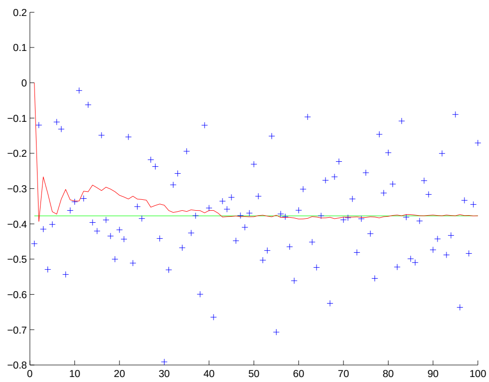

Figure 4.6: Simulation of a random constant, with $R = 0.16$. The true value of $x = -0.37727$ is given by the green line, the Kalman Filter by the red line and the crosses are the noisy measurements.

Chapter 5

The Kalman Filter for the Ornstein-Uhlenbeck process

To fit the state-space model, the state variables (which in this case are the spot prices and the convenience yield) are contained inside a state vector. The measurement equation, consisting of this vector and uncorrelated disturbances to account for possible errors in the data, links actual observations (in this case the future prices on several different maturities) with latent variables. These latent variables are assumed to be first-order Markovian processes and are related to systems (2.1) and (2.2). The Kalman filter will give an optimal prediction for the unobserved data by only considering the previously estimated value.
Consider again

$$F(S,\delta,\tau) = Se^{A(\tau)-B(\tau)\delta}, \tag{5.1}$$

where

$$A(\tau) = \left[\left(r-\tilde{\alpha}+\frac{\sigma^2}{2k^2}-\frac{\eta\sigma\rho}{k}\right)\tau\right] + \left[\frac{\sigma^2}{4}\frac{1-e^{-2k\tau}}{k^3}\right] + \left[\left(\tilde{\alpha}k+\sigma\eta\rho-\frac{\sigma^2}{k}\right)\left(\frac{1-e^{-k\tau}}{k^2}\right)\right],$$

and

$$B(\tau) = \frac{1-e^{-k\tau}}{k}, \quad \tilde{\alpha} = \alpha - (\frac{\lambda}{k}).$$

In state variable terms, (5.1) can be rewritten in

$$F_t(\tau) = \ln S_t + A(\tau) - B(\tau)\delta_t, \tag{5.2}$$

The measurement equation then reads

$$Y_t = d_t + Z_t[x_t,\delta_t]' + \epsilon_t, \quad t = 1,..,NOBS, \tag{5.3}$$

where[1]

- $Y_t = [\ln F_t(\tau_i)]$, for $i = 1,...,n$ is a $n \times 1$ vector for n maturities. τ_i are the maturity dates. $F_t(\tau)$ are observed from market data.
- $d_t = [A(\tau_i)]$ for $i = 1,..,n$ is a $n \times 1$ vector.
- $Z_t = [1, -B(\tau_i)]$, for $i = 1,...,n$ is a $n \times 2$ matrix.
- ϵ_t is a $n \times 1$ vector of uncorrelated disturbances and is assumed to be normal with zero mean and variance matrix H_t. H_t is a n $\times$ n diagonal matrix with h_i on its diagonal.

[1] NOBS is the number of observations.

The ϵ_t in (5.3) term is included to account for possible errors in the measurement. These errors especially occur when the state variables are unobservable. To get a feeling of the size of the error suppose that the OU model generates the prices and yields perfectly, and that the state variables can be observed from the market directly. The included error term in the measurement equation can be seen as the possibilities of bid-ask spreads, errors in the data etc. The error is assumed to be small in comparison to the variation of the yield. The matrix H_t is assumed to be diagonal for convenience in order to reduce the number of parameters to be estimated. The diagonal elements h_i are then estimated via the log-likelihood function.

In combination with the relationship $x_t = \ln S_t$ we have

$$\begin{cases} dx_t = (\mu - \delta_t - \frac{1}{2}\eta^2)dt + \eta dW_t, \\ d\delta_t = k(\alpha - \delta_t)dt + \sigma dZ_t, \end{cases} \tag{5.4}$$

From (5.4) the transition equation follows immediately,

$$\begin{bmatrix} x_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} (\mu - \frac{1}{2}\eta^2)\Delta t \\ k\alpha\Delta t \end{bmatrix} + \begin{bmatrix} 1 & -\Delta t \\ 0 & 1 - k\Delta t \end{bmatrix} \begin{bmatrix} x_{t-\Delta t} \\ \delta_{t-\Delta t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_t^1 \\ \xi_t^2 \end{bmatrix} \tag{5.5}$$

which for simplicity will be written as

$$\alpha_t = c_t + Q_t \alpha_{t-\Delta t} + R\xi_t. \tag{5.6}$$

where ξ_t is the transition error. Like ϵ_t, ξ_t is also assumed normal with mean zero and has a covariance-variance matrix given by

$$V_t = \begin{bmatrix} \sigma^2 \Delta t & \rho\sigma\eta\Delta t \\ \rho\sigma\eta\Delta t & \eta^2 \Delta t \end{bmatrix} \tag{5.7}$$

The matrices, H, Z, Q, c, d given above are parametrized by of the unknown parameter set $\varphi = \{k, \alpha\lambda, \eta, \sigma, \rho, \mu, h_i\}$. We dropped the time subscript for each matrix because they are time-independent. Note that V_t does not depend on the state variables $\{x_t, \delta_t\}$.

5.1 Iterative procedure

The iterative procedure consists of two main parts. Part one is the Kalman Filter which is used to update the systems matrices. Part two is estimating the parameter set $\varphi = \{k, \alpha\lambda, \eta, \sigma, \rho, \mu, h_i\}$. The algorithm of the Kalman Filter is given in pseudo code. Define $a_t = \mathbb{E}[\alpha_t]$ and $P_t = \mathbb{E}[(a_t - \alpha_t)(a_t - \alpha_t)']$. The tilde indicates predicted values.

while (number of iterations has not been reached , optimal φ has not been found) **do**
  (Kalman Filter)
  for $i = 1 : NOBS$ **do**
    (Prediction Equations)
    $\tilde{a}_{t_i} = Q * a_{t_{i-1}} + c;$
    $\tilde{P}_{t_i} = Q * P_{t_{i-1}} * Q' + R * V * R';$
    $\tilde{y}_{t_i} = Z * \tilde{a}_{t_i} + D;$
    (Innovations)
    $v_{t_i} = y_{t_i} - \tilde{y}_{t_i};$
    (Updating equations)
    $F_{t_i} = Z * \tilde{P}_{t_i} * Z' + H;$
    $a_{t_i} = \tilde{a}_{t_i} + \tilde{P}_{t_i} * Z' * F_{t_i}^{-1} * v_{t_i};$
    $P_{t_i} = \tilde{P}_{t_i} - \tilde{P}_{t_i} * Z' * F_{t_i}^{-1} * Z * \tilde{P}_{t_i};$
    $dF_{t_i} = \det F_{t_i};$
    if $dF_{t_i} \leqslant 0$ **then**

$dFt = 10^{-10}$;
end if
(Log-likelihood function per iteration)
$logl(i) = -(n/2) * \ln(2 * \pi) - 0.5 * \ln(dF_{t_i}) - 0.5 * v'_{t_i} * F_{t_i}^{-1} * v_{t_i}$;
end for
$LogL = \sum_i logl(i)$;
(Adjustment for φ via a Matlab optimization routine)
end while

Via an optimalisation routine, the vector φ is chosen such that the total sum of the log-likelihood function is maximized and the innovations minimized. After this is done, the matrices in the measurement and transition equation are updated and the Kalman Filter algorithm will then be repeated. This iterative procedure is repeated untill the optimized φ is found. With this optimized set of parameters the matrices will be updated once more and are used for the last time to generate the paths of the non-observable variables via the Kalman Filter. The total sum of the log-likelihood function is calculated as follows

$$\ln L(Y;\varphi) = -\frac{1}{2} n \ln 2\pi - \frac{1}{2}\sum_t \ln|F_t| - \frac{1}{2}\sum_t v'_t F_t^{-1} v_t, \tag{5.8}$$

where Y_t is the information vector at time t and it is assumed that Y_t conditional on Y_{t-dt} is normal with mean $\mathbb{E}[Y_t|Y_{t-dt}]$ and covariance matrix F_t.[2]

5.2 Numerical results for the Ornstein-Uhlenbeck process

We use weekly observation (on every friday to account for possible weekend effects) of light crude oil data from 01-02-2002 until 25-01-2008 (313 observations). At each observation seven monthly contracts (F1,F2,F3,F4,F5,F6,F7) ($n = 7$) are observed. Each diagonal element of the error matrix H corresponds to each maturity, i.e. h_1 corresponds to F1, h_2 corresponds to F2 and so on. The used marketdata is such that the time to maturities (τ_i) are constant for each forward contract i and for each observation data, i.e. for each observation date of contract i there exists a future price with maturity T_i.

5.2.1 Setting up the calibration

In order to start the iterative process, some difficulties must be overwon. Namely, the initial choice of the parameterset φ and the non-observable variables at time zero, i.e. x_0 and δ_0.
We start with $\varphi = \{k, \mu, \alpha, \lambda, \eta, \sigma, \rho, h_1, h_2, h_3, h_4, h_5, h_6, h_7\} = \Big\{0.3, 0.2, 0.06, 0.01, 0.4, 0.4, 0.8, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F1})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F2})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F3})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F4})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F5})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F6})]^2, \mathbb{V}\mathrm{ar}[(\ln \mathrm{F7})]^2\Big\}$. The first seven parameters are chosen arbitrarily. However, for the diagonal elements h_i of the error variance matrix H we choose the squares of the measurement variance. Therefore we expect the error to be small in comparison to the variance of the measurements. For the non-observable parameters (cf. [8]), the nearest future price is retained as the spot price S and

$$\delta_t^{\mathrm{implied}} = r - \frac{\ln(\mathrm{F}_t(\tau_1)) - \ln(\mathrm{F}_t(\tau_2))}{\tau_1 - \tau_2}. \tag{5.9}$$

with $\delta_0 = \delta_0^{\mathrm{implied}}$.

5.2.2 Results

The optimized parameter set is estimated for different initial sets and given in Table 5.1. To test the robustness of the calibration, we first choose randomly an initial parameter set. This gaves us the optimized set, which we then inserted as the initial one. Within a few interation steps the parameters converges to the same values.
Remark: In the following sections we use the second column of Table 5.1 to be the initial parameter set.

[2]This F_t is not the same as the future price

Parameters	Ini parset	Opti parset	Ini parset	Opti parset	Ini parset	Opti parset
k	0.3	1.4221 (0.0372)	0.3	1.4221 (0.0380)	2	1.4221 (0.0382)
μ	0.2	0.3733 (0.1471)	0.2	0.3733 (0.1376)	0.2	0.3733 (0.1382)
α	0.06	0.0699 (0.1128)	0.2	0.0699 (0.1025)	0.06	0.0699 (0.1082)
λ	0.01	-0.0183(0.1602)	0.1	-0.0183(0.1459)	0.01	-0.0183(0.1543)
η	0.4	0.3630 (0.0137)	0.4	0.3630 (0.0139)	0.5	0.3630 (0.0153)
σ	0.4	0.4028 (0.0165)	0.4	0.4028 (0.0172)	0.4	0.4028 (0.0181)
ρ	0.8	0.8378 (0.0162)	0.5	0.8378 (0.0164)	0.5	0.8378 (0.0177)
$\|h_1\|$	0.0246	0.0188 (0.0008)	0.0246	0.0188 (0.0007)	0.01	0.0188 (0.0007)
$\|h_2\|$	0.0268	0.0072 (0.0003)	0.0268	0.0072 (0.0003)	0.01	0.0072 (0.0003)
$\|h_3\|$	0.0291	0.0022 (0.0001)	0.0291	0.0022 (0.0001)	0.01	0.0022 (0.0001)
$\|h_4\|$	0.0313	0.0000 (0.0001)	0.0313	0.0000 (0.0001)	0.01	0.0000 (0.0001)
$\|h_5\|$	0.0336	0.0006 (0.0000)	0.0336	0.0006 (0.0000)	0.01	0.0006 (0.0000)
$\|h_6\|$	0.0357	0.0000 (0.0001)	0.0357	0.0000 (0.0001)	0.01	0.0000 (0.0001)
$\|h_7\|$	0.0377	0.0014 (0.0001)	0.0377	0.0014 (0.0001)	0.01	0.0014 (0.0001)
Log-Likelihood		8744.6479		8744.6479		8744.6479

Table 5.1: Optimized parameterset (Opti parset) for different initial parametersets (Ini parset). Standard errors in parentheses.

5.2.3 Results for different number of contracts

In this section we test the method for different number of contracts. We see that the standard error increases if only two contracts are used. This is expected because the Kalman Filter now has to estimate two unobservable values using by only two datasets.

Contracts	F1,F2,F3,F4,F5,F6,F7	F1,F3,F5,F7	F1,F7
k	1.4221 (0.0372)	1.4591 (0.0602)	1.7628 (0.4897)
μ	0.3733 (0.1471)	0.3654 (0.1354)	0.8134 (0.5488)
α	0.0699 (0.1128)	0.0653 (0.1004)	0.5169 (0.5381)
λ	-0.0183(0.1602)	-0.0331(0.1475)	1.8517 (2.2353)
η	0.3630 (0.0137)	0.3610 (0.0139)	0.3501 (0.0153)
σ	0.4028 (0.0165)	0.3995 (0.0183)	0.4108 (0.0705)
ρ	0.8378 (0.0162)	0.8418 (0.0174)	0.8172 (0.0234)
$\|h_1\|$	0.0188 (0.0008)	0.0161 (0.0007)	0.0188 (0.0012)
$\|h_2\|$	0.0072 (0.0003)	0.0000 (0.0003)	0.0081 (0.0031)
$\|h_3\|$	0.0022 (0.0001)	0.0021 (0.0001)	
$\|h_4\|$	0.0000 (0.0001)	0.0010 (0.0012)	
$\|h_5\|$	0.0006 (0.0000)		
$\|h_6\|$	0.0000 (0.0001)		
$\|h_7\|$	0.0014 (0.0001)		
Log-Likelihood	8744.6479	4001.8997	1440.5857

Table 5.2: Optimized parameterset (Opti parset) for different initial parametersets (Ini parset). Standard errors in parentheses.

5.2.4 Estimated future prices versus observered future prices

We plot the state variable x_t (the log of the spot price) versus the one-month maturity of the log of the future prices, i.e. $\ln(\mathrm{F}_t(\tau_i))$. The two curves are following each other quite good but are not identical to each other.

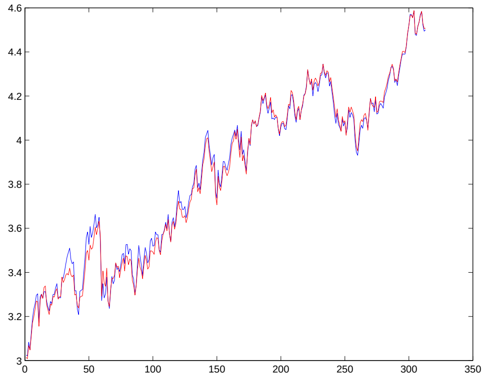

Figure 5.1: Blue line: State variable x_t, Red line:$\ln(\mathrm{F}_t(\tau_1))$.

5.2.5 Filtered convenience yield versus implied convenience yield

Nowadays traders imply the convenience yield via the future prices like in (5.9). The estimated convenience yield is plotted versus the implied convenience yield.

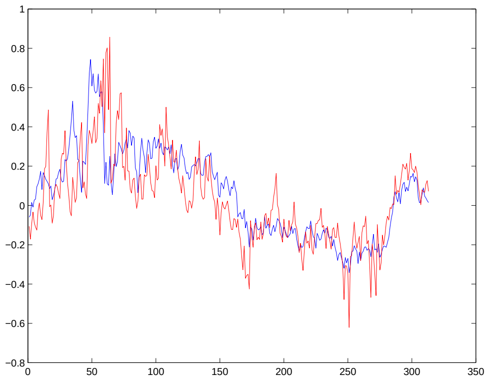

Figure 5.2: Blue line:Filtered δ_t, Red line:$\delta_t^{\text{implied}}$.

5.2.6 Innovations v_t

For each future contract we plot the innovation, i.e. $v_{t_i} = y_{t_i} - \tilde{y}_{t_i}$. We assumed the measurement errors, ϵ_t in equation (5.3) to be normal with mean zero and variance H. Considering the figures, this assumption is acceptable. Note that there are some transcend points. This could be caused by data errors.

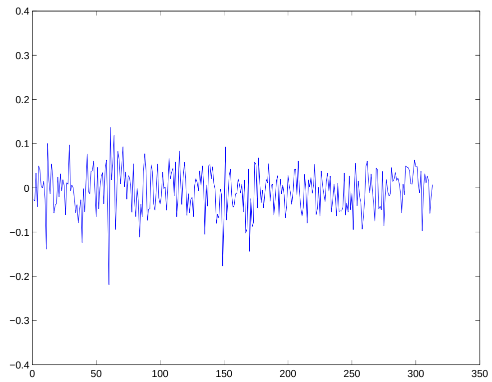

Figure 5.3: Innovation corresponding to F1. Mean = -8.1734e-004, Variance = 0.0022.

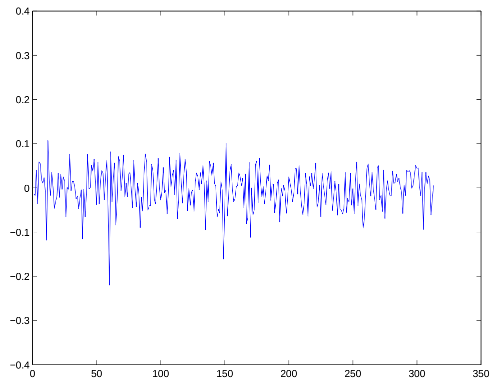

Figure 5.4: Innovation corresponding to F2. Mean = 8.6078e-004, Variance = 0.0019.

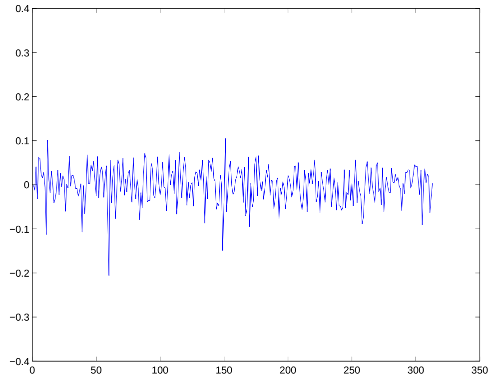

Figure 5.5: Innovation corresponding to F3. Mean = 6.4612e-004, Variance = 0.0017.

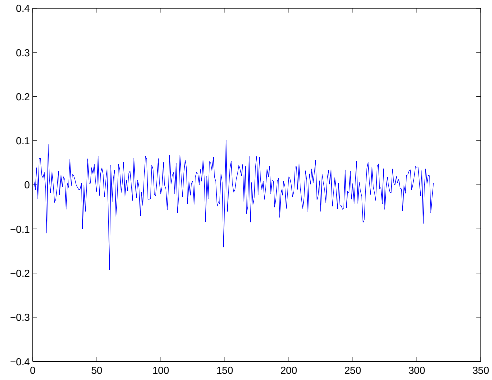

Figure 5.6: Innovation corresponding to F4. Mean = -1.0181e-004, Variance = 0.0015.

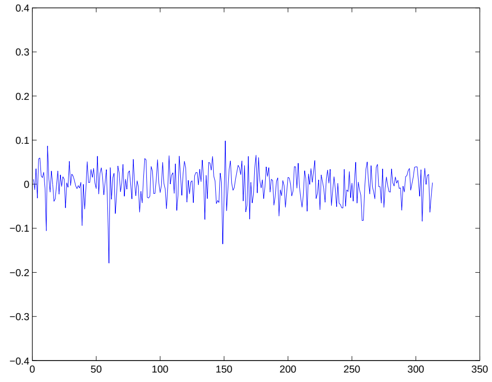

Figure 5.7: Innovation corresponding to F5. Mean = -5.5711e-004, Variance = 0.0014.

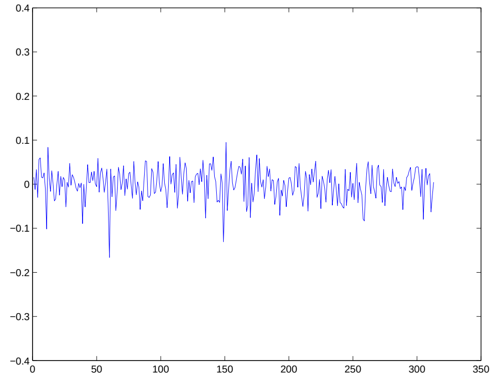

Figure 5.8: Innovation corresponding to F6. Mean = -3.4456e-004, Variance = 0.0013.

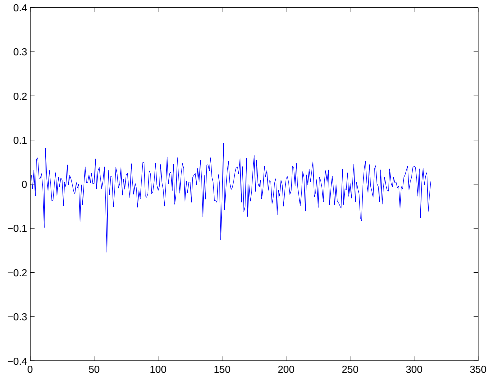

Figure 5.9: Innovation corresponding to F7. Mean = 6.1501e-004, Variance = 0.0012.

5.2.7 Parameter explanation

If we look at Table 5.1, we see an mean-reverting term, α of 0.0699. In figure 5.2 we can see that the convenience yield indeed reverts to this mean. The correlation between the two state variables is 0.8378. To find graphical evidence for this high number we made a scatterplot of the increments of both the state-variables, i.e. $\Delta x_t := x_t - x_{t-\Delta t}$ versus $\Delta \delta_t := \delta_t - \delta_{t-\Delta t}$.

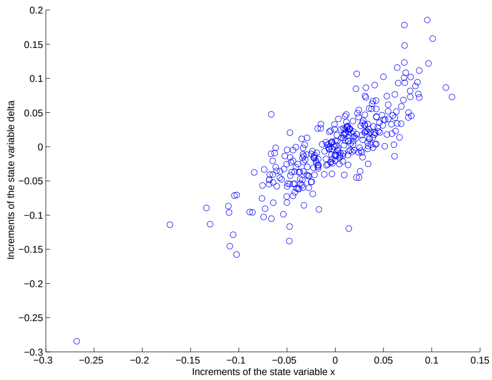

Figure 5.10: Scatterplot of $(\Delta x_t, \Delta \delta_t)$ for 313 observations.

Since the pattern of dots slopes from lower left to upper right, it suggests a positive correlation between the variables being studied, which is exactly what we expected.

5.2.8 Implementation issues

There are a few important implementation details worth mentioning. The log-likelihood function estimates the optimal parameterset. For that, it has control of a parameter set bound. We use the following bounds

```
          %[  mu sigmaS alpha   k sigmaC rho lambda  h1 h2 h3 h4 h5 h6 h7]:
lowerbound=[ -10  0.001   -10   0.001  0.001   0    -10  -1 -1 -1 -1 -1 -1 -1];
upperbound=[  10     10    10     10     10   1     10   1  1  1  1  1  1  1];
```

If this is not done, the iterative procedure may break down. In case of matrix H, consisting of the variances of the error terms, can become negative. But being negative, this can cause problems in the KF. In (??) F_t could converge to a negative diagonal matrix, becoming impossible to inverse. So we square each element of H before calculating F_t. If the parameters h_i come out negative, it is ok, because they effectively enter F_t as the variance, i.e. they will be squared. This is why h_i in Table 5.1 and in Table 5.2 are in absolute value.

5.2.9 Kalman Forecasting

We test a Kalman forecasting method on half of the observations. The first half, from observation i=1 until i=156 we perform the KF algorithm with the optimized parameter set given in the third column of Table 5.1. From i=157 until i=313 we forecast the state variables, x_t and δ_t by the following recursive equations

$$\begin{aligned} a_t &= Q_t a_t + c_t, \\ P_t &= Q_t P_t Q_t' + R_t V_t R_t', \\ y_t &= Z_t a_t + D_t. \end{aligned} \tag{5.10}$$

Forecasting for the state variable δ_t is not a succes. This can be explained by the sudden drop at 160.

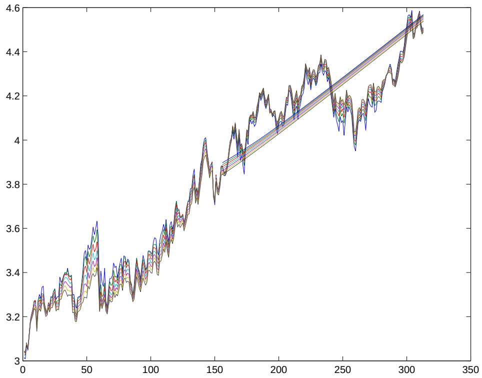

Figure 5.11: Forecast over half a sample of the log of the future prices (F1-F7).

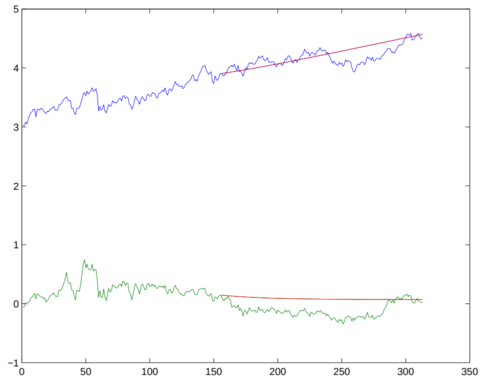

Figure 5.12: Forecast over half a sample of the state variables. Blue line: x_t, green line: δ_t, red lines: forecasting

The Kalman forecasting can not absorb this sudden drop in the convenience yield. For that, it follows the previous estimations. However, as noted earlier, the mean-reverting term is equal to 0.0699, which is approximately equal to the forecasting. Also for x_t, the sudden drop at 250 can not be absorbed by this forecasting method. This is a real disadvantage of this method. A way to get the method to absorb these drops, jumps can be inserted in the stochastic processes of both the state variables.

Chapter 6

The Kalman Filter model for the Cox-Ingersoll-Ross process.

In state variable terms, (3.7) can be written in

$$F_t(\tau) = \ln S_t + A(\tau) - B(\tau)\delta_t, \tag{6.1}$$

where $A(\tau)$ and $B(\tau)$ are given in (3.14) and (3.13). The measurement equation then reads

$$Y_t = d_t + Z_t[x_t, \delta_t]' + \epsilon_t, \quad t = 1, .., NOBS, \tag{6.2}$$

where

- $Y_t = [\ln F(\tau_i)]$, for $i = 1, ..., n$ is a $n \times 1$ vector for n maturities. τ_i are the maturity dates.
- $d_t = [A(\tau_i)]$ for $i = 1, .., n$ is a $n \times 1$ vector.
- $Z_t = [1, -B(t_i)]$, for $i = 1, ..., n$ is a $n \times 2$ matrix.
- ϵ_t is a $n \times 1$ vector of uncorrelated disturbances and is assumed to be normal with zero mean and variance matrix H_t.

Another motivation for including the error term is that perhaps the dynamics of the stochastic convenience yield does not include a square-root. If so, the yields implied by the CIR model will differ from the observed yields. Furthermore, since the CIR model forbids negativity of the convenience yield there are two solutions; firstly we can replace any negative element of the estimate δ_t with zero. Secondly, we could just skip the updating step, i.e. we set $\delta_t = \delta_{t-1}$ whenever δ_t is negative. The reason why this is done is just a matter of convenience and it seems rather difficult to impose the positivity constraint in the estimation procedure and thus a solution like these is needed.
Again, the vector $[x_t, \delta_t]'$ is the data which we want to estimate. From (3.1) and (3.5) the transition equation immediately follows

$$\begin{bmatrix} x_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} \mu\Delta t \\ k\alpha\Delta t \end{bmatrix} + \begin{bmatrix} 1 & -(1+\frac{1}{2}\eta^2)\Delta t \\ 0 & 1-k\Delta t \end{bmatrix} \begin{bmatrix} x_{t-\Delta t} \\ \delta_{t-\Delta t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_t^1 \\ \xi_t^2 \end{bmatrix}.$$

For notational simplicity we write

$$\alpha_t = c_t + Q_t\alpha_{t-1} + R_t\xi_t, \quad t = 1, ..NOBS, \tag{6.3}$$

where $\mathbb{E}[\xi_t^i] = 0$ for $i = 1, 2$ and where the covariance matrix of ξ_t^i, V_t, is given by

$$V_t = \begin{bmatrix} \mathbb{E}[(x_t - \mathbb{E}[x_t])^2 | x_{t-\Delta t}] & \mathbb{E}[(x_t - \mathbb{E}[x_t])(\delta_t - \mathbb{E}[\delta_t]) | x_{t-\Delta t}, \delta_{t-\Delta t}] \\ \mathbb{E}[(x_t - \mathbb{E}[x_t])(\delta_t - \mathbb{E}[\delta_t]) | x_{t-\Delta t}, \delta_{t-\Delta t}] & \mathbb{E}[(\delta_t - \mathbb{E}[\delta_t])^2 | \delta_{t-\Delta t}] \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{V}\text{ar}[x_t | x_{t-\Delta t}] & \sqrt{\mathbb{V}\text{ar}[x_t | x_{t-\Delta t}]}\sqrt{\mathbb{V}\text{ar}[\delta_t | \delta_{t-\Delta t}]} \\ \sqrt{\mathbb{V}\text{ar}[x_t | x_{t-\Delta t}]}\sqrt{\mathbb{V}\text{ar}[\delta_t | \delta_{t-\Delta t}]} & \mathbb{V}\text{ar}[\delta_t | \delta_{t-\Delta t}] \end{bmatrix}$$

$$= \begin{bmatrix} \eta^2 \Delta t \delta_{t-\Delta t} & \rho\eta\sqrt{\Delta t}\sqrt{\delta_{t-\Delta t}}\sqrt{\mathbb{V}\mathrm{ar}[\delta_t|\delta_{t-\Delta t}]} \\ \rho\eta\sqrt{\Delta t}\sqrt{\delta_{t-\Delta t}}\sqrt{\mathbb{V}\mathrm{ar}[\delta_t|\delta_{t-\Delta t}]} & \mathbb{V}\mathrm{ar}[\delta_t|\delta_{t-\Delta t}] \end{bmatrix}.$$

The difference between V_t for the OU process and the CIR process is the time dependency. Implementing the Kalman Filter for the CIR process includes feedback from the kalman filter to update V_t with $\delta_{t|t-1}$. As mentioned earlier, since the square root is taking of the convenience yield, we replace any negative element of the estimate $\delta_{t|t-1}$ with zero.

General remark: it would be better to use $\delta_t + \gamma$, γ a constant parameter.

Chapter 7

Conclusion

We implemented the KF for the OU process. Both the CY as well as the state-variable x (log of the spot prices) seems to follow the implied yield and the market prices (resp.) quite good. Also, different initial values for the parameterset will eventually converge to the optimized set with the same value of the log-likelihood. This is a good result and tests the robustness of the method.
The main difference bewteen the systmes matrices of both processes is the transition error covariance-variance matrix V_t. For the CIR process, we simply replaced any negative element of the CY by zero, but since it is negative for a large number of observations, this will probably give rise to large standard errors in the optimized parameterset φ.

7.1 Further research

Finish the implementation of the Kalman Filter for the CIR process and compare the results with the OU process. If the CY in the CIR gives problems because it can not be negative we could insert a jump-constant in the CY. Perhaps it is possible to extend the two factor stochastic process by a third factor, e.g. stochastic volatility. The problem with this is that the solution of the future prices is assumed to have an affine form solution. This makes the Kalman Filter considerable easier. Including stochastic volatility could mess this up. Another idea is to price options on commodities using the optimized parameterset.

Bibliography

[1] F.H.C. Naber, *Fast solver for the three-factor Heston-Hull-White problem*, 2006-2007, Wholesale Banking Amsterdam ING.

[2] A. Krul, L^∞ *Error estimates for the solution of the Black-Scholes and Hull-White models on truncated domains with a payoff boundary condition*, 2007, Wholesale Banking Amsterdam ING.

[3] P. Bjerksund, *Contingent Claims Evaluation when the Convenience Yield is stochastic: Analytical Results*, 1991.

[4] D.R. Ribeiro and S. D. Hodges, *A Two-Factor Model for Commodity Prices and Futures Valuation*, 2004.

[5] R. Gibson and E. S. Schwartz, *Stochastic Convenience Yield and the Pricing of Oil Contingent Claims*, The Journal of Finance, Volume 45, Issue 3(Jul.1990), 959-976.

[6] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, July 24, 2006.

[7] A. Heemink , *http://ta.twi.tudelft.nl/wagm/users/heemink/ch5dataas.doc*, Delft University of Technology.

[8] D. Lautier, *The Kalman Filter In Finance: An Application to Term Structure Models of Commodity Prices and a Comparison between the Simple and the Extended Filters*, 06/12/02, working paper.

[9] D. Lautier, *The Informational Value of Crude Oil Futures Prices*, June 2003.