A Parametric Approach To Bayesian Drift Estimation of Discretely Observed Price Diffusion Processes[1]

Alexandry AUGUSTIN [2]

September 19, 2011

Department of Mathematics
Imperial College of Science, Technology and Medicine
London SW7 2BZ

Project report submitted in partial fulfilment of the degree of MSc in Mathematics and Finance at Imperial College London.

Abstract

The present report suggests an approach to parametric estimation of the drift function $\mu(t)$ in a one-dimensional Itô stochastic differential equation representing a stock price with a variable diffusion parameter σ such that $dS_t = \mu(t)S_t dt + \sigma S_t dW_t$ from a portfolio consisting of the stock, a call option and a forward contract. The parametric charactrisation of $\mu(t)$ will be derived from non-arbitrage arguments used to price derivatives - Black-Scholes (1970). The stock price process will be expressed in terms of discretely observed derivative prices and the unobservable discrete *market risk premium*, the later modeled using the Ornstein-Uhlenbeck process. To address the time variation of volatility σ, the *Black-Scholes implied volatility* of the option will be used. The resulting system of prices and market risk premium dynamics will then be cast into the *state space* framework from which the time varying conditional distribution of the market risk premium can be estimated using the *Kalman filter* - Kalman (1960). Numerical analysis will be carried out using simulated market prices.

Keywords: Itô Diffusion, Market price of risk, Risk-neutral valuation, Euler-Maruyama scheme, Linear recursive Baysian estimation, Kalman filter, Stochastic filtering.

[1]I would like to cordially thank Dr. Chris Barnett for his advice and help along the project.
[2]alexandry.augustin@gmail.com

Contents

1 Introduction

The recovery of information from measurements corrupted by uncertainty has long been a struggle endured by many investors. Typical to financial asset prices, the volatility is usually very large compare to the drift resulting in difficult accurate estimation.

Stochastic processes are a natural choice to model the time evolution of financial assets subject to random influences. Historically, Louis Bachelier is credited with being the first to model stock prices with stochastic processes (using a Brownian motion), which was part of his PhD thesis *The Theory of Speculation* (1900). However, price in his model could assume negative values and therefore couldn't reflect the reality of the market. Subsequently, Black-Scholes developed in *The Pricing of Options and Corporate Liabilities* (1973) where the stock price follows a geometric brownian motion with drift (see appendix A). While derivative valuation models presuppose knowledge of the parameters that underlie diffusion processes, in practice these parameters are unknown and must be estimated (commonly referred as *model calibration*).

The development of *stochastic filtering* theory in the early 1940s pioneered by Wiener (1942) and Kolmogorov (1941) provides us with the necessary tools to make reasonable inferences about the unknown diffusion parameters. In particular, *Bayesian filtering* is aimed to apply Bayes statistics to stochastic filtering problems. It culminated with the publication of *A new approach to linear fitering and prediction problem* by Kalman (1960) where the classic *Kalman filter* was introduced [3].

The first full implementation of the Kalman filter was in 1960 at the *Ames Research Center* (NASA) where it was used for the trajectory estimation and control problem of the Apollo project. In particular, the Apollo 11 lunar module that landed to the moon was guided by a Kalman filter. The filter performed inference on the module position based on multiple embeded sensors which was then compared to earth-based estimates. If they had disagreed too much the mission would have had to be aborted - see Mohinder and Angus (2001) for historical accounts and appendix E for an example from classical mecanics. Nowadays, Kalman filters have been applied in various scientific areas, including communications, machine learning, finance, economics, etc. It has been said that the Kalman filter is the greater discoveries in the history of statistical estimation theory.

The point of view taken here is that scientific inferences made about financial phenomena are not fundamentally different from inference made about phenomena in other areas of science (namely *engineering*). Thinking aside leads to a richer variety of ideas from which combinations can be formed.

[3] Bucy (1961) derived the continuous time counterpart independently in what is now called the *Kalman-Bucy filter*.

"The unity of all science consists alone in its method, not in its material."
- Karl Pearson (1892), The Grammar of Science.

Although many methods currently exist to estimate diffusion processes, the method proposed in this report is applicable to a large class of discretely observed diffusion processes, including those for which observations are irregularly spaced, those in which the data are observed with error due to factors such as market microstructure.

The report is organized as follows. Chapter 1 provide the fundation for the parametric system based on the risk-neutral valuation framework. Chapter 2 is a self contained derivation of the Kalman filtering equations from a Bayesian filtering standpoint. Chapter 3 addresses the estimation problem of the diffusion parameter μ_k identified in Chapter 1. In each Chapters particular areas have been addressed where the views clearly advanced the general objective. Otherwise, detailed proofs and further more-detailes have been relegated to the appendix section so as not to interupt the flow of reading. The select bibliography is intended mainly to give references to specific secondary literature cited in the text, although other influential works have been included as well.

2 Market Price of Risk in the Risk-Neutral Valuation Framework

In this section will be described the risk-neutral relationship between the prices of the traded derivatives and their underlying assets.

Let $(\Omega, \mathscr{F}_t, P)$ be a complete filtered probability space where the filtration $\mathscr{F}_t$ satisfies the *usual conditions*:

(i) non-decreasing: $\mathscr{F}_s \subseteq \mathscr{F}_t \subset \mathscr{F}$ for $s \leq t$,
(ii) right continuity: $\mathscr{F}_t = \mathscr{F}_{t^+} = \cap_{u>t}\mathscr{F}_u$,
(iii) completness: $\mathscr{F}_0$ contains all P-null sets of $\mathscr{F}$.

Such a complete filtered probability space satisfying the usual conditions is called a *stochastic basis*.

On this stochastic basis, the filtration $\mathscr{F}_t$ is the sigma-field generated by a standard one-dimensional Brownian motion $(W_t)_{t>0}$ such that

$$\mathscr{F}_t = \sigma\{W_t\}.$$

We model the stock index price as a one-dimensional time-inhomogeneous Itô diffusion process

$$dS_t = \mu(t)S_t dt + \sigma S_t dW_t. \tag{1}$$

We denote $\mu(t)$ the unknown time–dependent *drift function* and σ^2 the *variance*.

We recall that the dynamic of a riskless bond is given by

$$dB_t = B_t r(t) dt. \tag{2}$$

We assume the existance of a *forward contract* on $(S_t)_{t>0}$ such that

$$dF_t = \mu_f(t)F_t dt + \sigma_f F_t dW_t. \tag{3}$$

Suppose that there is a *call option* on $(F_t)_{t>0}$ and its price process is $(C_t)_{t>0}$. We assume $(F_t)_{t>0}$ is adapted to the filtration $\mathscr{F}_t$, and in particular that $(C_t)_{t>0}$ is fully characterised by an $\mathscr{F}_T$-measurable terminal value $C_T = (F_t - K)^+$. We can write under P,

$$dC_t = \mu_c(t)C_t dt + \sigma_c C_t dW_t \tag{4}$$

for the price dynamics of the option.

In summary, we are left with the following price dynamics to estimate $\mu(t)$

$$\boxed{\left\{\begin{array}{rcrcr} dS_t & = & \mu(t)S_t dt & + & \sigma S_t dW_t \\ dF_t & = & \mu_f(t)F_t dt & + & \sigma_f F_t dW_t \\ dC_t & = & \mu_c(t)C_t dt & + & \sigma_c C_t dW_t. \end{array}\right.} \tag{5}$$

Before proceeding further, we give the following definitions.

2.1 Preliminary Definitions

Definition 2.1. A *portfolio process* $(V_t)_{t>0}$ in the market $\mathscr{M} = (S_t, F_t, C_t, B_t)$ is a stochastic linear combination of S_t, F_t, C_t, B_t such that

$$V_t = \phi_t S_t + \varphi_t F_t + \psi_t C_t + \xi_t B_t \tag{6}$$

where ϕ_t, φ_t, ψ_t, ξ_t are progressively measurable adapted processes indicating the number of respective assets held at time t. We suppose that the assets are infinitely divisible and can be borrowed ("shorted") so that it is possible to hold any positive or negative quantities of either assets.

Definition 2.2. The portfolio process $(V_t)_{t>0}$ is *self-financing* if

$$dV_t = \phi_t dS_t + \varphi_t dF_t + \psi_t dC_t + q\phi_t S_t dt + \xi_t dB_t. \tag{7}$$

That is, eventual incomes (selling assets, receiving dividends, ...) are exactly offset by required payments (buying additional assets, transaction costs, ...). And therefore the change in the portfolio value is only due to the trading profits and losses.

The no-arbitrage argument is crucial for the existence of a risk-neutral measure. In fact, the condition of no-arbitrage is equivalent to the existence of a risk-neutral measure[4].

Definition 2.3. An *arbitrage opportunity* is a self-financing portfolio $(V_t)_{t>0}$ with the property that

$$V_0 = 0,$$
$$P(V_T \geq 0) = 1,$$
$$P(V_T > 0) > 0.$$

In other word, starting with no initial capital a $t = 0$, the portfolio process $(V_t)_{t>0}$ has a positive probability of gaining value at time $t = T$ with no exposure to risk.

To avoid arbitrage opportunities, possible strategies need to be restricted to a subset whose wealth is bounded uniformly from below by a constant[5].

Definition 2.4. A strategy $(\phi_t, \varphi, \psi, \xi)$ is *admissible* if for an integrable, nonnegative and $\mathscr{F}_T$-measurable random variable X,

$$\frac{V_t}{B_t} \geq -E^P[X|\mathscr{F}_t]. \tag{8}$$

[4] See *fundamental theorem of asset pricing.*

[5] See Biagini and Cerny (2011) for a complete presentation.

2.2 The Risk-Neutral Framework

In this section we will derive an expression for all the parameters of the diffusion processes (5) under the risk-neutral assumption.

Proposition 2.5. *The conditions for the portfolio consisting of the stock index S_t paying dividends at a constant rate q, the forward on the stock index F_t and the call option on the forward C_t to be risk-neutral are given by*

$$\mu(t) = (r - q) + \lambda(t)\sigma \tag{9a}$$

$$\mu_f(t) = \lambda(t)\sigma_f \tag{9b}$$

$$\mu_c(t) = r + \lambda(t)\sigma_c \tag{9c}$$

where $\lambda(t)$ is known as the time-dependent markert price of risk.

Proof. Let the portfolio $V_t = \phi_t S_t + \varphi_t F_t + \psi_t C_t + \xi_t B_t$ consisting of a stock S_t and a forward option on the stock F_t, a call option on the forward C_t and a riskless bond B_t.

It should be stressed at this point that dividend paying underlyings have divergent treatments in the literature. We will asume here that the accumulated value of the dividend stream $q\phi_t S_t dt$ is reinvested into the bond as to ensure that the portfolio is self-financing. We can write its dynamic as

$$dV_t = \phi_t dS_t + \varphi_t dF_t + \psi_t dC_t + q\phi_t S_t dt + \xi_t dB_t.$$

Substituting (5) from this expression, we get

$$\begin{aligned} dV_t = &\left[\phi_t \mu(t) S_t + \varphi_t \mu_f(t) F_t + \psi_t \mu_c(t) C_t + q\phi_t S_t + \xi_t r(t) B_t\right] dt \\ &- \left[\phi_t \sigma S_t + \varphi_t \sigma_f F_t + \psi_t \sigma_c C_t\right] dW_t. \end{aligned}$$

The portfolio is rendered riskless if

$$\left[\phi_t \sigma S_t + \varphi_t \sigma_f F_t + \psi_t \sigma_c C_t\right] = 0. \tag{10}$$

The riskless dynamic of V_t is then given by

$$\frac{dV_t}{V_t} = \left(\frac{\phi_t \mu(t) S_t + \varphi_t \mu_f(t) F_t + \psi_t \mu_c(t) C_t + q\phi_t S_t + \xi_t r B_t}{\phi_t S_t + \varphi_t F_t + \psi_t C_t + \xi_t B_t} \right) dt.$$

A riskless portfolio must earn the risk-free rate r, therefore

$$\frac{\phi_t \mu(t) S_t + \varphi_t \mu_f(t) F_t + \psi_t \mu_c(t) C_t + q\phi_t S_t + \xi_t r(t) B_t}{\phi_t S_t + \varphi_t F_t + \psi_t C_t + \xi_t B_t} = r \tag{11}$$

giving

$$(\mu(t) + q - r)\phi_t S_t + (\mu_f(t) - r)\varphi_t F_t + (\mu_c(t) - r)\psi_t C_t = 0.$$

Defining

$$\begin{cases} X & = \phi_t S_t \neq 0 \\ Y & = \varphi_t F_t \neq 0 \\ Z & = \psi_t C_t \neq 0 \end{cases}$$

we can write the conditions (10) and (11) as

$$\sigma X + \quad \sigma_f Y + \quad \sigma_c Z = 0 \tag{12a}$$

$$(\mu(t) + q - r)X + \quad (\mu_f(t) - rY + \quad (\mu_c(t) - r)Z = 0. \tag{12b}$$

Eliminating X by subtracting $\sigma \times$ (12b) from $(\mu(t) + q - r) \times$ (12a) we get

$$\begin{aligned} &[\sigma_f(\mu(t) + q - r) - \sigma(\mu_f(t) - r)]\, Y \\ &+ [\sigma_c(\mu(t) + q - r) - \sigma(\mu_c(t) - r)]\, Z = 0. \end{aligned}$$

From the first term we get

$$\frac{\mu(t) + q - r}{\sigma} = \frac{\mu_f(t) - r}{\sigma_f}$$

and from the second term we get

$$\frac{\mu(t) + q - r}{\sigma} = \frac{\mu_c(t) - r}{\sigma_c}.$$

Therefore

$$\frac{\mu(t) + q - r}{\sigma} = \frac{\mu_f(t) - r}{\sigma_f} = \frac{\mu_c(t) - r}{\sigma_c} = \lambda(t).$$

We observe here that our result differs from the one obtained by Bhar, Chiarella and Runggaldier (2001) i.e.

$$\frac{\mu(t) + q - r}{\sigma} = \frac{(\mu_f(t) + r) - r}{\sigma_f} = \frac{\mu_c(t) - r}{\sigma_c} = \lambda(t).$$

It seems that the traitement of dividends in the modeling of the assets is the source of the discrepancy. Further investigations needs to be done in order to complete accurately the proof (the approach taken by Musiela and Rutkowski in *Martingale methods in financial modelling* seems to be of particular interest).

Furthermore, the bond included in our portfolio (where the dividends are being re-invested) doesn't seems to be present in Bhar, Chiarella and Runggaldier (2001) derivations. We might question the importance of its inclusion in our portfolio. □

Proposition 2.6. *The volatility of the Forward σ_f is the same as it's underlying (the stock index price) σ,*

$$\boxed{\sigma_f = \sigma} \tag{13}$$

Proof. The discounted conditional expectation of the forward payoff under the risk-neutral measure is given by

$$\begin{aligned} f_t &= E_t^{\mathbb{Q}}[(S_T - K)e^{-r(T-t)}] \\ &= e^{-r(T-t)}(E_t^{\mathbb{Q}}[S_T] - K). \end{aligned}$$

The forward price process F_t is the value of K such that $f_t = 0$. Therefore,

$$\begin{aligned} F_t &= E_t^{\mathbb{Q}}[S_T] \\ &= E_t^{\mathbb{Q}}[S_t e^{(r-q-\frac{\sigma^2}{2})(T-t)+\sigma(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}] \\ &= S_t e^{(r-q-\frac{\sigma^2}{2})(T-t)} E_t^{\mathbb{Q}}[e^{\sigma(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}]. \end{aligned}$$

Identifying the expectation as the moment generating function[6] of $W_T^{\mathbb{Q}} - W_t^{\mathbb{Q}} \sim N(0, T-t)$ we get

$$\begin{aligned} F_t &= S_t e^{(r-q-\frac{\sigma^2}{2})(T-t)} e^{\frac{\sigma^2}{2}(T-t)} \\ &= S_t e^{(r-q)(T-t)}. \end{aligned} \tag{14}$$

Applying the *Itô 's product rule* gives

$$dF_t = S_t de^{(r-q)(T-t)} + e^{(r-q)(T-t)} dS_t + \left\langle S_t, e^{(r-q)(T-t)} \right\rangle$$

with

$$\frac{de^{(r-q)(T-t)}}{dt} = -(r-q)e^{(r-q)(T-t)}$$

and

$$\left\langle S_t, e^{(r-q)(T-t)} \right\rangle = 0$$

as $e^{(r-q)(T-t)}$ is a continuous finite variation processes. Therefore

$$dF_t = -(r-q)S_t e^{(r-q)(T-t)} dt + e^{(r-q)(T-t)}(\mu(t)S_t dt + \sigma S_t dW_t)$$

and using (14) we deduce that

$$dF_t = [\mu(t) - (r-q)] F_t dt + \sigma F_t dW_t.$$

Consequently, by identification with (5) we get

$$\mu_f(t) = \mu(t) - (r-q)$$

and using (9a) complete the proof. □

[6]We recall that the moment generating function of a normally distributed random variable $X \sim N(\mu, \sigma)$ is given by $E[e^{tX}] = e^{t\mu + \frac{1}{2}t^2\sigma^2}$.

Proposition 2.7. *The volatility of the option σ_c is given by*

$$\sigma_c = \sigma \frac{F_t}{C_t} e^{-r(T-t)} N(d_1(t, F_t)) \tag{15}$$

with

$$d_1(t, F_t) = \frac{ln\left(\frac{F_t}{K}\right) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}} \tag{16}$$

and $N(.)$ the normal cumulative distribution.

Proof. The index option is assumed to be a function $C(t, F_t)$ of the futures price and time where $C(t, x)$ is a $C^{1,2}$ function of the time t and its spacial variable x. Application of *Itô 's lemma* gives,

$$dC(t, F_t) = \frac{\partial C(t, F_t)}{\partial t} dt + \frac{\partial C(t, F_t)}{\partial x} dF_t + \frac{1}{2} \frac{\partial^2 C(t, F_t)}{\partial x^2} (dF_t)^2$$

with dF_t given by

$$dF_t = \mu_f(t) F_t dt + \sigma F_t dW_t.$$

and

$$(dF_t)^2 = \sigma^2 F_t^2 dt.$$

We therefore obtain,

$$dC(t, F_t) = \left(\frac{\partial C(t, F_t)}{\partial t} + \mu_f(t) F_t \frac{\partial C(t, F_t)}{\partial x} + \frac{1}{2} \sigma^2 F_t^2 \frac{\partial^2 C(t, F_t)}{\partial x^2} \right) dt + \sigma F_t \frac{\partial C(t, F_t)}{\partial x} dW_t.$$

Identifying the terms with

$$dC_t = \mu_c(t) C_t dt + \sigma_c C_t dW_t$$

we get[7],

$$\mu_c(t) C_t = \frac{\partial C(t, F_t)}{\partial t} + \mu_f(t) F_t \frac{\partial C(t, F_t)}{\partial x} + \frac{1}{2} \sigma^2 F_t^2 \frac{\partial^2 C(t, F_t)}{\partial x^2}$$

and

$$\sigma_c C_t = \sigma F_t \frac{\partial C(t, F_t)}{\partial x}.$$

Using the Black's formula,

$$C(t, F_t) = e^{-r(T-t)} [F_t N(d_1(t, F_t)) - K N(d_2(t, F_t))]$$

with

$$d_1(t, F_t) = \frac{ln\left(\frac{F_t}{K}\right) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}}$$

and

$$d_2(t, F_t) = d_1(t, F_t) - \sigma\sqrt{T-t} \tag{17}$$

[7]We observe a discrepancy with equation (35) in Bhar, Chiarella and Runggaldier (2001).

we deduce that

$$\begin{aligned}\frac{\partial C(t,F_t)}{\partial x} &= e^{-r(T-t)}\left[\frac{\partial(F_tN(d_1(t,F_t)))}{\partial x} - K\frac{\partial N(d_2(t,F_t))}{\partial x}\right] \\ &= e^{-r(T-t)}\left[N(d_1(t,F_t)) + F_t\frac{\partial N(d_1(t,F_t))}{\partial x} - K\frac{\partial N(d_2(t,F_t))}{\partial x}\right].\end{aligned}$$

Substituting equation (17) on the last term we get

$$\begin{aligned}\frac{\partial N(d_2(t,F_t))}{\partial d_2(t,F_t)} &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2(t,F_t)^2} = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_1(t,F_t)^2}\right)e^{d_1(t,F_t)\sigma\sqrt{T-t}-\frac{1}{2}\sigma^2(T-t)} \\ &= \frac{\partial N(d_1(t,F_t))}{\partial d_1(t,F_t)}exp\left\{\left[ln\left(\frac{F_t}{K}\right)+\frac{1}{2}\sigma^2(T-t)\right]-\frac{1}{2}\sigma^2(T-t)\right\} \\ &= \frac{\partial N(d_1(t,F_t))}{\partial d_1(t,F_t)}\frac{F_t}{K}\end{aligned}$$

which cancels out with the second term leaving

$$\frac{\partial C(t,F_t)}{\partial x} = e^{-r(T-t)}N(d_1(t,F_t)).$$

therefore completing the proof.

□

Combining the drift and diffusion parameters derived in propositions 2.5, 2.6 and 2.7 into (5) yields

$$\boxed{\left\{\begin{array}{rcrcr} dS_t &=& (r-q+\sigma\lambda(t))S_tdt &+& \sigma S_tdW_t \\ dF_t &=& \sigma\lambda(t)F_tdt &+& \sigma F_tdW_t \\ dC_t &=& (r+\sigma_c\lambda(t))C_tdt &+& \sigma_cC_tdW_t \end{array}\right.} \tag{18}$$

with σ_c given by (15). This result is the risk-neutral evolution of the price dynamics in the risk-neutral framework.

As will see in chapter 4, it will be computationally convenient to express (18) in its logarithm form. Therefore using Itô 's lemma with $f(.) = ln(.)$, we obtain

$$\boxed{\left\{\begin{array}{rcrcr} ds_t &=& (r-q+\sigma\lambda(t)-\frac{\sigma^2}{2})dt &+& \sigma dW_t \\ df_t &=& (\sigma\lambda(t)-\frac{\sigma^2}{2})dt &+& \sigma dW_t \\ dc_t &=& (r+\sigma_c\lambda(t)-\frac{\sigma_c^2}{2})dt &+& \sigma_cdW_t \end{array}\right.} . \tag{19}$$

2.3 Black-Scholes Implied Volatility

As observed by Bhar, Chiarella and Runggaldier (2001), it is well known that the assumption of a constant volatility σ is not valid in the light of empirical facts. As a practical solution to handling the non-constancy of σ we will be using the implied volatility calculated from market prices using the Black-Scholes model.

Definition 2.8. Assuming that an underlying asset follows a geometric Brownian motion with constant volatility σ, the Black-Scholes formula gives the non-arbitrage price of an option on that underlying. By solving the formula for an unknown volatility parameter σ and taking as given the price of a call option C_t (all other variables being fixed), one can infer the volatility implicit in the observed market price called the *Black-Scholes implied volatility* denoted as $\hat{\sigma}$ such that

$$C_t = f(r, T, K, S_t, \hat{\sigma}).$$

Note that the actual value of the implied volatility of the option determined in this way depends on the choice of the model. Of the five variables necessary to specify the model (namely the interest rate r, maturity T, strike price K, underlying price S_t and volatility σ), all are directly observable except for the underlying volatility σ.

Unfortunately, there is no explicit solution to invert the Black-Scholes formula in order to recover $\hat{\sigma}$ from the market prices of the call option therefore one needs to make use of numerical method to solve

$$f(\sigma) - C_t = 0.$$

It can be shown that

$$\lim_{\sigma \to 0} f(\sigma) = (S_t e^{-q(T-t)} - K e^{-r(T-t)})^+$$

and

$$\lim_{\sigma \to +\infty} f(\sigma) = S_t e^{-q(T-t)}.$$

Therefore, because of the monotonicity of the Black-Scholes formula in the volatility parameter (strictly increasing over $(0, \infty)$) - called the *vega* measure - it is the unique solution for a given price.

Proposition 2.9. *The* vega *of an option is the sentitivity of the option price to a change in volatility of the underlying. The vega of a call option satisfies*

$$f'(\sigma) = \frac{\partial f(\sigma)}{\partial \sigma} = S_t e^{-q(T-t)} \sqrt{T-t} \frac{\partial N(d_1)}{\partial d_1} > 0 \tag{20}$$

Proof. Recall from appendix A that the Black-Scholes formula is given by

$$C_t = e^{-q(T-t)} S_t N(d_1) - e^{-r(T-t)} K N(d_2)$$

with

$$d_1 = \frac{ln\left(\frac{S_t}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \; d_2 = d_1 - \sigma\sqrt{T-t}.$$

Differentiating with respect to σ gives

$$\frac{\partial C_t}{\partial \sigma} = e^{-q(T-t)} S_t \frac{\partial N(d_1)}{\partial \sigma} - e^{-r(T-t)} K \frac{\partial N(d_2)}{\partial \sigma}.$$

Using the *chain rule*[8] in both terms on the right-hand side we get

$$\frac{\partial C_t}{\partial \sigma} = e^{-q(T-t)} S_t \left(\frac{\partial d_1}{\partial \sigma} \frac{\partial N(d_1)}{\partial d_1} \right) - e^{-r(T-t)} K \left(\frac{\partial d_2}{\partial \sigma} \frac{\partial N(d_2)}{\partial d_2} \right)$$

with

$$\frac{\partial d_1}{\partial \sigma} = \frac{\partial d_2}{\partial \sigma} - \sqrt{T-t}$$

and

$$\begin{aligned}
\frac{\partial N(d_2)}{\partial d_2} &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_2^2} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \right) e^{d_1 \sigma \sqrt{T-t} - \frac{1}{2}\sigma^2 (T-t)} \\
&= \frac{\partial N(d_1)}{\partial d_1} exp \left\{ \left[ln \left(\frac{S_t}{K} e^{(r-q)(T-t)} \right) + \frac{1}{2}\sigma^2 (T-t) \right] - \frac{1}{2}\sigma^2 (T-t) \right\} \\
&= \frac{\partial N(d_1)}{\partial d_1} \frac{S_t}{K} e^{(r-q)(T-t)}.
\end{aligned}$$

Combining gives

$$\begin{aligned}
\frac{\partial C_t}{\partial \sigma} &= e^{-q(T-t)} S_t \left(\frac{\partial d_2}{\partial \sigma} + \sqrt{T-t} \right) \frac{\partial N(d_1)}{\partial d_1} - e^{-r(T-t)} K \frac{\partial d_2}{\partial \sigma} \left(\frac{\partial N(d_1)}{\partial d_1} \frac{S_t}{K} e^{(r-q)(T-t)} \right) \\
&= e^{-q(T-t)} S_t \sqrt{T-t} \frac{\partial N(d_1)}{\partial d_1}.
\end{aligned}$$

All the terms involved are strictly positive therefore $\frac{\partial C_t}{\partial \sigma} > 0$ which complete the proof. $\square$

While there are many techniques for finding the root, one of the most commonly used is the *Newton-Raphson algorithm* given by

$$\sigma_{n+1} = \sigma_n - \frac{f(\sigma_n) - C_t}{f'(\sigma_n)}.$$

The Newton-Raphson's algorithm provides a rapid convergence but requires a vega of closed form (which is true for the Black-Scholes model but not for most practical pricing models). It should also be noted that without the correct initial value, solution that do exist can be overlooked. Manaster and Koehler (1982) gave the initial value for the first iteration to ensure convergence (see reference [14] for the proof)

$$\sigma_1^2 = \left| ln \left(\frac{S_t e^{-q(T-t)}}{K} \right) + r(T-t) \right| \frac{2}{T-t}.$$

[8]For two differentiable functions f and u: $\frac{df(u(x))}{dx} = \frac{du(x)}{dx} \frac{df(u(x))}{du(x)}$

2.4 The Market Price of Risk Dynamics

Following Bhar, Chiarella and Runggaldier (2001), the Ornstein–Uhlenbeck process will be considered for the dynamics of the market price of risk process $(\lambda_t)_{t>0}$.

Definition 2.10. Let (W_t) be a standard Brownian motion. A stochastic process $(\lambda_t)_{t\geq 0}$ is an *Ornstein–Uhlenbeck process* if it satisfies the following stochastic differential equation

$$\boxed{d\lambda_t = \kappa(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dW_t} \tag{21}$$

where $\bar{\lambda} > 0$ is the *mean reversion level* and $\kappa > 0$ the *mean reversion rate.*

We will show in chapter 5 that in order to estimate the parameters κ, $\bar{\lambda}$ and σ_λ of the Ornstein-Uhlenbeck process, we will be using the maximum likelihood estimator.

Proposition 2.11. *The solution of the Ornstein-Uhlenbeck process is given by*

$$\lambda_t = \lambda_0 e^{-\kappa t} + \bar{\lambda}(1 - e^{-\kappa t}) + \sigma_\lambda \int_0^t e^{\kappa(s-t)} dW_s.$$

Proof. The solution can be obtained by using the Itô product rule to the function $f(\lambda_t, t) = \lambda_t e^{\kappa t}$

$$\begin{aligned} df(\lambda_t, t) &= e^{\kappa t} d\lambda_t + \lambda_t de^{\kappa t} \\ &= e^{\kappa t}\left(\kappa(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dW_t\right) + \lambda_t \kappa e^{\kappa t} dt \\ &= e^{\kappa t}\kappa\bar{\lambda} dt + \sigma_\lambda e^{\kappa t} dW_t. \end{aligned}$$

Integrating from 0 to t gives

$$\begin{aligned} \lambda_t e^{\kappa t} &= \lambda_0 + \kappa\bar{\lambda} \int_0^t e^{\kappa s} ds + \sigma_\lambda \int_0^t e^{\kappa s} dW_s \\ &= \lambda_0 + \bar{\lambda}(e^{\kappa t} - 1) + \sigma_\lambda \int_0^t e^{\kappa s} dW_s \end{aligned}$$

which complete the proof. □

Proposition 2.12. *The first moment about* 0 *of the Ornstein-Uhlenbeck process is given by*

$$E[\lambda_t] = \lambda_0 e^{-\kappa t} + \bar{\lambda}(1 - e^{-\kappa t}).$$

Proof. By linearity of the expectation, it suffices to prove that

$$E\left[\int_0^t e^{\kappa(s-t)} dW_s\right] = 0.$$

Because $e^{\kappa(s-t)}$ is square-integrable over $[0, t]$ - i.e. $\int_0^t \left|e^{\kappa(s-t)}\right|^2 ds < \infty$, the Itô integral $\int_0^t e^{\kappa(s-t)} dW_s$ is a Martingale starting at 0 which complete the proof. □

Proposition 2.13. *The second moment about the mean of the Ornstein-Uhlenbeck process is given by*

$$cov(\lambda_t, \lambda_s) = \frac{\sigma_\lambda^2}{2\kappa} e^{-\kappa(s+t)}(e^{2\kappa(s\wedge t)} - 1).$$

Proof. By definition

$$\begin{aligned}
cov(\lambda_t, \lambda_s) &= E[(\lambda_s - E[\lambda_s])(\lambda_t - E[\lambda_t])] \\
&= \sigma_\lambda^2 E\left[\int_0^s e^{\kappa(u-s)} dW_u \int_0^t e^{\kappa(v-t)} dW_v\right] \\
&= \sigma_\lambda^2 e^{-\kappa(s+t)} E\left[\int_0^s e^{\kappa u} dW_u \int_0^t e^{\kappa v} dW_v\right] \\
&= \sigma_\lambda^2 e^{-\kappa(s+t)} E\left[\left(\int_0^{s\wedge t} e^{\kappa u} dW_u\right)^2\right].
\end{aligned}$$

Using the *Itô isometry* we get

$$\begin{aligned}
cov(\lambda_t, \lambda_s) &= \sigma_\lambda^2 e^{-\kappa(s+t)} E\left[\int_0^{s\wedge t} e^{2\kappa u} du\right] \\
&= \sigma_\lambda^2 e^{-\kappa(s+t)} \frac{(e^{2\kappa(s\wedge t)} - 1)}{2\kappa}
\end{aligned}$$

which complete the proof. □

Remark 2.14. If $s < t$ then

$$cov(\lambda_t, \lambda_s) = \frac{\sigma_\lambda^2}{2\kappa}(e^{-\kappa(t-s)} - e^{-\kappa(s+t)}).$$

2.5 The Euler-Maruyama Scheme

So far, we have been studying the price dynamics in the continous-time context. However the information available from the market are collected at discrete time instants. Therefore, in order to make inference from the observations, we need to discretize the set of equations (19).

Consider the general Itô process

$$dX_t = f(X_t)dt + g(X_t)dW_t \tag{22}$$

for $0 \leq t \leq T$ and with initial value $X_0 = x_0$. To apply the *Euler-Maruyama approximation*, we first uniformly discretize the interval [0, T] into N equal subintervals of width $\Delta t = \frac{T}{N}$ such that $0 = \tau_0 \leq \tau_1 \leq ... \leq \tau_k \leq ... \leq \tau_N = T$.

Remark 2.15. For conveniance X_{τ_k} will be denoted X_k.

Definition 2.16. The Euler-Maruyama approximation of the Itô process $(X_t)_{t>0}$ is the Markov chain satisfying the iterative equation,

$$X_{k+1} = X_k + f(X_k)\Delta t + g(X_k)\Delta W$$

with

$$\Delta W = W_{\tau_{k+1}} - W_{\tau_k} \sim N(0, \Delta t). \tag{23}$$

Following the Euler-Maruyama approximation scheme, the system (19) becomes

$$\boxed{\begin{cases} \lambda_{k+1} & = & \lambda_k & + & \kappa(\bar{\lambda} - \lambda_k)\Delta t & + & \sigma_\lambda \Delta W \\ s_{k+1} & = & s_k & + & (r - q + \sigma\lambda_k - \frac{\sigma^2}{2})\Delta t & + & \sigma\Delta W \\ f_{k+1} & = & f_k & + & (\sigma\lambda_k - \frac{\sigma^2}{2})\Delta t & + & \sigma\Delta W \\ c_{k+1} & = & c_k & + & (r + \sigma_c\lambda_k - \frac{\sigma_c^2}{2})\Delta t & + & \sigma_c\Delta W \end{cases}}. \tag{24}$$

2.6 Hidden Markov Model

We recall that the Markov property, named after the Andrey Markov (1856-1922), refers to the memoryless characteristic of a stochastic process.

Definition 2.17. A discrete-time stochastic process $(X_k)_{k>0}$, has the *Markov property* if it satisfies for all $k \in \mathbb{N}$

$$P(X_{k+1} = x_{k+1}|X_k = x_k, ..., X_0 = x_0) = P(X_{k+1} = x_{k+1}|X_k = x_k).$$

We define a *Markov process* as follow.

Definition 2.18. A *Markov process* is a discrete-time stochastic process $(X_k)_{k\geq 0}$ on the uncountable state-space $\mathbb{G} \in \mathbb{R}$, for which the *Markov property* holds for $x_k \in \mathbb{G}$.

The values (or *states*[9]) taken by $(\lambda_k)_{k>0}$ are not directly observable from the market data. Therefore $(\lambda_k)_{k>0}$ follows an *unobservable* Markov process, and the system (19) follows what is called an *Hidden Markov Model*. A Hidden Markov Model is a Markov model in which the Markov states are not observable. This means that the unobservable states λ_k can only be deduced from other dependent observable stochastic processes, in our case s_k, f_k and c_k.

[9]We recall that a random variable X is a measurable function $X : \Omega \to S$ from the sample space Ω to another measurable space S called the *state space*.

2.7 Summary

We have seen in this chapter the framework in which our modeling lies. In particular we derived the equations (24) which gives the evolution of the price processes in the risk-neutral framework and showed that they follow a hidden Markov model.

In the next chapter, we are going to present and derive the Kalman filter from the Bayes' theorem which will enable us to make inference about the $(\lambda_k)_{k>0}$ from the price observations in chapter 4.

3 Discrete Bayesian Filtering and the Kalman Filter

Although the Kalman filter was originally derived using the *orthogonal projection* method (see Kalman (1960)), it should be stressed that the Kalman filter has several interpretations. Other derivations have been from a *maximum-likelihood* standpoint - Rauch (1965), from a *recursive least-squares* standpoint or from a *Bayesian* standpoint - Ho and Lee (1964), Meinhold and Singpurwalla (1983).

Unfortunatly, much of the literature on the Kalman filter uses as engineering language, notation and style unfamiliar to most researchers outside the field. From a mathematical standpoint, it may be better understood if cast as a problem of recursive Bayesian inference.

"There are a number of mathematically equivalent ways of writing the Kalman filter equations. This can be confusing. You might read two different papers or books that present the Kalman filter equations, and the equations might look completely different. You may not know if one of the equations has a typographical error, or if they are mathematically equivalent. So you try to prove the equivalence of the two sets of equations only to arrive at a mathematical dead end, because it is not always easy to prove the equivalence of two sets of equations." - D. Simon, Optimal State Estimation, 2006.

It will be shown in this section that the Kalman filter is the optimal Baysian solution of linear models where the noises are normally distributed.

3.1 State-Space Equations

Definition 3.1. The general discrete-time *stochastic state-space model* is given by[10]

$$x_{n+1} = f(x_n, w_n) \tag{25}$$

$$y_n = g(x_n, v_n) \tag{26}$$

where equation (25) is called the *state equation* and equation (26) the *observation equation*. The processes w_n and v_n are independent *white noises* with unknown density probability.

Definition 3.2. A discrete-time process $\{w_k\}$ is a *white noise process* if and only if $E[w_k] = 0$, $E[w_k^2] = \sigma^2$ and $E[w_k w_{k+\tau}] = 0$ (distinct w_k uncorrelated, but may be dependent) for $t \neq \tau$.

Remark 3.3. If $w_k \sim N(0, \sigma^2)$, the process is said to be a *Gaussian white noise.*

[10]In the context stochastic filtering, the system input u_k generally added to the state-space equation is irrelevant. Its inclusion is only justified in *stochastic control* problems.

Remark 3.4. If a stronger restriction of independence is assumed, the process is said to be an *independent white noise process.*

Remark 3.5. For normally distributed random variables, uncorrelatedness and independence are equivalent.

Definition 3.6. When the system is linear, the equations (25) and (26) reduce to the following

$$x_{n+1} = F_n x_n + G_n w_n \tag{27}$$

$$y_n = H_n x_n + v_n. \tag{28}$$

3.2 Stochastic Filtering

As A. Bain and D. Crisan (2009) describs it, the process of using partial observations and a stochastic model to make inference about an evolving system is known as *stochastic filtering.*

There are three classes of stochastic estimation methods each of which may contains various *estimators* (Bayesian or frequentist).

Definition 3.7. Let a discrete *signal* process $X = \{X_k, k \in \mathbb{N}\}$ and a discrete *observation* process $Y = \{Y_n, n \in \mathbb{N}\}$. Given a realization of the sequence of observation Y, the discrete estimation problem consists of computing an estimate of X_k based on Y. If $k < n$, the problem is called a discrete *smoothing*[11] problem. If $k = n$, it is called a discrete *filtering* problem. And if $k > n$, it is called a discrete *prediction* problem.

The aim of solving the filtering problem is to determine the conditional distribution of the process X given the observation σ-algebra $\mathscr{Y}_k$ such that

$$\mathscr{Y}_k = \sigma\{y_i : 0 \leq i \leq k\}.$$

Remark 3.8. Stochastic filtering forms an integral part of *stochastic optimal control* in the case where the feedback signal in only allowed to depend on noisy observations rather than assuming that we can precisely observe the state of the system.

Stochastic filtering is an inverse problem, given Y and provided $f(.)$ in (25) and $g(.)$ in (26) are known, one needs to find the *optimal* or *suboptimal* estimate $\hat{x}_n$ (see section 3.5).

3.3 The Probability Measure and its Interpretation

The foundation of modern probability theory involve three basic axioms known as the *Kolmogorov axiom system* (1933) based on the Lebesgue theory of measure (1902).

[11] The benefit gained by waiting for more data to accumulate is that smoothing can yield a more accurate estimate than filtering.

Axiom 1. *Let A be an event from the sample space Ω. The probability function P satisfies*
1. $P(\emptyset) = 0$, $P(\Omega) = 1$,
2. Non-negativity: $P(A) \geq 0$ *for all* A,
3. Countable additivity: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

We deduce that the *probability function* P is a *measure* on the probability space $(\Omega, \mathscr{F}, P)$ satisfying $P(\Omega) = 1$ (unit measure). These axioms impose rigorous restrictions on the *probabilities*, but do not specify any particular meaning, nor any particular method of assigning them. Therefore many interpretations appeared in the literature. Does probability measure the real tendency of an event to occur (*frequency* probability), or is it just a measure of how strongly one believes it will occur (*Bayesian* probability)?

"It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. " - Leonard J. Savage (1954). The Foundations of Statistics.

The holder of frequency views assume that probabilities can only be assigned to results of experiments that can, at least in principle, be conducted repeatedly. This is in sharp contrast with the Bayesian view where the observations can be unique and non-repeatable.

Baysian inference focuses on the modification of belief in the light of new data. The beliefs about the unknown quantities are expressed directly in terms of probabilities. The revision of beliefs held prior to obtaining the data, to those posterior to the new data, is carried out through the use of the *Bayes' theorem.* The revised beliefs, expressed in terms of subjective probability, constitute the inferencial result.

Although discussion of the interpretation of probability seems remote for our purpose, it is necessary for understanding the different aims and methods of classical and Bayesian statistics. In discussing the Bayesian filter here, we will downplay the controversy concerning its application.

3.4 Bayesian Filtering

Bayesian statistics takes its name from the English Thomas Bayes (1701-1761) who published a paper after his death (1763) entitled *An Essay Towards Solving a Problem in the Doctrine of Chances.* The paper is concerned with the problem of inductive reasoning in statistics that is, using observations to make probability statements about hypotheses giving rise to those observations. In particular, the paper contains a version of an equality among several probabilities that today is known as the *Bayes' Theorem.*

Unfortunately, Bayesian theory did not gain attention until its modern form was discovered independently and developed much more thoroughly by Laplace in *Essai philosophique sur les probabilites* (1814). Two hundred year later Bayes' theorem has taken on a new importance and now provides the foundation for *Bayesian statistical inference*.

Proposition 3.9. *On a probability space* $(\Omega, \mathscr{F}, P)$*,* Bayes' theorem *relates the conditional and marginal probabilities of two random variables* x *and* y *such that*

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{29}$$

provided that the probability of y *does not equal zero.*

Proof. Bayes' theorem follows directly from the axioms of probability and the definition of conditional probability.

From the definition of conditional probability, we have

$$P(x, y) = P(x|y)P(y), \text{ and } P(x, y) = P(y|x)P(x).$$

Combining the two expressions complete the proof. □

In Bayesian inference, x is called the *parameter*, y the *observation*, $P(y|x)$ is the *likelihood function*, $P(x)$ the *prior density function* and $P(x|y)$ the *posterior density function.*

Remark 3.10. Since $P(y)$ is a normalizing constant independent of x (called the *evidence*) that serves to ensure that $P(x|y)$ integrate to one, Bayes' theorem is often written as

$$P(x|y) \propto P(y|x)P(x)$$

where $\propto$ indicates a proportionality relationship.

Remark 3.11. Properties of the posterior density $P(x|y)$ such as its maxima and its shape remain unchanged after normalization.

Bayes' theorem allows us to revise the probability $P(x)$ to $P(x|y)$ given the observation y. In a narrow sense, it is simply a rule for calculating conditional probabilities and by itself is not controversial. However in a broad sense it is the essence of a theory of learning from new observations involving the subjective nature of the prior probabilities.

From a Bayesian inference perspective, the posterior density $P(x|y)$ should always be regarded as the most general solution to the estimation problem, as it describes everything worth knowking after the observation. Even though the parameters are regarded as random, a single estimate is often of more practical interest than the conditional density itself. Given $P(x|y)$, what would be a reasonable choice for an estimate $\hat{x}$? The next section will answer this question.

3.5 Point-Estimation and Optimal Filtering

Definition 3.12. A filter is *optimal* if its average error is minimized for a given *point-estimator*.

Several *point-estimators* can be used to derive the Kalman filter and it can be shown that the Kalman filter is *optimal* for virtually any criterion. This is due to the fact that for a Gaussian distribution, the *mean*, the *mode*, the *median* and any choice of optimal estimate coincide, so there is a unique *best* estimate $\hat{x}$.

Definition 3.13. A *point-estimator* of a random sample $\xi = \{y_1, ..., y_n\}$ is the real-valued function g that maps the sample space to a set of sample estimates Θ such that

$$\hat{\theta} = g(\xi)$$

where $\hat{\theta} \in \Theta$ is called the *point-estimate* of the population parameter θ.

Remark 3.14. An *estimator* is a function of the sample ξ, while an *estimate* is the realized value of an estimator that is obtained when a sample is actually taken.

Remark 3.15. For brevity we will be using the same notation $\hat{\theta}$ for the *estimator* and the *estimate*. The interpretation should be clear from the context.

Desirable properties of an estimator are:

- *Unbiasedness.* An estimator $\hat{x}$ is said to be *unbiased* if $E[\hat{x}] - x = 0$.
- *Consistency.* An estimator $\hat{x}$ is said to be *consistent* if $\lim_{n\to\infty} \hat{x}_n = x$ in probability. In other words, the estimator should tend toward the true value x as the sample size increases.
- *Efficiency.* An estimator $\hat{x}$ is said to be *efficient* if its variance is less than that of alternative estimators[12].
- *Sufficiency.* An estimator $\hat{x}$ is said to be *sufficient* if it contains all of the information relevant to x that is possible to get from the sample.

In the Bayesian inference setting, two common choices of estimators are the *Minimum Mean-Squared Error* (MMSE) and the *Maximum A Posteriori* (MAP) which are detailed in the following sections.

[12]The *Cramer-Rao inequality* establishes the smallest possible variance that an unbiased estimator can have.

3.5.1 Minimum Mean-Squared Error Estimator

The MMSE estimator is aimed to find the *mean* of the posterior distribution.

Definition 3.16. The *Mean Square Error* (MSE) of an estimator $\hat{x}$ is given by

$$MSE(\hat{x}) = E[|x - \hat{x}|^2]. \tag{30}$$

Remark 3.17. We will denote the *argument of the supremum* and the *argument of the infinum* respectively by

$$\arg\max_\theta f(\theta) = \{\theta | \forall u : f(u) \leq f(\theta)\}$$

and

$$\arg\min_\theta f(\theta) = \{\theta | \forall u : f(u) \geq f(\theta)\}.$$

Proposition 3.18. *The* Minimal Mean Square Error *(MMSE) estimator $\hat{x}_{MMSE}$ is the estimator that minimizes the Mean Square Error (MSE) and is given by*

$$\hat{x}_{MMSE} = E[x|y]. \tag{31}$$

Proof.

$$\begin{aligned}
\hat{x}_{MMSE} &= \arg\min_{\hat{x}} E[|x - \hat{x}|^2] \\
&= \arg\min_{\hat{x}} \int\int |x - \hat{x}|^2 P(x, y) dx dy \\
&= \arg\min_{\hat{x}} \int |x - \hat{x}|^2 P(x|y) dx.
\end{aligned}$$

Differentiating with respect to $\hat{x}$ and equating to zero in order find the minimum gives

$$\frac{\partial}{\partial \hat{x}} \int |x - \hat{x}|^2 P(x|y) dx = \int 2(x - \hat{x}) p(x|y) dx = 0.$$

Therefore,

$$\hat{x} \int P(x|y) dx = \int x P(x|y) dx$$

and the proof is completed as

$$\int P(x|y) dx = E[1|y] = 1.$$

□

Remark 3.19. MMSE estimators may be inappropriate for multi-modal distributions.

Proposition 3.20. *The MMSE estimator is unbiased.*

Proof. From the *law of total expectation*

$$E[\hat{x}_{MMSE}] = E[E[x|y]] = E[x].$$

□

Proposition 3.21. *For two jointly distributed random vectors X and Y, the MMSE estimator is given by*

$$\hat{x}_{MMSE} = E[X] + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y]) \tag{32}$$

Proof. The proof is straightforward from proposition B.1 in appendix B. □

3.5.2 Maximum A Posteriori Estimator

The MAP estimator is aimed to find the *mode* of the posterior distribution.

Definition 3.22. The *Maximum A Posteriori* (MAP) estimator $\hat{x}_{MAP}$ is the estimator that maximizes the probability of the posterior probability function $P(x|y)$ and is given by

$$\hat{x}_{MAP} = \arg\max_x P(x|y). \tag{33}$$

Remark 3.23. From the Bayes' theorem (29) and the remark 3.11, the MAP estimator is equivalently characterized by

$$\hat{x}_{MAP} = \arg\max_x P(y|x)P(x) = \arg\max_x P(x,y) \tag{34}$$

where $P(y|x)$ is the likelihood function.

Remark 3.24. From remark 3.23, we observe that the MAP estimator coincides with the *maximum-likelihood* (ML) estimator when the prior $P(x)$ is uniform (see chapter 5).

Proposition 3.25. *For two jointly distributed random vectors X and Y, the MAP estimator is given by*

$$\hat{x}_{MAP} = E[X] + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y]) \tag{35}$$

Proof. For a normal distribution, its maximum is located at its first moment about 0. The mean of the conditional density is provided in proposition B.1 which complete the proof. □

3.6 Recursive Bayesian Filtering

The Bayes' theorem (29) can be applied sequentially.

Proposition 3.26. *Given the filtration* $\mathscr{Y}_k = \sigma\{y_i : 0 \le i \le k\}$ *of independent observations* y_i*, the* recursive Bayes' theorem *is given by*

$$P(x_{k+1}|\mathscr{Y}_{k+1}) = \frac{P(y_{k+1}|x_{k+1})P(x_{k+1}|\mathscr{Y}_k)}{P(y_{k+1}|\mathscr{Y}_k)}. \tag{36}$$

where $P(x_{k+1}|\mathscr{Y}_{k+1})$ *is the* posterior density *and* $P(x_{k+1}|\mathscr{Y}_k)$ *is the* prior density*.*

Proof. From the Bayes' theorem

$$P(x_{k+1}|y_{k+1}, \mathscr{Y}_k) = \frac{P(y_{k+1}, \mathscr{Y}_k|x_{k+1})P(x_{k+1})}{P(y_{k+1}, \mathscr{Y}_k)}.$$

But the independence of the observations implies

$$P(y_{k+1}, \mathscr{Y}_k|x_{k+1}) = P(y_{k+1}|x_{k+1})P(\mathscr{Y}_k|x_{k+1}),$$

and by definition of conditional probability

$$P(y_{k+1}, \mathscr{Y}_k) = P(y_{k+1}|\mathscr{Y}_k)P(\mathscr{Y}_k).$$

Hence

$$P(x_{k+1}|\mathscr{Y}_{k+1}) = \frac{P(y_{k+1}|x_{k+1})P(\mathscr{Y}_k|x_{k+1})P(x_{k+1})}{P(y_{k+1}|\mathscr{Y}_k)P(\mathscr{Y}_k)}$$

and using the Bayes' theorem on the right-hand side complete the proof. □

3.7 Conjugate Distributions

Definition 3.27. For a given likelihood function $P(y|x)$ in Bayes' theorem (29), if the posterior distribution $P(x|y)$ is in the same family as the prior distribution $P(x)$, the prior and posterior are then called *conjugate distributions*, and the prior is called a *conjugate prior for the likelihood.*

Conjugate priors are useful for sequential estimation, where the posterior of the current measurement is used as the prior in the next measurement. Unless a conjugate prior distribution is used, the posterior distribution typically becomes more complex with each added measurement, and the Bayes estimator cannot usually be calculated without resorting to numerical methods.

3.8 The Kalman Filter

We saw in the preceding section that if the prior distribution and likelihood function are conjugate distributions, then the posterior distribution has a closed-form expression.

The celebrated *Kalman filter* is the analytical solution to the recursive Bayesian estimation (36) when the state-space equations are linear and the noises are white, Gaussian and independent from each other.

Theorem 3.28. *Given a linear state-space model,*

$$\begin{aligned} x_{k+1} &= F_k x_k + G_k w_k \\ y_k &= H_k x_k + v_k \end{aligned}$$

where both the noises (w_k and v_k) and the initial state x_0 are normally distributed, the noises are white, and

$$E\left(\begin{bmatrix} w_k \\ v_k \\ x_0 \end{bmatrix} \begin{bmatrix} w_k^T & v_k^T & x_0^T & 1 \end{bmatrix}\right) = \begin{bmatrix} Q_k & 0 & 0 & 0 \\ 0 & R_k & 0 & 0 \\ 0 & 0 & P_0 & x_0 \end{bmatrix} \tag{37}$$

The conditional density for the one-step ahead prediction problem $P(x_{k+1}|\mathscr{Y}_k)$ is Gaussian. Initiated by $\hat{x}_0 = x_0$, P_0, the mean obeys the recursion

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - H_{k+1}\hat{x}_{k+1}^-), \tag{38}$$

and the posterior covariance is

$$P_{k+1} = (I - K_{k+1}H_{k+1})P_{k+1}^- \tag{39}$$

where

$$K_{k+1} = P_{k+1}^- H_{k+1}^T (H_{k+1}P_{k+1}^- H_{k+1}^T + R_{k+1})^{-1} \tag{40}$$

is the Kalman gain,

$$\hat{x}_{k+1}^- = F_k \hat{x}_k \tag{41}$$

and

$$P_{k+1}^- = F_k P_k F_k^T + G_k Q_k G_k^T \tag{42}$$

the covariance of the observation provided that $(H_{k+1}P_{k+1}^- H_{k+1}^T + R_{k+1})^{-1}$ exists.

Proof. [13] From propostion (36) we need to identify the three density functions $P(y_{k+1}|x_{k+1})$, $P(x_{k+1}|\mathscr{Y}_k)$ and $P(y_{k+1}|\mathscr{Y}_k)$ in order to recover $P(x_{k+1}|\mathscr{Y}_{k+1})$.

$$E[y_{k+1}|x_{k+1}] = H_{k+1}x_{k+1}$$

$$\begin{aligned} Cov(y_{k+1}|x_{k+1}) &= E[(y_{k+1} - E[y_{k+1}|x_{k+1}])(y_{k+1} - E[y_{k+1}|x_{k+1}])^T|x_{k+1}] \\ &= E[v_{k+1}v_{k+1}^T|x_{k+1}] = R_{k+1} \end{aligned}$$

[13] The full proof is also provided in the papper *A Bayesian Approach to Problems in Stochastic Estimation and Control* by Ho and Lee (1964).

Similarly,

$$\hat{x}_{k+1}^{-} \triangleq E[x_{k+1}|\mathscr{Y}_k] = F_k E[x_k|\mathscr{Y}_k] = F_k \hat{x}_k$$

$$\begin{aligned}
P_{k+1}^{-} &\triangleq Cov(x_{k+1}|\mathscr{Y}_k) = E[(x_{k+1} - E[x_{k+1}|\mathscr{Y}_k])(x_{k+1} - E[x_{k+1}|\mathscr{Y}_k])^T|\mathscr{Y}_k] \\
&= E[(F_k(x_k - \hat{x}_k) + G_k w_k)(F_k(x_k - \hat{x}_k) + G_k w_k)^T|\mathscr{Y}_k] \\
&= F_k E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T|\mathscr{Y}_k]F_k^T + G_k E[w_k w_k^T|\mathscr{Y}_k]G_k^T \\
&= F_k P_k F_k^T + G_k Q_k G_k^T
\end{aligned}$$

Finally,

$$\begin{aligned}
E[y_{k+1}|\mathscr{Y}_k] &= H_{k+1} E[x_{k+1}|\mathscr{Y}_k] \\
&= H_{k+1} F_k E[x_k|\mathscr{Y}_k] = H_{k+1} F_k \hat{x}_k
\end{aligned}$$

$$\begin{aligned}
Cov(y_{k+1}|\mathscr{Y}_k) &= E[(y_{k+1} - E[y_{k+1}|\mathscr{Y}_k])(y_{k+1} - E[y_{k+1}|\mathscr{Y}_k])^T|\mathscr{Y}_k] \\
&= E[(H_{k+1}(x_{k+1} - F_k\hat{x}_k) + v_{k+1})(H_{k+1}(x_{k+1} - F_k\hat{x}_k) + v_{k+1})^T|\mathscr{Y}_k] \\
&= H_{k+1} E[(x_{k+1} - F_k\hat{x}_k)(x_{k+1} - F_k\hat{x}_k)^T|\mathscr{Y}_k]H_{k+1}^T + E[v_{k+1}v_{k+1}^T|\mathscr{Y}_k] \\
&= H_{k+1} P_{k+1}^{-} H_{k+1}^T + R_{k+1}
\end{aligned}$$

Combining using equation (36) gives

$$\begin{aligned}
P(x_{k+1}|\mathscr{Y}_{k+1}) \quad = \quad & \frac{|H_{k+1}P_{k+1}^{-}H_{k+1}^T + R_{k+1}|^{1/2}}{(2\pi)^{n/2}|R_{k+1}|^{1/2}|P_{k+1}^{-}|^{1/2}} exp\left\{-\frac{1}{2}\right[\\
& (x_{k+1} - F_k\hat{x}_k)^T(P_{k+1}^{-})^{-1}(x_{k+1} - F_k\hat{x}_k) \\
& +(y_{k+1} - H_{k+1}x_{k+1})^T R_{k+1}^{-1}(y_{k+1} - H_{k+1}x_{k+1}) \\
& -(y_{k+1} - H_{k+1}F_k\hat{x}_k)^T(H_{k+1}P_{k+1}^{-}H_{k+1}^T + R_{k+1})^{-1}(y_{k+1} - H_{k+1}F_k\hat{x}_k) \\
&]\}
\end{aligned}$$

By completing the squares we obtain

$$P(x_{k+1}|\mathscr{Y}_{k+1}) = \frac{|H_{k+1}P_{k+1}^{-}H_{k+1}^T + R_{k+1}|^{1/2}}{(2\pi)^{n/2}|R_{k+1}|^{1/2}|P_{k+1}^{-}|^{1/2}} exp\left\{-\frac{1}{2}(x_{k+1} - \hat{x}_{k+1})^T P_{k+1}^{-1}(x_{k+1} - \hat{x}_{k+1})\right\}$$

where

$$\hat{x}_{k+1} = F_k\hat{x}_k + P_{k+1}^{-}H_{k+1}^T(H_{k+1}P_{k+1}^{-}H_{k+1}^T + R_{k+1})^{-1}(y_{k+1} - H_{k+1}F_k\hat{x}_k)$$

and

$$P_{k+1}^{-1} = (P_{k+1}^{-})^{-1} + H_{k+1}^T R_{k+1} H_{k+1}.$$

Lemma 3.29.

$$(A+CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}$$

Proof.

$$\begin{aligned}
&(A+CBD)(A^{-1} - A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1})\\
=&(A+CBD)A^{-1} - (A+CBD)A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}\\
=&I + CBDA^{-1} - (C+CBDA^{-1}C)(B^{-1}+DA^{-1}C)^{-1}DA^{-1}\\
=&I + CBDA^{-1} - CB(B^{-1}+DA^{-1}C)(B^{-1}+DA^{-1}C)^{-1}DA^{-1}\\
=&I + CBDA^{-1} - CBDA^{-1} = I
\end{aligned}$$

□

Therefore,

$$P_{k+1} = P^-_{k+1} - P^-_{k+1}H^T_{k+1}(H_{k+1}P^-_{k+1}H^T_{k+1} + R_{k+1})^{-1}H_{k+1}P^-_{k+1}.$$

which complete the proof. □

Remark 3.30. The difference $y_{k+1} - H_{k+1}F_k\hat{x}_k$ is called the *measurement innovation*. The innovation reflects the discrepancy between the predicted value and the actual measurement.

Remark 3.31. When $Q = cov(w_k)$ and $R = cov(v_k)$ are unknown, the solution is not unique and leads to a sub-optimal filter. When Q and R are constant over time, Oussalah and De Schutter (2003) provide a solution using the autocorrelation function of the innovation sequence to provide values for Q and R leading to an optimal filter.

Remark 3.32. For the scalar Kalman filter, we deduce from (40)

$$\lim_{R_{k+1}\to+\infty} K_{k+1} = 0 \Rightarrow \left\{ \begin{array}{l} \hat{x}_{k+1} = \hat{x}^-_{k+1} \\ P_{k+1} = P^-_{k+1} \end{array} \right. . \tag{43}$$

This means that we will put little confidence in a very noisy observation and would therefore weight it lightly. Similarly,

$$\lim_{Q_{k+1}\to+\infty} K_{k+1} = H^{-1}_{k+1} \Rightarrow \left\{ \begin{array}{l} \hat{x}_{k+1} = H^{-1}_{k+1}y_{k+1} \\ P_{k+1} = 0 \end{array} \right. . \tag{44}$$

In that case we are not certain of the ouput of the state model and therefore would weight the measurement heavily.

In chapter 4 we will derive a similar set of equations when the filter will be applied to our drift estimation problem. In that case, the noise will be correlated and the state space equations will be shifted by a constant.

3.9 Alternative Derivation of the Kalman Filter

[14] To avoid the tedious calculation involved when completing the square in the previous section, we observe that the following statement holds true at time $k+1$

$$P(x_{k+1}|\mathscr{Y}_{k+1}) = N\left(E[x_{k+1}|\mathscr{Y}_{k+1}], cov(x_{k+1}|\mathscr{Y}_{k+1})\right)$$

for a given $E[x_{k+1}|\mathscr{Y}_{k+1}]$ and $cov(x_{k+1}|\mathscr{Y}_{k+1})$ that needs to be identified. An important property of the multivariate Gaussian distribution is that if two sets of normally distributed variable have a conditional Gaussian distribution (one set conditioned over the other), then the joint distribution of is again Gaussian.

Lemma 3.33. *For two normally distributed random variables x_k, y_k and the σ-field $\mathscr{Y}_k = \sigma\{y_i : 0 \le i \le k\}$, if*

$$P(x_{k+1}|\mathscr{Y}_k, y_{k+1}|\mathscr{Y}_k) = N\left(E\left(\begin{bmatrix} x_{k+1}|\mathscr{Y}_k \\ y_{k+1}|\mathscr{Y}_k \end{bmatrix}\right), \begin{bmatrix} cov(x_{k+1}|\mathscr{Y}_k) & cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k) \\ cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k)^T & cov(y_{k+1}|\mathscr{Y}_k) \end{bmatrix}\right)$$

then

$$P(x_{k+1}|\mathscr{Y}_{k+1}) = N\left(E[x_{k+1}|\mathscr{Y}_{k+1}], cov(x_{k+1}|\mathscr{Y}_{k+1})\right)$$

and

$$\begin{cases} E[x_{k+1}|\mathscr{Y}_{k+1}] & = E[x_{k+1}|\mathscr{Y}_k] + K_{k+1}(y_{k+1} - E[y_{k+1}|\mathscr{Y}_k]) \\ \\ cov(x_{k+1}|\mathscr{Y}_{k+1}) & = cov(x_{k+1}|\mathscr{Y}_k) - K_{k+1} cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k)^T \end{cases}$$

where

$$K_{k+1} = cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k) cov(y_{k+1}|\mathscr{Y}_k)^{-1}.$$

Proof. The proof is straightforward from proposition B.1 when the random variables are pre-conditioned - see appendix B. □

Therefore if we find all the terms involved in $P(x_{k+1}|\mathscr{Y}_k, y_{k+1}|\mathscr{Y}_k)$ we can then completely identify $P(x_{k+1}|\mathscr{Y}_{k+1})$. As seen in the previous section

$$E[x_{k+1}|\mathscr{Y}_k] = \hat{x}^-_{k+1}$$

$$cov(x_{k+1}|\mathscr{Y}_k) = P^-_{k+1},$$

and

$$E[y_{k+1}|\mathscr{Y}_k] = H_{k+1}\hat{x}^-_{k+1}$$

$$cov(y_{k+1}|\mathscr{Y}_k) = H_{k+1}P^-_{k+1}H^T_{k+1} + R_{k+1}.$$

The only remaining terms which needs to be computed are

$$\begin{aligned} cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k) &= E[(x_{k+1} - E[x_{k+1}|\mathscr{Y}_k])(y_{k+1} - E[y_{k+1}|\mathscr{Y}_k])^T|\mathscr{Y}_k] \\ &= E[(x_{k+1} - E[x_{k+1}|\mathscr{Y}_k])(H_{k+1}(x_{k+1} - E[x_{k+1}|\mathscr{Y}_k]) + v_{k+1})^T|\mathscr{Y}_k] \\ &= P^-_{k+1}H^T_{k+1} \end{aligned}$$

[14] This approach has been suggested by Meinhold and Singpurwall (1983).

and

$$cov(x_{k+1}, y_{k+1}|\mathscr{Y}_k)^T = H_{k+1}P_{k+1}^-.$$

Combining the terms complete the proof.

3.10 Interpretation of the Kalman Filter

A common interpretation of the Kalman filter is in terms of the prediction/correction paradigm. In this approach, the model equation is used to predict one step forward what the next value should be, and then compares that estimate to the actual measurement. The actual measurement is used in conjunction with the prediction and noise characteristics to form the final estimate $\hat{x}$ and update a characterization of the noise (measure of how much the measurements are differing from the model).

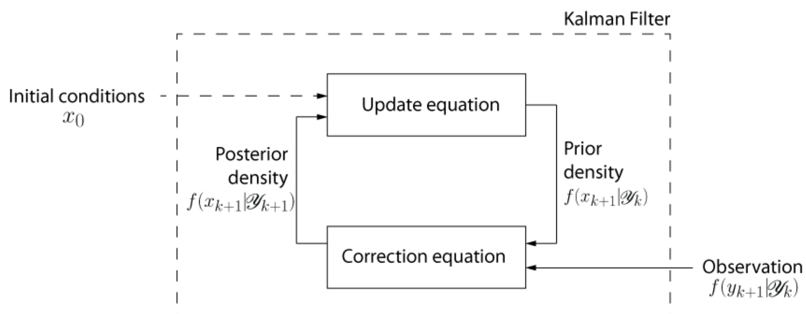

Figure 1: Ongoing Kalman filter cycle.

The *update equation* is responsible for projecting one step forward in time (in an unbiased fashion) the current state and error covariance estimates to obtain the prior estimates for the next time step.

The *correction equation* is responsible for incorporating a new measurement into the prior estimate to obtain the posterior estimate.

The Kalman filter generates new state estimates as new observations becomes available, thus opening the possibility of real-time estimation. As a by-product, the filter generates the estimation error covariance matrix P_{k+1}, which measures the uncertainty in the estimate.

Remark 3.34. P_k (and similarly P_k^-) can be seen as the covariance matrix of the posterior (prior) estimation error $e_k = x_k - \hat{x}_k$ (respectively $e_k^- = x_k - \hat{x}_k^-$), or the covariance matrix of the random variable x_k conditioned on $\mathscr{Y}_k$.

3.11 Application to Pair Trading

To illustrate the use of the Kalman filter in a quantitative trading strategy setting, we will develop a short example inspired from *statistical arbitrage* - see Dunis et al. (2010), Ajay Jasra (2011) or Helliott et al. (2005).

Pair trading was pioneered by Gerry Bamberger and later by Nunzio Tartaglia's quantitative group at Morgan Stanley in the 1980s. The strategy involves forming a portfolio of two *co-integrated* stocks whose relative pricing is away from its *equilibrium* state. It is assumed that over time they will move back to a rational equilibrium (mean revert) and therefore weightning appropriately the number of stocks held should result in a profit which will be a function of the *spread*.

We give the following preliminary definitions.

Definition 3.35. A process $\{X_k\}$ is said to be *strictly stationary*, if its finite-dimentional distributions are invariant under time-shifts - that is, for all $k \in \mathbb{N}$ and $\tau \in \mathbb{N}$ (called the *lag*), $\{X_k, ..., X_{k+n}\}$ and $\{X_{k+\tau}, ..., X_{k+n+\tau}\}$ have the same distribution

$$F(X_k, ..., X_{k+n}) = F(X_{k+\tau}, ..., X_{k+n+\tau})$$

where $F(.)$ is the joint cumulative distribution function.

Definition 3.36. A process $\{X_k\}$ is said to be *weakly stationary*, if for each $k \in \mathbb{N}$, its mean $E[X_k] = \mu$ is constant over time and it covariance $Cov(X_k, X_{k+\tau}) = \gamma(\tau)$ independs only on the lag τ and not on the time of k.

Remark 3.37. Because only the first two moments of the stochastic process have to be defined and independent of the time, this process is also referred to as being *second-order stationary* or *covariance stationary.*

Remark 3.38. A strictly stationary process is always weakly stationary. The converse is false in general but true for a normally distributed processes.

We deduce that a stationary process has a clear tendency to return to a constant value over time (no trend).

Definition 3.39. A process X_k is said to be *integrated of order one*, $I(1)$, if $\{X_k\}$ is non-stationary, but $\{X_k - X_{k-1}\}$ is stationary.

Introduced by Engle and Granger (1987), co-integration is a measure of the long-term dependencies in stochastic processes.

Definition 3.40. [15] Two processes $\{X_k\}$, $\{Y_k\}$ are said to be *co-integrated* if they are both $I(1)$ and there exist $\beta \in \mathbb{R}$ such that the *spread* $Y_k - \beta X_k$ is stationary in the weak sense.

The figure below illustrate clearly co-integrated processes[16], the two *move* together.

[15] See Banerjee et al. (2003) for a complete presentation.
[16] Source data: http://finance.yahoo.com

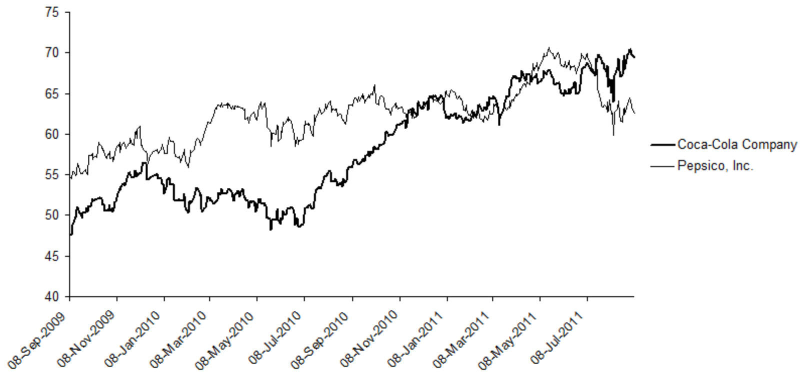

Figure 2: Adjusted daily closing prices of The Coca-Cola Company (NYSE: KO) and Pepsico, Inc. (NYSE: PEP) from 08-Sept-2011 to 06-Sept-2011.

Remark 3.41. Other potentially co-integrated pairs includes:

- Wal-Mart (WMT) and Target Corporation (TGT),
- Dell (DELL) and Hewlett-Packard (HPQ),
- Ford (F) and General Motors (GM).

.

The strategy consists of forming a corresponding pair of stocks (x, y) from the same industry (as they face similar systematic risks) and then evaluate whether the pair is co-integrated in the sample period [17].

Subsequently, the parameter β is estimated using the Kalman filter given the following state-space model

$$\begin{aligned} \beta_{k+1} &= \beta_k + w_k \\ y_k &= x_k \beta_k + v_k \end{aligned}$$

where $v_k \sim N(0, P)$ and $w_k \sim N(0, Q)$ are independent Gaussian white noises.

Following Theorem 3.28, the solution is given by

$$\hat{\beta}_{k+1} = \hat{\beta}_k + K_{k+1}(y_{k+1} - x_k \hat{\beta}_k)$$

$$P_{k+1} = (I - K_{k+1} x_{k+1}) P_{k+1}^-$$

[17]The main methods of testing for co-integration are the *Engle-Granger* (1987), the *Johansen procedure* (1988,1991) and the *Phillips-Ouliaris Cointegration Test*. These tests assume that the cointegrating vector remains constant during the period of study.

where the Kalman gain is

$$K_{k+1} = P_{k+1}^{-} x_{k+1}^{T} (x_{k+1} P_{k+1}^{-} x_{k+1}^{T} + R)^{-1}$$

and

$$P_{k+1}^{-} = P_k + Q.$$

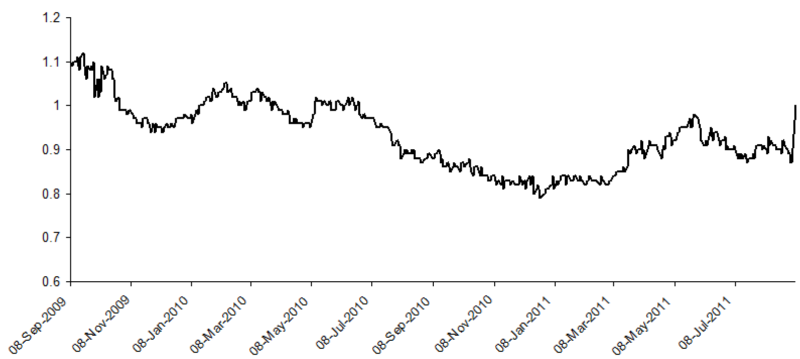

Figure 3: Estimated β_k of The Coca-Cola Company
(NYSE: KO) and Pepsico, Inc. (NYSE: PEP) from 08-Sept-2011 to
06-Sept-2011.
$Q = 0.003$, $R = 1$, $\beta_0 = 1$ and $P_0 = 1$.
(See appendix D for the source code.)

Because of its stationarity, the spread will mean revert to a constante μ over time:
(i) if $y_k - \beta_k x_k > \mu$ buy β_k of x_k and short-sell one unit of y_k,
(ii) if $y_k - \beta_k x_k < \mu$ buy one unit of y_k and short-sell β_k of x_k.

This strategy is *market neutral* meaning that it is profitable regardless of the market price direction.

However simplistic this strategy seems to be at first glance, it requires leverage to be truly profitable. And if the mean revertion rate is not high enough, the position will be too costly to hold on to.

4 Drift Estimation Problem and the Kalman Filter

We recall from the risk-neutral analysis in chapter 2 that our objective is to estimate the unknown drift function

$$\mu_k = (r - q) + \lambda_k \sigma \tag{45}$$

where

$$\lambda_{k+1} = \lambda_k + \kappa(\bar{\lambda} - \lambda_k)\Delta t + \sigma_\lambda \Delta W. \tag{46}$$

We say that the drift function μ_k is *parametric* and we are left with four parameters (see table 1) that we need to estimate in order to recover μ_k from the observations.

Parameter	Framework	Estimation Method
λ_k	Bayesian	Kalman filter
$\kappa, \bar{\lambda}, \sigma_\lambda$	Classical	Maximum likelihood

Table 1: Parameter Estimation.

In this section we will make use of the Kalman filter derived in chapter 3 to estimate λ_k. The remaining parameters $\kappa, \bar{\lambda}$ and σ_λ will be estimated in chapter 5 using the *maximum likelihood estimator*.

4.1 The State-Space Equations

In order to cast equation (24) into the *state space* model (27) and (28), we observe from equation (23) that $\Delta W = \sqrt{\Delta t}.\epsilon$ where $\epsilon \sim N(0,1)$. Therefore equations (24) can be written

$$\left\{ \begin{array}{lclclcl} \lambda_{k+1} & = & \lambda_k & + & \kappa(\bar{\lambda} - \lambda_k)\Delta t & + & \sigma_\lambda \sqrt{\Delta t}.\epsilon_k \\ s_{k+1} & = & s_k & + & (r - q + \sigma\lambda_k - \frac{\sigma^2}{2})\Delta t & + & \sigma\sqrt{\Delta t}.\epsilon_k \\ f_{k+1} & = & f_k & + & (\sigma\lambda_k - \frac{\sigma^2}{2})\Delta t & + & \sigma\sqrt{\Delta t}.\epsilon_k \\ c_{k+1} & = & c_k & + & (r + \sigma_c\lambda_k - \frac{\sigma_c^2}{2})\Delta t & + & \sigma_c\sqrt{\Delta t}.\epsilon_k \end{array} \right. . \tag{47}$$

4.1.1 The State Transition Equation

Writting $X_k = \lambda_k$ for the state variable, the equation (47a) gives us the *state transition equation*

$$\boxed{X_{k+1} = a + BX_k + R\epsilon_k} \tag{48}$$

where

$$\left\{ \begin{array}{ll} a & = \kappa\bar{\lambda}\Delta t \\ B & = 1 - \kappa\Delta t \\ R & = \sigma_\lambda\sqrt{\Delta t} \end{array} \right. \tag{49}$$

are scalar constants.

4.1.2 The Observation Equation

There are three different observations related to the same state variable X_k namely

$$\left\{\begin{array}{lcrcrcl} \Delta s_k & = & (r-q-\frac{\sigma^2}{2})\Delta t & + & \sigma\Delta t X_k & + & \sigma\sqrt{\Delta t}.\epsilon_k \\ \Delta f_k & = & -\frac{\sigma^2}{2}\Delta t & + & \sigma\Delta t X_k & + & \sigma\sqrt{\Delta t}.\epsilon_k \\ \Delta c_k & = & (r-\frac{\sigma_c^2}{2})\Delta t & + & \sigma_c\Delta t X_k & + & \sigma_c\sqrt{\Delta t}.\epsilon_k \end{array}\right. . \tag{50}$$

As a result, we will allow for a vector valued observation Y such that

$$Y_k = d_k + D_k X_k + H_k \epsilon_k \tag{51}$$

where

$$Y_k = \begin{bmatrix} \Delta s_k \\ \Delta f_k \\ \Delta c_k \end{bmatrix}, d_k = \begin{bmatrix} (r-q-\frac{\hat{\sigma_k}^2}{2})\Delta t \\ -\frac{\hat{\sigma_k}^2}{2}\Delta t \\ (r-\frac{\hat{\sigma_{c,k}}^2}{2})\Delta t \end{bmatrix}, H_k = \begin{bmatrix} \hat{\sigma_k}\sqrt{\Delta t} \\ \hat{\sigma_k}\sqrt{\Delta t} \\ \hat{\sigma_{c,k}}\sqrt{\Delta t} \end{bmatrix}, D_k = H_k\sqrt{\Delta t}. \tag{52}$$

In addition to the system noise ϵ_k, we will also assume the existance of a *market microstructure noise* term $Q_k\eta_k$ where $\eta_k \sim N(0,1)$. The market microstructure noise arise from the impact that the financial intermediaries linking buyers and sellers (such as brokers, market makers, ...) has on the market. These market participants are subject to certain rules that each exchange and national legislation impose on them, called the *market structure*. The market structure affects the price formation process both spatially and temporally in the form of bid-ask spread and nonsynchronicity of observed data. An introduction to the general idea of market microstructure can be found in O'Hara (1997). Although the structural forms of financial markets are different around the world, we will follow Bhar, Chiarella and Runggaldier (2004) and assumed a weigtning factor of the form $Q_k = \begin{bmatrix} 0.001 \\ 0.001 \\ 0.001 \end{bmatrix}$ which reflects a typical bid-ask spread.

Consequently, the observation equation can be written

$$\boxed{Y_k = d_k + D_k X_k + v_k} \tag{53}$$

where

$$v_k = H_k\epsilon_k + Q_k\eta_k. \tag{54}$$

Remark 4.1. We observe that we are in the case where there is correlation between the system noise and observation noise [18]

$$G_k = E[\epsilon_k \mathbf{v}_k^T] = \begin{bmatrix} \hat{\sigma_k}\sqrt{\Delta t} & \hat{\sigma_k}\sqrt{\Delta t} & \hat{\sigma_{c,k}}\sqrt{\Delta t} \end{bmatrix} = H_k^T. \tag{55}$$

[18] For two matrices A and B we have the following properties of transpose: $(A+B)^T = A^T + B^T$ and $(AB)^T = B^T A^T$.

4.2 The Kalman Filtering Solution

We introduce the following common notations used by Bhar, Chiarella and Runggaldier (2004) - see table 2, along with the parameters dimension - see table 3.

		Definition
$X_{i\|j}$	$E[X_i\|\mathscr{Y}_j]$	Expectation of X_i based on observations up to time j.
$P_{i\|j}$	$cov(X_i\|\mathscr{Y}_j)$	Covariance of X_i based on observations up to time j.
F_{k+1}	$cov(y_{k+1}\|\mathscr{Y}_k)$	Covariance of the observations.

Table 2: Summary of the notations.

Parameter	Dimension
$X_{k\|k}$, $X_{k+1\|k}$, $X_{k+1\|k+1}$, $P_{k+1\|k}$, $P_{k+1\|k+1}$, a_k, B, R, ϵ_k, η_k	1x1
P_k^{yx}, Y_k, d_k, H_k, D_k, v_k, Q_k, θ	3x1
P_k^{xy}, G_k, K_k	1x3
F_k, V_k	3x3

Table 3: Parameters dimension.

The Kalman filter solution for the state-space equations (48) and (53) described above is summarized in the following theorem.

Theorem 4.2. *The optimal filter for the state space* (48), (53) *consists of the* update equations

$$\begin{cases} X_{k+1|k} = a + BX_{k|k} \\ P_{k+1|k} = BP_{k|k}B^T + RR^T \end{cases} \tag{56}$$

and the correction equations

$$\begin{cases} X_{k+1|k+1} & = X_{k+1|k} + K_{k+1}(Y_{k+1} - D_{k+1}X_{k+1|k} - d_{k+1}) \\ P_{k+1|k+1} & = (1 - K_{k+1}D_{k+1})P_{k+1|k} + K_{k+1}G_{k+1}^T R \end{cases} \tag{57}$$

where

$$K_{k+1} = (P_{k+1|k}D_{k+1}^T + RG_{k+1})F_{k+1}^{-1} \tag{58}$$

is the Kalman gain and

$$F_{k+1} = D_{k+1}P_{k+1|k}D_{k+1}^T + D_{k+1}RG_{k+1} + G_{k+1}^T R^T D_{k+1}^T + H_{k+1}H_{k+1}^T + Q_{k+1}Q_{k+1}^T \tag{59}$$

is the covariance of the observations provided that F_{k+1}^{-1} exists[19]*.*

[19] We may replace F_{k+1}^{-1} by its pseudo-inverse.

Proof. The proofs of the Kalman filter provided in the chapter 3 assumed uncorelation between the model and the observation noises. Therefore following remark 4.1, we need to derive theorem 4.2 taking into account this new fact.

Following section 3.9, we need to identify the three density functions $P(Y_{k+1}, X_{k+1}|\mathscr{Y}_k)$, $P(X_{k+1}|\mathscr{Y}_k)$ and $P(Y_{k+1}|\mathscr{Y}_k)$ in order to recover $P(X_{k+1}|\mathscr{Y}_{k+1})$ from lemma 3.33.

(i) The update density is defined as

$$P(X_{k+1}|\mathscr{Y}_k) = N(X_{k+1|k}, P_{k+1|k})$$

with the expectation given by

$$\begin{aligned} X_{k+1|k} &= E[X_{k+1}|\mathscr{Y}_k] \\ &= a + BE[X_k|\mathscr{Y}_k] \\ &= a + BX_{k|k}. \end{aligned}$$

The variance is given by

$$P_{k+1|k} = E[(X_{k+1} - X_{k+1|k})(X_{k+1} - X_{k+1|k})^T|\mathscr{Y}_k]$$

with

$$\begin{aligned} X_{k+1} - X_{k+1|k} &= (a + BX_k + R\epsilon_k) - (a + BX_{k|k}) \\ &= B(X_k - X_{k|k}) + R\epsilon_k. \end{aligned}$$

Therefore

$$\begin{aligned} P_{k+1|k} =& E[B(X_k - X_{k|k})(B(X_k - X_{k|k}))^T|\mathscr{Y}_k] + \\ & E[B(X_k - X_{k|k})(R\epsilon_k)^T|\mathscr{Y}_k] + \\ & E[(R\epsilon_k)(B(X_k - X_{k|k}))^T|\mathscr{Y}_k] + \\ & E[(R\epsilon_k)(R\epsilon_k)^T|\mathscr{Y}_k] \\ =& BE[(X_k - X_{k|k})(X_k - X_{k|k})^T|\mathscr{Y}_k]B^T + \\ & BE[(X_k - X_{k|k})\epsilon_k^T|\mathscr{Y}_k]R^T + \\ & RE[\epsilon_k(X_k - X_{k|k})^T|\mathscr{Y}_k]B^T + \\ & RE[\epsilon_k\epsilon_k^T|\mathscr{Y}_k]R^T. \end{aligned}$$

Observing that,

$$E[(X_k - X_{k|k})(X_k - X_{k|k})^T|\mathscr{Y}_k] = P_{k|k},$$

$$E[(X_k - X_{k|k})\epsilon_k^T|\mathscr{Y}_k] = 0,$$

$$E[\epsilon_k (X_k - X_{k|k})^T | \mathscr{Y}_k] = 0,$$
$$E[\epsilon_k \epsilon_k^T | \mathscr{Y}_k] = 1.$$

We get

$$P_{k+1|k} = BP_{k|k}B^T + RR^T.$$

(ii) The observation density is given by

$$P(Y_{k+1}|\mathscr{Y}_k) = N(E[Y_{k+1}|\mathscr{Y}_k], F_{k+1})$$

with expectation given by

$$\begin{aligned} E[Y_{k+1}|\mathscr{Y}_k] &= d_{k+1} + D_{k+1}E[X_{k+1}|\mathscr{Y}_k] \\ &= d_{k+1} + D_{k+1}X_{k+1|k}. \end{aligned}$$

The variance is given by

$$F_{k+1} = E[(Y_{k+1} - Y_{k+1|k})(Y_{k+1} - Y_{k+1|k})^T | \mathscr{Y}_k]$$

with

$$\begin{aligned} Y_{k+1} - Y_{k+1|k} &= (d_{k+1} + D_{k+1}X_{k+1} + v_{k+1}) - (d_{k+1} + D_{k+1}X_{k+1|k} + E[v_{k+1}|\mathscr{Y}_k]) \\ &= D_{k+1}(X_{k+1} - X_{k+1|k}) + v_{k+1}. \end{aligned}$$

Therefore

$$\begin{aligned} F_{k+1} =& E[D_{k+1}(X_{k+1} - X_{k+1|k})(D_{k+1}(X_{k+1} - X_{k+1|k}))^T | \mathscr{Y}_k] + \\ & E[D_{k+1}(X_{k+1} - X_{k+1|k})v_{k+1}^T | \mathscr{Y}_k] + \\ & E[v_{k+1}(D_{k+1}(X_{k+1} - X_{k+1|k}))^T | \mathscr{Y}_k] + \\ & E[v_{k+1}v_{k+1}^T | \mathscr{Y}_k] \\ =& D_{k+1}E[(X_{k+1} - X_{k+1|k})(X_{k+1} - X_{k+1|k})^T | \mathscr{Y}_k]D_{k+1}^T + \\ & D_{k+1}E[(X_{k+1} - X_{k+1|k})v_{k+1}^T | \mathscr{Y}_k] + \\ & E[v_{k+1}(X_{k+1} - X_{k+1|k})^T | \mathscr{Y}_k]D_{k+1}^T + \\ & E[v_{k+1}v_{k+1}^T | \mathscr{Y}_k] \end{aligned}$$

with

$$\begin{aligned} E[v_{k+1}v_{k+1}^T | \mathscr{Y}_k] =& E[(H_{k+1}\epsilon_{k+1} + Q_{k+1}\eta_{k+1})(H_{k+1}\epsilon_{k+1} + Q_{k+1}\eta_{k+1})^T | \mathscr{Y}_k] \\ =& H_{k+1}E[\epsilon_{k+1}\epsilon_{k+1}^T | \mathscr{Y}_k]H_{k+1}^T + \\ & H_{k+1}E[\epsilon_{k+1}\eta_{k+1}^T | \mathscr{Y}_k]Q_{k+1}^T + \\ & Q_{k+1}E[\eta_{k+1}\epsilon_{k+1}^T | \mathscr{Y}_k]H_{k+1}^T + \\ & Q_{k+1}E[\eta_{k+1}\eta_{k+1}^T | \mathscr{Y}_k]Q_{k+1}^T + \\ =& H_{k+1}H_{k+1}^T + Q_{k+1}Q_{k+1}^T \end{aligned}$$

and

$$
\begin{aligned}
E[(X_{k+1} - X_{k+1|k})v_{k+1}^T|\mathscr{Y}_k] &= E[(B(X_k - X_{k|k}) + R\epsilon_k)v_{k+1}^T|\mathscr{Y}_k] \\
&= RE[\epsilon_k v_{k+1}^T|\mathscr{Y}_k] = RG_{k+1}.
\end{aligned}
$$

Combining gives

$$
F_{k+1} = D_{k+1}P_{k+1|k}D_{k+1}^T + D_{k+1}RG_{k+1} + G_{k+1}^T R^T D_{k+1}^T + H_{k+1}H_{k+1}^T + Q_{k+1}Q_{k+1}^T.
$$

(iii) The joint density of the state and observation at time $k+1$ given prior knowledge is

$$
P(X_{k+1}, Y_{k+1}|\mathscr{Y}_k)
$$

$$
\begin{aligned}
cov(X_{k+1}, Y_{k+1}|\mathscr{Y}_k) &= E[(X_{k+1} - X_{k+1|k})(Y_{k+1} - Y_{k+1|k})^T|\mathscr{Y}_k] \\
&= E[(X_{k+1} - X_{k+1|k})(D_{k+1}(X_{k+1} - X_{k+1|k}) + v_{k+1})^T|\mathscr{Y}_k] \\
&= E[(X_{k+1} - X_{k+1|k})((X_{k+1} - X_{k+1|k})^T D_{k+1}^T + v_{k+1}^T)|\mathscr{Y}_k] \\
&= P_{k+1|k}D_{k+1}^T + E[(X_{k+1} - X_{k+1|k})v_{k+1}^T|\mathscr{Y}_k] \\
&= P_{k+1|k}D_{k+1}^T + RG_{k+1}
\end{aligned}
$$

(iv) We finally deduce the transpose of the covariance of the joint density $P(X_{k+1}, Y_{k+1}|\mathscr{Y}_k)^T$,

$$
cov(X_{k+1}, Y_{k+1}|\mathscr{Y}_k)^T = D_{k+1}P_{k+1|k}^T + R^T G_{k+1}^T.
$$

(v) Making use of lemma (3.33) complete the proof. □

4.3 Discussion

It is important to ask what, if anything, can be gained from fitering the data. How well the state is known is measured by the estimation error covariance $P_{k+1|k} = E[(X_{k+1} - X_{k+1|k})(X_{k+1} - X_{k+1|k})^T|\mathscr{Y}_k]$. If little is gained from filtering, then we should consider remodeling the system (24). It should be noted that $P_{k+1|k}$ also depends on the initial condition P_0 and does not reflect the uncertainty in the estimate of the filtering alone.

5 Parameter Estimation using Maximum Likelihood

We have discussed Bayesian estimators in section 3.5. We will show in this section that the non-time-dependent parameters $\theta = \begin{bmatrix} \kappa & \bar{\lambda} & \sigma_\lambda \end{bmatrix}$ can be estimated using the *Maximum Likelihood Estimator* (MLE).

Contrary to Bayesian estimators where the parameters are regarded as random variables, in non-Bayesian setting the sought paramters are considered as fixed and the *Maximum Likelihood Estimator* (MLE) is a common choice of estimator.

Definition 5.1. The *likelihood function* is given by

$$L(\theta) = P(y|\theta) \tag{60}$$

where $P(\theta|y)$ is regarded as a function of the parameters θ.

Definition 5.2. The maximum likelihood estimator $\hat{\theta}_{ML}$ is the estimator that maximizes the likelihood function $L(\theta)$

$$\hat{x}_{ML}(y) = \arg\max_{\theta} L(\theta) \tag{61}$$

.

Remark 5.3. As for the *Maximum a Posteriori* (MAP), the MLE estimator need not exist nor be unique.

.

Therefore given the observations $\mathscr{Y}_N$, the parameter vector θ corresponding to the density most likely generating the observation is chosen.

We recall that the observations Y_i are independent and normally distributed (i.e. $Y \sim N(Y_{k+1|k}, F_{k+1})$). The likelihood function is given by

$$\begin{aligned}
L(\theta) = P(\theta|\mathscr{Y}_N) = P(\theta|Y_1)\cdots P(\theta|Y_N) &= \prod_{k=0}^{N-1} P(\theta|Y_{k+1}) \\
&= \prod_{k=0}^{N-1} \frac{1}{(2\pi)^{3/2}|F_{k+1}|^{1/2}} e^{-\frac{1}{2}(Y_{k+1}-Y_{k+1|k})^T F_{k+1}^{-1}(Y_{k+1}-Y_{k+1|k})} \\
&= \frac{1}{(2\pi)^{3N/2}} \prod_{k=0}^{N-1} \frac{1}{|F_{k+1}|^{1/2}} e^{-\frac{1}{2}(Y_{k+1}-Y_{k+1|k})^T F_{k+1}^{-1}(Y_{k+1}-Y_{k+1|k})}.
\end{aligned}$$

Maximizing the logarithm of the likelihood function (called *log-likelihood*) is equivalent to maximizing the likelihood function as log(.) is monotonous,

therefore

$$
\begin{aligned}
l(\theta) = log(L(\theta)) &= -\frac{3N}{2} log(2\pi) + \sum_{k=0}^{N-1} log\left(\frac{1}{|F_{k+1}|^{1/2}} e^{-\frac{1}{2}(Y_{k+1}-Y_{k+1|k})^T F_{k+1}^{-1}(Y_{k+1}-Y_{k+1|k})}\right) \\
&= -\frac{3N}{2} log(2\pi) + \sum_{k=0}^{N-1}\left(-\frac{1}{2} log|F_{k+1}| - \frac{1}{2}(Y_{k+1}-Y_{k+1|k})^T F_{k+1}^{-1}(Y_{k+1}-Y_{k+1|k})\right) \\
&= -\frac{3N}{2} log(2\pi) - \frac{1}{2}\sum_{k=0}^{N-1} log|F_{k+1}| - \frac{1}{2}\sum_{k=0}^{N-1}(Y_{k+1}-Y_{k+1|k})^T F_{k+1}^{-1}(Y_{k+1}-Y_{k+1|k}).
\end{aligned}
$$

The standard analytical method of finding the MLE is to take the first partial derivatives of the log-likelihood with respect to each parameter θ_i. Therefore from here, we will resort to numerical methods to find the roots of $\frac{\partial l}{\partial \theta_i}$ and deduce the best approximation of $\hat{\theta}_{ML}$.

6 Discussion

It may be argue that the assumptions inherent of this approach may seems too restrictive for practical applications. On one hand, we hypothesized that the assets on the market assumed a parametric form driven by non-arbitrage arguments. The Black-Scholes model is indeed an elegant model but it does not perform very well in practice. It is well known that stock prices jump on occasions and do not always move in the smooth manner predicted by the geometric Brownian motion model. Stock prices also tend to have fatter tails than those predicted by this model.

The simplifying assumption of constant interest rate must also be acknowledged. A tremendous amount of research are currently taking place to model the yield curve accurately. Including such model would have necessarily increase the complexity of our approach and should be considered as further improvement.

On the other hand, because of its real-time recursive processing and its easily implemented algorithm, the Kalman filter will continue to play a very important role in finance. Bearing in mind that the Kalman filter is limited by its assumptions (linearity of the state-space equations and normality of the noises), numerous nonlinear filtering methods along its line have been developed to overcome its limitations. As an example, the *extended Kalman Filter* is an important result in non-linear filtering.

The more generic non-parametric approach seems more appealing. This is an essentially a more difficult problem being in general infinite-dimensional - see M. H. Davis and Steven I. Marcus [18] for an introduction. There is an extensive literature on non-parametric estimation for the drift in diffusion processes, see Ian W. McKeague (1989), Isao Shoji (2003), etc.

Quantitative trading, also known as *algorithmic trading*, now accounts for a large portion of the world trading volume. It is the trading of securities based strictly on the buy and sell decisions of automated computer algorithms. The question arise as to how the instantaneous drift information can be used in investment decision-making. Suppose we monitor continuously the market, we want to be able to detect as quickly as possible a change in the drift estimated. An investor would find such information interesting so then he could adjust his investments according to the market regimes. In conjuction with other engineering areas, such as *pattern recognition* in signal processing or *machine learning* in computer science, further synergies could be developed.

Nevertheless, even if the model described here gives some interesting results, it is my belief that the explanation and prediction power stays very low and cannot be used for an investment purpose.

However, according to Jansen and Hagenaars (2006) Bayesian methods hold great promises for model calibration. Characteristically, complex models draw their information from diverse sources: observations at several spatial and temporal scales, experiments with sub-models, information from literature, etc. Bayesian methods may bring conceptual clarity in the calibration of complex models because they enable combination of heterogeneous information about parameter values accompanied by an indication of their accuracy. In addition, assumptions may be used as prior information.

APPENDIX

A The Black's Formula

Proposition A.1. *On a stochastic basis* $(\Omega, \mathscr{F}_t, P)$ *with* $t \in [0, T]$ *live a risky asset* S *and a riskless bond* B *such that,*

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{62a}$$
$$dB_t = r B_t dt. \tag{62b}$$

The parameters μ, σ, r *are constante and the filtration* $\mathscr{F}_t$ *is the sigma-field generated by a standard one-dimensional Brownian motion* $(W_t)_{t>0}$ *such that*

$$\mathscr{F}_t = \sigma\{W_t\}.$$

The Black-Scholes formula *is given by*

$$C_0 = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2) \tag{63}$$

where

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}} \tag{64}$$

and

$$d_2 = d_1 - \sigma\sqrt{T}. \tag{65}$$

Proof. The Black-Scholes formula can be derived from either the martingale pricing approach or the replicating strategy/partial differential equation (PDE) approach. While the original derivation from Fischer Black and Myron Scholes in *The pricing of options and corporate liabilities* (1970) was based on PDE, the martingale pricing approach is more general and tractable.

The argument goes as follow. Under the Risk-neutral measure $\mathbb{Q}$, the discounted option payoff must be a Martingale,

$$\begin{aligned} C_0 &= E^{\mathbb{Q}}\left[\frac{(S_T - K)^+}{e^{rT}}\right] \\ &= E^{\mathbb{Q}}\left[(S_T - K) I_{\{S_T \geq K\}} e^{-rT}\right] \\ &= E^{\mathbb{Q}}[\tilde{S}_T I_{\{S_T \geq K\}}] - K e^{-rT} E^{\mathbb{Q}}[I_{\{S_T \geq K\}}]. \end{aligned}$$

where $\tilde{S}_T = S_T e^{-rT}$. A straightforward use of Itô 's lemma on the equations (62a) and (62b) yields

$$\begin{aligned} S_t &= S_0 e^{(\mu - q - \frac{\sigma^2}{2})t + \sigma W_t} \\ B_t &= B_0 e^{rt}. \end{aligned}$$

Therefore

$$\tilde{S}_T = S_T e^{-rT} = S_0 e^{(r-q-\frac{\sigma^2}{2})T+\sigma W_T^{\mathbb{Q}}} e^{-rT} = S_0 e^{-(q+\frac{\sigma^2}{2})T+\sigma W_T^{\mathbb{Q}}}.$$

Let us first look at the second expectation,

$$\begin{aligned} E^{\mathbb{Q}}[I_{\{S_T \geq K\}}] &= \mathbb{Q}(S_T \geq K) \\ &= \mathbb{Q}\left(S_0 e^{\left(r-q-\frac{\sigma^2}{2}\right)T+\sigma W_T^{\mathbb{Q}}} \geq K\right) \\ &= \mathbb{Q}\left(W_T^{\mathbb{Q}} \geq \frac{ln\left(\frac{K}{S_0}\right) - \left(r-q-\frac{\sigma^2}{2}\right)T}{\sigma}\right). \end{aligned}$$

As $W_T^{\mathbb{Q}}$ is normally distributed with zero mean, taking advantage of the symmetry gives

$$\mathbb{Q}\left(Z^{\mathbb{Q}} \leq \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q-\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right) = N(d_2)$$

where $N(x)$ is the standard cumulative normal distribution

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du$$

and

$$d_2 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q-\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}.$$

Let's now look at the first expectation

$$E^{\mathbb{Q}}[\tilde{S}_T I_{\{S_T \geq K\}}] = S_0 e^{-qT} E^{\mathbb{Q}}\left[e^{-\frac{\sigma^2}{2}T+\sigma W_T^{\mathbb{Q}}} I_{\{S_T \geq K\}}\right].$$

Let us define the measure

$$\mathbb{R}(E) = \int_E e^{-\frac{\sigma^2}{2}T+\sigma W_T^{\mathbb{Q}}} d\mathbb{Q}$$

under which

$$E^{\mathbb{Q}}\left[e^{-\frac{\sigma^2}{2}T+\sigma W_T^{\mathbb{Q}}} I_{\{S_T \geq K\}}\right] = E^{\mathbb{R}}\left[I_{\{S_T \geq K\}}\right].$$

The *Girsanov theorem* tells us that $W_t^{\mathbb{R}} = W_t^{\mathbb{Q}} - \sigma t$ is a Brownian motion under $\mathbb{R}$. Using the same argument as earlier under $\mathbb{R}$ instead of $\mathbb{Q}$ we obtain

$$
\begin{aligned}
\mathbb{R}(S_T \geq K) &= \mathbb{R}\left(S_0 e^{\left(r-q-\frac{\sigma^2}{2}\right)T+\sigma W_T^{\mathbb{Q}}} \geq K\right) \\
&= \mathbb{R}\left(S_0 e^{\left(r-q+\frac{\sigma^2}{2}\right)T+\sigma (W_T^{\mathbb{Q}}-\sigma T)} \geq K\right) \\
&= \mathbb{R}\left(S_0 e^{\left(r-q+\frac{\sigma^2}{2}\right)T+\sigma W_T^{\mathbb{R}}} \geq K\right) \\
&= \mathbb{R}\left(W_T^{\mathbb{R}} \geq \frac{ln\left(\frac{K}{S_0}\right) - \left(r-q+\frac{\sigma^2}{2}\right)T}{\sigma}\right) \\
&= \mathbb{R}\left(Z^{\mathbb{R}} \leq \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q+\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right) \\
&= N(d_1).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
C_0 &= E^{\mathbb{Q}}[\tilde{S}_T I_{\{S_T \geq K\}}] - Ke^{-rT}E^{\mathbb{Q}}[I_{\{S_T \geq K\}}] \\
&= S_0 e^{-qT} E^{\mathbb{R}}\left[I_{\{S_T \geq K\}}\right] - Ke^{-rT}E^{\mathbb{Q}}[I_{\{S_T \geq K\}}]
\end{aligned}
$$

giving the Black-Scholes formula.

Finally we observe that

$$
\begin{aligned}
d_1 - \sigma\sqrt{T} &= \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q+\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} - \sigma\sqrt{T} \\
&= \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q+\frac{\sigma^2}{2}\right)T - \sigma^2 T}{\sigma\sqrt{T}} \\
&= \frac{ln\left(\frac{S_0}{K}\right) + \left(r-q-\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \\
&= d_2.
\end{aligned}
$$

$\square$

A slightly more general model can be derived when the price process start at time t instead of 0.

Corollary A.2. *Under the same assumptions as in proposition A.1, the* Black-Scholes formula *is alternatively given over* $[t, T]$ *by*

$$C_t = e^{-q(T-t)} S_t N(d_1) - e^{-r(T-t)} K N(d_2) \tag{66}$$

where

$$d_1 = \frac{ln\left(\frac{S_t}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{67}$$

and

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{68}$$

Proof. For $h > 0$,

$$\begin{aligned} S_{t+h} &= S_0 e^{(\mu - q - \frac{\sigma^2}{2})(t+h) + \sigma W_{t+h}} \\ &= \left(S_0 e^{(\mu - q - \frac{\sigma^2}{2})t + \sigma W_t}\right)\left(e^{(\mu - q - \frac{\sigma^2}{2})h + \sigma W_{t+h} - \sigma W_t}\right). \end{aligned}$$

Identifying the the first term as S_t and defining $u = t + h > 0$ we get

$$S_u = S_t e^{(\mu - q - \frac{\sigma^2}{2})(u-t) + \sigma(W_u - W_t)}.$$

Lemma A.3. *Let* $h > 0$*, if the* W *is a standard Brownian motion, then the process* $B_t = W_{t+h} - W_h$ *is also a standard Brownian motion independent of* $(W_t)_{0 \le t \le h}$*.*

Remark A.4. The *strong Markov property* is an extension of the *reborn* property of the previous lemma from fixed times h to random *stopping times* τ.

Therefore for $t \le u \le T$

$$W_u - W_t \sim N(0, u - t)$$

is independent of t over $[t, T]$. And following identical steps as in proposition A.1 but with all expectations conditioned on $\mathscr{F}_t = \sigma\{S_t\}$ complete the proof. □

Proposition A.5. *The* Black's formula *is given by*

$$C_t = B_t[F_t N(d_1) - K N(d_2)] \tag{69}$$

where

$$F_t = S_t e^{(r-q)(T-t)} \tag{70}$$

is the forward price *at time* t *for delivery at time* T*,*

$$B_t = e^{-r(T-t)} \tag{71}$$

the discount factor*,*

$$d_1 = \frac{ln\left(\frac{F_t}{K}\right) + \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}} \tag{72}$$

and

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{73}$$

Proof. The proof is straightforward from corollary A.2.

□

B Conditional Multivariate Normal Distribution

Proposition B.1. *If X and Y are two jointly normal vectors of respective dimensions p and q with $p+q=n$ such that*

$$P(X,Y) = N\left(E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right), \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$$

then

$$P(X|Y) = N\left(E[X|Y], Var[X|Y]\right)$$

with

$$\begin{cases} E[X|Y] & = E[X] + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y]) \\ Var[X|Y] & = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \end{cases}.$$

The converse is also true.

Proof. Note that the symmetry $\Sigma^T = \Sigma$ of the covariance matrix implies that Σ_{xx} and Σ_{yy} are symmetric, while $\Sigma_{xy}^T = \Sigma_{yx}$.

The joint multivariate normal distribution can be written as the following

$$P(X,Y) = Ce^{-\frac{1}{2}Q(X,Y)}$$

where

$$Q(X,Y) = \left(\begin{bmatrix} X \\ Y \end{bmatrix} - E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)\right)^T \Sigma^{-1} \left(\begin{bmatrix} X \\ Y \end{bmatrix} - E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)\right)$$

and

$$C = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}.$$

If we can find the terms C_1, C_2, $Q_1(X,Y)$ and $Q_2(Y)$ such that

$$\begin{aligned} P(X,Y) &= \left(C_1 e^{-\frac{1}{2}Q_1(X,Y)}\right)\left(C_2 e^{-\frac{1}{2}Q_2(Y)}\right) \\ &= N(X, b, A).N(E[Y], \Sigma_{yy}) \end{aligned}$$

for some vector b and matrix A, we can then identify the expression with the *product rule of probability*

$$P(X,Y) = P(X|Y)P(Y).$$

It will be convenient to work with the inverse of the covariance matrix

$$\Lambda = \Sigma^{-1}$$

and its partitioned form

$$\Lambda = \begin{bmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{bmatrix}.$$

Substituing Λ in the expression of $Q(X,Y)$ yields

$$\begin{aligned}
Q(X,Y) =& \left(\begin{bmatrix} X \\ Y \end{bmatrix} - E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)\right)^T \Lambda \left(\begin{bmatrix} X \\ Y \end{bmatrix} - E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)\right) \\
=& \begin{bmatrix} X - E[X] \\ Y - E[Y] \end{bmatrix}^T \begin{bmatrix} \Lambda_{xx} & \Lambda_{xy} \\ \Lambda_{yx} & \Lambda_{yy} \end{bmatrix} \begin{bmatrix} X - E[X] \\ Y - E[Y] \end{bmatrix} \\
=&(X - E[X])^T \Lambda_{xx}(X - E[X]) + (X - E[X])^T \Lambda_{xy}(Y - E[Y]) + \\
& (Y - E[Y])^T \Lambda_{yx}(X - E[X]) + (Y - E[Y])^T \Lambda_{yy}(Y - E[Y])
\end{aligned}$$

Because the inverse of a symmetric matrix is also symmetric, we see that Λ_{xx} and Λ_{yy} are symmetric, while $\Lambda_{xy}^T = \Lambda_{yx}$. Therefore the third term (of dimension one) can be written

$$\begin{aligned}
(Y - E[Y])^T \Lambda_{yx}(X - E[X]) &= \left[(X - E[X])^T \Lambda_{xy}(Y - E[Y])\right]^T \\
&= (X - E[X])^T \Lambda_{xy}(Y - E[Y])
\end{aligned}$$

giving

$$\begin{aligned}
Q(X,Y) =&(X - E[X])^T \Lambda_{xx}(X - E[X]) \\
&+ 2(X - E[X])^T \Lambda_{xy}(Y - E[Y]) \\
&+ (Y - E[Y])^T \Lambda_{yy}(Y - E[Y]).
\end{aligned}$$

It should be stressed that for instance Λ_{xx} is not simply given by the inverse of Σ_{xx}. In fact, the relation between the inverse of a partitioned matrix and the inverses of its partitions is given by the following lemma,

Lemma B.2. *The inverse of a non-singular partitioned matrix A is given by*

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} M & -MA_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}M & A_{22}^{-1} + A_{22}^{-1}A_{21}MA_{12}A_{22}^{-1} \end{bmatrix}$$

where

$$M = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}.$$

M^{-1} is known as the Schur complement *of A^{-1} with respect to A_{22}.*

Proof. Consider an $n \times n$ matrix A partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

The inverse matrix $B = A^{-1}$ can also be divided partitioned as

$$B = A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

It follows that

$$AA^{-1} = AB = I_n$$

and

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}.$$

Therefore,

$$\left\{ \begin{array}{ll} A_{11}B_{11} + A_{12}B_{21} & = I_p \\ A_{11}B_{12} + A_{12}B_{22} & = 0 \\ A_{21}B_{11} + A_{22}B_{21} & = 0 \\ A_{21}B_{12} + A_{22}B_{22} & = I_q \end{array} \right. \Leftrightarrow \left\{ \begin{array}{ll} B_{11} & = A_{11}^{-1} - A_{11}^{-1}A_{12}B_{21} \\ B_{12} & = -A_{11}^{-1}A_{12}B_{22} \\ B_{21} & = -A_{22}^{-1}A_{21}B_{11} \\ B_{22} & = A_{22}^{-1} - A_{22}^{-1}A_{21}B_{12} \end{array} \right. \tag{74}$$

Substituting (3)→(1) to obtain B_{11} and (2)→(4) to obtain B_{22} we get

$$\left\{ \begin{array}{ll} B_{11} & = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} \\ B_{12} & = -A_{11}^{-1}A_{12}\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \\ B_{21} & = -A_{22}^{-1}A_{21}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} \\ B_{22} & = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{array} \right. \tag{75}$$

Equivalently, $BA = I_n$ gives the following expression

$$\left\{ \begin{array}{ll} B_{11} & = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} \\ B_{12} & = -\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}A_{12}A_{22}^{-1} \\ B_{21} & = -\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}A_{21}A_{11}^{-1} \\ B_{22} & = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{array} \right. . \tag{76}$$

Defining $B_{11} = M$ we can write

$$\left\{ \begin{array}{ll} B_{11} & = M \\ B_{12} & = -MA_{12}A_{22}^{-1} \\ B_{21} & = -A_{22}^{-1}A_{21}M \\ B_{22} & = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{array} \right. .$$

We will make use of the following lemma to express B_{22} as a function of M.

Lemma B.3.

$$(A - CBD)^{-1} = A^{-1} + A^{-1}C(B^{-1} - DA^{-1}C)^{-1}DA^{-1} \tag{77}$$

Proof.

$$\begin{aligned} &(A - CBD)(A^{-1} + A^{-1}C(B^{-1} - DA^{-1}C)^{-1}DA^{-1}) \\ =&(A - CBD)A^{-1} + (A - CBD)A^{-1}C(B^{-1} - DA^{-1}C)^{-1}DA^{-1} \\ =&I - CBDA^{-1} + (C - CBDA^{-1}C)(B^{-1} - DA^{-1}C)^{-1}DA^{-1} \\ =&I - CBDA^{-1} + CB(B^{-1} - DA^{-1}C)(B^{-1} - DA^{-1}C)^{-1}DA^{-1} \\ =&I - CBDA^{-1} + CBDA^{-1} = I \end{aligned}$$

$\square$

Therefore,

$$B_{22} = A_{22}^{-1} + A_{22}^{-1} A_{21} M A_{12} A_{22}^{-1}$$

which complete the proof.

$\square$

From the above lemma, we deduce the following expression

$$\begin{cases} \Lambda_{xx} & = \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} \\ \Lambda_{xy} & = -\left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ \Lambda_{yy} & = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \end{cases}$$

giving

$$\begin{aligned} Q(X,Y) &= (X - E[X])^T \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} (X - E[X]) \\ &- 2(X - E[X])^T \left(\left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\right)(Y - E[Y]) \\ &+ (Y - E[Y])^T \left(\Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}\left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\right)(Y - E[Y]) \\ &= (X - E[X])^T \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} (X - E[X]) \\ &- 2(X - E[X])^T \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} \left(\Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y])\right) \\ &+ \left((Y - E[Y])^T\Sigma_{yy}^{-1}\Sigma_{yx}\right)\left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1}\left(\Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y])\right) \\ &+ (Y - E[Y])^T\Sigma_{yy}^{-1}(Y - E[Y]). \end{aligned}$$

We define the last term as $Q_2(Y) = (Y - E[Y])^T\Sigma_{yy}^{-1}(Y - E[Y])$ and observe that:

Lemma B.4. *For any vector u and v of dimension p and a symmetric matrix $A = A^T$ of dimension p-by-p*

$$u^T Au - 2u^T Av + v^T Av = (u - v)^T A(u - v).$$

Proof.

$$\begin{aligned} u^T Au - 2u^T Av + v^T Av &= u^T Au - u^T Av - u^T Av + v^T Av \\ &= u^T A(u - v) - (u - v)^T Av. \end{aligned}$$

Using the symmetry of $A = A^T$ and observing that $(u - v)^T Av$ is of dimension one, the last term can be written

$$(u - v)^T Av = (v^T A(u - v))^T = v^T A(u - v).$$

Finally, factoring by $A(u - v)$ on the right side completes the proof. $\square$

Therefore

$$Q = Q_1(X,Y) + Q_2(Y)$$

with

$$Q_1(X,Y) = (X-b)^T A^{-1}(X-b),$$
$$b = E[X] + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - E[Y]),$$
$$A = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}.$$

Now for C we will make use of the following lemma.

Lemma B.5. *The determinant of a partitioned symmetric matrix is given by*

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{22}|.|A_{11} - A_{12}A_{22}^{-1}A_{12}^T|.$$

Proof. Observe that

$$\begin{aligned}[A] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} &= \begin{bmatrix} A_{11} & 0 \\ A_{12}^T & I \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & A_{22} - A_{12}^T A_{11}^{-1} A_{12} \end{bmatrix} \\ &= \begin{bmatrix} I & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{12}^T & 0 \\ A_{22}^{-1}A_{21} & I \end{bmatrix}.\end{aligned}$$

From the following properties of matrix determinant

$$|AB| = |A|.|B|$$

and

$$\begin{vmatrix} B & 0 \\ C & D \end{vmatrix} = \begin{vmatrix} B & C \\ 0 & D \end{vmatrix} = |B|.|D|$$

the lemma is proved. □

Therefore

$$|\Sigma| = |\Sigma_{yy}|.|A|$$

and

$$\begin{aligned} C &= \frac{1}{(2\pi)^{(p+q)/2}|\Sigma_{yy}|^{1/2}|A|^{1/2}} \\ &= C_1.C_2 \end{aligned}$$

with

$$C_1 = \frac{1}{(2\pi)^{q/2}|A|^{1/2}},$$
$$C_2 = \frac{1}{(2\pi)^{p/2}|\Sigma_{yy}|^{1/2}}.$$

As

$$C_1 e^{-\frac{1}{2}Q_1(X,Y)} = N(X, b, A)$$

and

$$C_2 e^{-\frac{1}{2}Q_2(Y)} = N(E[Y], \Sigma_{yy})$$

we deduce that $E[X|Y] = b$ and $Var[X|Y] = A$ which complete the proof. $\square$

C Conditional Univariate Normal Distribution

Proposition C.1. *If x and y are two jointly distributed normal random variables*

$$P(x,y) = N\left(E\left(\begin{bmatrix}x\\y\end{bmatrix}\right), \begin{bmatrix}var(x) & cov(x,y)\\cov(x,y) & var(y)\end{bmatrix}\right)$$

then

$$P(x|y) = N\left(E[x|y], var(x|y)\right)$$

with

$$\begin{cases} E[x|y] & = E[x] + \frac{cov(x,y)}{var(y)}(y - E[y])) \\ \\ var(x|y) & = var(x) - \frac{cov(x,y)^2}{var(y)} \end{cases}$$

Proof. The proof is straightforward from proposition B.1 for $p = q = 1$. $\square$

Corollary C.2. *Defining the Pearson's correlation coefficient* $\rho = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$ *we obtain*

$$\begin{cases} E[x|y] & = E[x] + \rho(y - E[y])\frac{\sqrt{var(x)}}{\sqrt{var(y)}} \\ \\ var(x|y) & = var(x)(1-\rho^2) \end{cases}$$

Corollary C.3. *If x and y are two jointly distributed normal random variables of the form*

$$\begin{cases} y & = x + \sigma_y \epsilon \quad\quad \epsilon \sim N(0,1) \\ \\ x & = E[x] + \sigma_x \delta \quad \delta \sim N(0,1) \end{cases}$$

with ϵ and δ uncorrelated random variables. Then

$$\begin{cases} E[x|y] & = \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}E[x] + \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}y \\ \\ var(x|y) & = \frac{\sigma_x^2\sigma_y^2}{\sigma_y^2+\sigma_x^2} \end{cases}$$

Proof. We observe that

$$E[y] = E[x].$$

Using the *law of total variance* we get

$$Var(y) = E[Var(y|x)] + Var(E[y|x]) = \sigma_y^2 + \sigma_x^2.$$

And by definition of the covariance

$$\begin{aligned} Cov(x,y) &= E[(x - E[x])(y - E[y])] \\ &= E[(x - E[x])(y - E[x])] \\ &= E[(\sigma_x\delta)(\sigma_x\delta + \sigma_y\epsilon)] \\ &= \sigma_x^2 E[\delta^2] + \sigma_x\sigma_y E[\epsilon\delta] \\ &= \sigma_x^2. \end{aligned}$$

Therefore, using proposition C.1 gives

$$\begin{cases} E[x|y] & = E[x] + \frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}(y - E[x])) \\ \\ var(x|y) & = \sigma_x^2 - \frac{\sigma_x^4}{\sigma_y^2+\sigma_x^2} \end{cases}$$

which complet the proof. $\square$

D Pair Trading Example Simulation (Matlab)

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% beta(k+1)=beta(k)+w(k)          Model equation
% y(k)=x(k)*beta(k)+v(k)          Observation equation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function pairTrading

y=load('KO.txt');                 %Load Coca-Cola prices from file
y=flipud(y);

x=load('PEP.txt');                %Load Pepsico prices from file
x=flipud(x);

N=min(size(x), size(y));          %Sample size

Q=0.003;                          %Process noise covariance
R=1;                              %Measurement noise covariance
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Initial conditions
beta = ones(1, N);                %State
P = ones(1, N);                   %Estimation covariance

M = ones(1, N);                   %Prediction covariance
K = ones(1, N);                   %Kalmna Gain
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for k=1:N-1,
    M(k+1)=P(k)+Q;
    K(k+1)=M(k+1)*x(k+1)'*inv(x(k+1)*M(k+1)*x(k+1)'+R);
    P(k+1)=M(k+1)-K(k+1)*x(k+1)*M(k+1);
    beta(k+1)=beta(k)+K(k+1)*(y(k+1)-x(k)*beta(k));
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clf;                              %Clear current figure window
plot((1:N)', beta, 'k-');         %Plot the graph of beta
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fid = fopen('beta.txt', 'w');             %Open output file
                                          %with write permission
fprintf(fid, '%4.2f\r\n', beta);          %Write the data
fclose(fid);                              %Close output file
```

E Example From Classical Mechanics

E.1 Preliminaries

The *acceleration* is defined by

$$a(t) = \frac{dv}{dt}.$$

The *velocity* is defined by

$$v(t) = \frac{dx}{dt}$$

where $x(t)$ is the *position*.

For constant acceleration

$$\begin{aligned} \int_0^T a dt &= v(T) - v(0) \\ v(T) &= v(0) + aT, \end{aligned}$$

and

$$\begin{aligned} \int_0^T v(t) dt &= x(T) - x(0) \\ \int_0^T (v(0) + at) dt &= x(T) - x(0) \\ [v(0)t + \frac{1}{2}at^2]_0^T &= x(T) - x(0) \\ x(T) &= x(0) + v(0)T + \frac{1}{2}aT^2. \end{aligned}$$

E.2 A Guided Example

To illustrate the use of the Kalman filter in a straightforward and familiar situation we will develop an example inspired from Maybeck (1979).

Suppose you are lost in a large city so you do not know your true location x_t at any given time (we assume a one-dimensional location for simplicity). While looking for your way around, you are traveling at a constant velocity u_t defined according to the law of classical mechanics as the derivative of the position x_t with respect to time t

$$\frac{dx_t}{dt} = u_t.$$

To account for model error, first-order approximation, etc we introduce an additive Gaussian noise w_t such that

$$\frac{dx_t}{dt} = u_t + w_t.$$

Discretizing the SDE using the Euler-Maruyama scheme as show in section 2.5 we obtain

$$x_{k+1} = x_k + u_k \Delta t + \Delta w_k \tag{78}$$

where $\Delta w_k = w((k+1)\Delta t) - w(k\Delta t)$, $w \sim N(0, \sigma_w^2)$. This is our *state transition equation* where x is acting as the state variable.

According to our model of motion (78), the prior estimate at time $k+1$ given the filtration $\mathscr{Y}_k = \sigma\{y_k\}$, written $\hat{x}_{k+1}^-$, is

$$\hat{x}_{k+1}^- = E[x_{k+1}|\mathscr{Y}_k].$$

Therefore

$$\hat{x}_{k+1}^- = E[x_k + u_k \Delta t + \Delta w_k|\mathscr{Y}_k] = \hat{x_k} + u_k \Delta t \tag{79}$$

with variance

$$\begin{aligned}(\sigma_{k+1}^-)^2 &= E[(x_{k+1} - \hat{x}_{k+1}^-)^2|\mathscr{Y}_k] \\ &= E[(x_k + v_k\Delta t + \Delta w_k) - (\hat{x}_k + v_k \Delta t))^2|\mathscr{Y}_k] \\ &= E[(x_k - \hat{x}_k)^2 + 2(x_k - \hat{x}_k)\Delta w_k + (\Delta w_k)^2|\mathscr{Y}_k] \\ &= \sigma_x^2 + \sigma_w^2. \end{aligned} \tag{80}$$

(79) and (80) constitute our *update equations* (also called *prediction equations*).

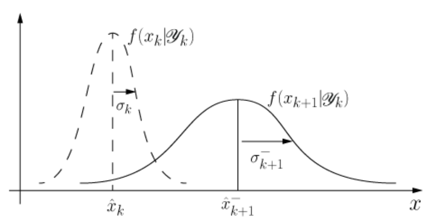

Figure 4: Density propagation.

You now observe your position at regular intervals Δt with a noisy sensor that we will model as follow.

$$y_k = x_k + v_k \tag{81}$$

where $v \sim N(0, \sigma_v^2)$. (81) is our *observation equation.* We will asume a unique source of measurement for simplicity. The random variables Δw_k and v_k represent respectively the model and measurement noise and are assumed to be independent of each other (both assumed to have a constant distribution for simplicity).

At this point, we have two Gaussian densities containing information available for estimating your position x_k. One encompassing all the information before the measurement $N(\hat{x}_{k+1}^{-}, (\sigma_{k+1}^{-})^2)$, and the other being the information provided by the measurement itself $N(x_{k+1}, \sigma_v^2)$. As shown in proposition C.1 from appendix C, the combined density is Gaussian with mean

$$\begin{aligned}\hat{x}_{k+1} &= \frac{(\sigma_{k+1}^{-})^2}{(\sigma_{k+1}^{-})^2 + \sigma_v^2}\hat{x}_{k+1}^{-} + \frac{\sigma_v^2}{(\sigma_{k+1}^{-})^2 + \sigma_v^2} y_{k+1} \\ &= \hat{x}_{k+1}^{-} + K_{k+1}(y_{k+1} - \hat{x}_{k+1}^{-}) \end{aligned} \tag{82}$$

and variance

$$\begin{aligned}\sigma_{k+1}^2 &= \frac{(\sigma_{k+1}^{-})^2 \sigma_v^2}{(\sigma_{k+1}^{-})^2 + \sigma_v^2} \\ &= (\sigma_{k+1}^{-})^2 - K_{k+1}(\sigma_{k+1}^{-})^2 \end{aligned} \tag{83}$$

with

$$K_{k+1} = \frac{(\sigma_{k+1}^{-})^2}{(\sigma_{k+1}^{-})^2 + \sigma_v^2}$$

called the *Kalman gain* and $(y_{k+1} - \hat{x}_{k+1}^{-})$ is called the *measurement innovation* or *residual*.

Remark E.1. The first term in equation (82) is the state estimate at time $k+1$. This would be the state estimate if we didn't have a measurement. The second term is called the correction term, and it represents how much to correct the propagated estimate due to the observation.

Remark E.2. Note that, from (83), $\sigma_{k+1} \leq \sigma_{k+1}^{-}$ and $\sigma_{k+1} \leq \sigma_v$, which means that the uncertainty in the estimate of position has been reduced by combining the information y_{k+1}.

Remark E.3. The state vector has two values at any given time k, that is the the prior value x_k^{-} (predicted value before the update) and the posterior value x_k (the corrected value after the update).

The *model equation* (78) along with the *observation equation* (81) constitute the *linear state space model*. The equations (82) and (83) characterising recursively the posterior density $N(\hat{x}_{k+1}, \sigma_{k+1}^2)$ defines the Kalman filter.

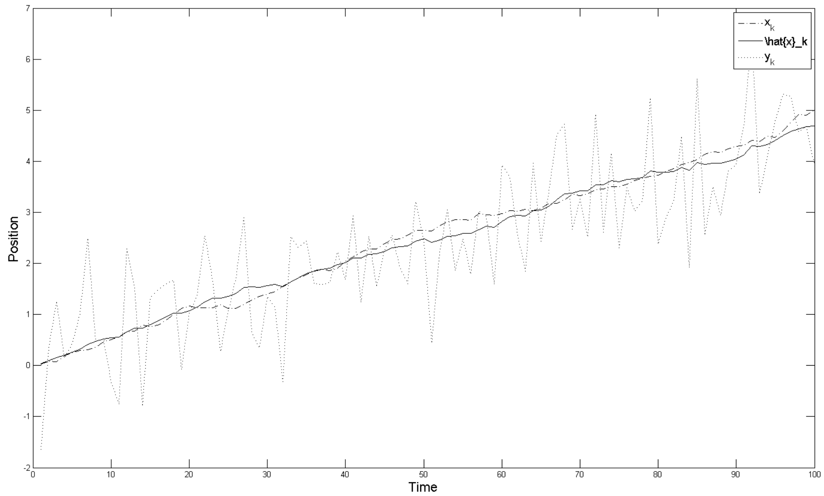

Figure 5: Estimated position

Parameters: $T = 1$, $dt = 1/100$; Volatilities: $\sigma_w = 0.3\%$, $\sigma_v = 1$; Initial conditions: $x_0 = 0$, $y_0 = 0$

		Definition
x_k		State (one-dimensional location).
u_k		Speed.
w_k		Process noise.
y_k		Measurement.
v_k		Measurement noise.
$\hat{x}_k^-$	$\hat{x}_{k\|k}$	Prior estimate of x_k, conditioned on all prior measurements except the one at time k.
$(\sigma_k^-)^2$		Prior variance of x_k.
$\hat{x}_k$	$\hat{x}_k^+$	Posterior estimate of x_k, conditioned on all available measurements at time k.
σ_k^2		Posterior variance of x_k.

Table 4: Notations.

E.3 Matlab Algorithm

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% N                         Time steps
% x(k)=A*x(k-1)+u+w(k)      Model equation
% y(k)=H*x(k)+v(k)          Observation equation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
function kalmanPosition(N)
randn('state',0);

A=1; H=1;
Q=0.003; %process noise variance
R=1;     %measurement noise variance
u=.05;   %velocity

X = []; % true position array
Xhat = []; % estimated position array
Y = []; % measured position array

%Initial conditions
x = 0; % initial state

xhat = x; % initial state estimate
p_pos=Q;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for k=1:N,
    %Model equation
    w=sqrt(Q)*randn; %process noise
    x=A*x+u+w;

    %Observation equation
    v=sqrt(R)*randn; %measurement noise
    y=H*x+v;

          %Update equation
    x_prior=A*xhat+u;
    % Innovation
    Inn = y - x_prior;
    %Covariance of Innovation
    p_prior = A*A*p_pos+Q;
    %Gain matrix
    K = H*p_prior/(H*H*p_prior+R);
    %State estimate
    xhat = A * x_prior + K * Inn;
    %Covariance of prediction error
    p_pos=p_prior*(1-H*K);
    %Save some parameters in vectors for plotting later
    X = [X; x];
    Xhat = [Xhat; xhat];
    Y = [Y; y];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
plot((1:N)',X,'k-.',k,Xhat,'k',k,Y,'k:');

hLegend = legend('x_k','\hat{x}_k', 'y_k');
set(hLegend,'FontSize',14);
xlabel('Time','FontSize',16)
ylabel('Position','FontSize',16)
```

F Drift Estimation Simulation

F.1 Main Function

```
function main(N)
randn('state',988);

K=1000;
T=1;
dt=T/500;

simulate(K,T);  %Simulation of monthly return over N months.
                %Also generate historical price files.

%%%%%%%%%%%%%%%%%%%% Load data from Files
S=load('IndexPrice.txt');
Q=load('IndexDividend.txt');
    q=Q(1);
F=load('ForwardPrice.txt');
C=load('OptionPrice.txt');
R=load('ThreeMonthRate.txt');
    r=R(1);
Lambda=load('Lambda.txt');

dim=[size(S,1), size(Q,1), size(F,1), size(C,1), size(R,1), size(Lambda,1)];
N=min(dim);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Computation of the Implied volatility
Sigma_implied=ImpliedVolatility(F, R, C, K, T, N);

if sum(isnan(Sigma_implied)) > 0
    error('Unexpected situation');
end

Sigma_c=ones(length(Sigma_implied),1);
for k=1:length(Sigma_implied)
    d1=(log(F(k)/K)+Sigma_implied(k)^2/2*(T-(k-1)*dt))/(Sigma_implied(k)*sqrt(T-(k-1)*dt));
    Sigma_c(k)=Sigma_implied(k)*F(k)/C(k)*exp(-r*(T-(k-1)*dt))*normpdf(d1);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%maximum likelihood estimation for kappa, lambda_bar and sigma_lambda
Theta=mle(, 'distribution', 'normal');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%estimate lambda(t) using kalman filtering
[Lambda_hat, Epsilon]=estimate(Sigma_implied, Sigma_c, Theta, S, r, C, F, q, dt, N);

Mu=ones(N,1);
for k=1:N
    Mu(k)=(r-q)+Sigma_implied(k)*Lambda_hat(k);
end

S_hat=ones(N,1);
S_hat(1)=1000;
for k=2:N
    S_hat(k)=S_hat(k-1)+Mu(k-1)*S_hat(k-1)*dt+Sigma_implied(k-1)*S_hat(k-1)*sqrt(dt)*Epsilon(k
end
```

```
Drift=ones(N,1);
Drift(1)=1000;
for k=2:N
    Drift(k)=Drift(k-1)+Mu(k)*S(k);
end

%%%%%%%%%%%%%%%%%%%%%%%%%% Save Estimation on files
%'\n' for all UNIX applications, Microsoft Word and WordPad
%'\r\n' for Microsoft Notepad

%Open the file with write permission
fid = fopen('lambda_hat.txt', 'w');
fprintf(fid, '%4.2f\r\n', Lambda_hat);
fclose(fid);
```

F.2 Price Simulation

The following algorithm is simulating the following processes

$$\begin{cases} \lambda_{k+1} & = & \lambda_k & + & \kappa(\bar{\lambda}-\lambda_k)\Delta t & + & \sigma_\lambda \Delta W_k \\ S_{k+1} & = & S_k & + & (r-q+\sigma\lambda_k)S_k\Delta t & + & \sigma S_k \Delta W_k \\ F_{k+1} & = & F_k & + & \sigma\lambda_k F_k \Delta t & + & \sigma F_k \Delta W_k \\ C_{k+1} & = & C_k & + & (r+\sigma_c\lambda_k)C_k\Delta t & + & \sigma_c C_k \Delta W_k \end{cases}$$

and save them on file.

```
function simulate(K,T)
N=500;dt=T/N;

sigma=.1; %Volatility for the index and forward

%%%%%%%%%%%%%%%%%%%%%%Brownian Motion
dW=sqrt(dt)*randn(N,1);
W =cumsum(dW);
%%%%%%%%%%%%%%%%%%%%%%Market Price of Risk
lambda_bar=1; %mean reversion level
sigma_lambda=1;
kappa=10;      %mean reversion rate

Lambda=zeros(N,1); %pre-allocation
Lambda(1)=1; %initial condition

for k=2:length(Lambda)
    %Ornstein Uhlenbeck process
    Lambda(k)=Lambda(k-1)+kappa*(lambda_bar-Lambda(k-1))*dt+sigma_lambda*dW(k-1);
end

%%%%%%%%%%%%%%%%%%%%%%Interest Rate
r=.02; %risk-free interest rate
R=zeros(N,1); %pre-allocation

for k=1:N
    R(k)=r;
end

%%%%%%%%%%%%%%%%%%%%%%Index Dividend
```

```
q=0;
Q=zeros(N,1); %pre-allocation

for k=1:N
    Q(k)=q;
end

%%%%%%%%%%%%%%%%%%%%Index Price
S=zeros(N,1); %pre-allocation
S(1)=1000; %initial condition

for k=2:length(S)
    S(k)=S(k-1)+(r-q+sigma*Lambda(k-1))*S(k-1)*dt+sigma*S(k-1)*dW(k-1);
end

%%%%%%%%%%%%%%%%%%%%Forward Price
F=zeros(N,1); %pre-allocation
F(1)=1000; %initial condition

for k=2:length(F)
    F(k)=F(k-1)+sigma*Lambda(k-1)*F(k-1)*dt+sigma*F(k-1)*dW(k-1);
end

%%%%%%%%%%%%%%%%%%%%Call Price
d1=zeros(N,1); %pre-allocation
sigma_c=zeros(N,1); %pre-allocation

C=zeros(N,1); %pre-allocation
C(1)=1000; %initial condition

for k=2:length(C)
    d1(k-1)=(log(F(k-1)/K)+sigma^2/2*(T-(k-1)*dt))/(sigma*sqrt(T-(k-1)*dt));
    sigma_c(k-1)=sigma*F(k-1)/C(k-1)*exp(-r*(T-(k-1)*dt))*normpdf(d1(k-1)); %option volatility

    C(k)=C(k-1)+(r+sigma_c(k-1)*Lambda(k-1))*C(k-1)*dt+sigma_c(k-1)*C(k-1)*dW(k-1);
end

%%%%%%%%%%%%%%%%%%%%Call Price
for k=1:length(C)
     C(k)=blsprice(F(k), K, r, T-(k-1)*dt, sigma);
end

%%%%%%%%%%%%%%%%%%%%Save Simulations on files
%'\n' for all UNIX applications, Microsoft Word and WordPad
%'\r\n' for Microsoft Notepad

fid = fopen('lambda.txt', 'w'); % open the file with write permission
fprintf(fid, '%4.2f\r\n', Lambda);
fclose(fid);

fid = fopen('ThreeMonthRate.txt', 'w'); %Open the file with write permission
fprintf(fid, '%4.2f\r\n', R);
fclose(fid);

fid = fopen('IndexPrice.txt', 'w'); %Open the file with write permission
fprintf(fid, '%4.2f\r\n', S);
fclose(fid);
```

```

fid = fopen('IndexDividend.txt', 'w'); %Open the file with write permission
fprintf(fid, '%4.2f\r\n', Q);
fclose(fid);

fid = fopen('ForwardPrice.txt', 'w'); %Open the file with write permission
fprintf(fid, '%4.2f\r\n', F);
fclose(fid);

fid = fopen('OptionPrice.txt', 'w'); %Open the file with write permission
fprintf(fid, '%4.2f\r\n', C);
fclose(fid);
```

F.3 Implied Volatility

```
function [Sigma_implied]=ImpliedVolatility(F, R, C, K, T, N)
Sigma_implied=zeros(N,1); %pre-allocation

dt=T/N;

for k=1:N
    Sigma_implied(k)=blsimpv(F(k), K, R(k), T-(k-1)*dt, C(k));
end
```

F.4 Kalman Filtering Estimation

```
function [X_hat, Epsilon]=estimate(sigma_implied, Sigma_c, Theta, S, r, C, F, q, dt, N)

% Theta(1)=kappa
% Theta(2)=lambda_bar
% Theta(3)=sigma_lambda

Q=[0.001; 0.001; 0.001]; %3x1
a=Theta(1)*Theta(2)*dt;     %1x1
B=1-Theta(1)*dt;            %1x1
R=Theta(3)*sqrt(dt); %1x1

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

X = []; %True state array
X_hat = []; %Estimated state array
Y = []; %Measured state array

Epsilon = []; %Measured state array

%Initial conditions
x = 1; % initial state

x_hat = x; %Initial state estimate
p_pos=dt;

for k=1:N
    H=[sigma_implied(k); sigma_implied(k); Sigma_c(k)]*sqrt(dt);
    d=[(r-q-sigma_implied(k)^2/2); -sigma_implied(k)^2/2; (r-Sigma_c(k)^2/2)]*dt;
    D=H*sqrt(dt);
```

```
    G=H';
    %Model equation
    epsilon=randn; %process noise
    x=a+B*x+R*epsilon;
    %Observation equation
    eta=randn;
    v=H*epsilon+Q*eta; %measurement noise
    y=d+D*x+v;
    %
    x_prior=a+B*x_hat;
    % Innovation
    Inn = y - D*x_prior-d;
    % Covariance of Innovation
    p_prior = B*p_pos*B'+R*R';
    % Kalman Gain
    F=D*p_prior*D'+D*R*G+G'*R'*D'+H*H'+Q*Q';
    K = (p_pos*D'+R*G)*pinv(F); %pseudoinverse
    % State estimate
    x_hat = x_prior + K * Inn;
    % Covariance of prediction error
    p_pos=(1-K*D)*p_prior+K*G'*R;

    % Save some parameters in vectors for plotting later
    X = [X; x];
    Y = [Y; y'];
    X_hat = [X_hat; x_hat];
    Epsilon=[Epsilon; epsilon];
end
```

G References

1. Finite Difference Methods, course notes, M. Davis.

2. Mathematical Option pricing, course notes, C. Barnet.

3. Estimation of Diffusion Parameters by Nonparametric Drift Function Model, Isao Shoji.

4. Nonparametric state estimation of diffusion process, Isao Shoji.

5. Approximating a Diffusion by a Hidden Markov Model, I. Kontoyiannis, S.P. Meyn, 2009.

6. Bond pricing in a hidden Markov model of the short rate, Camilla Landen, 2000.

7. Filtering Equity Risk Premia from Derivative Prices, Ramaprasad Bhar , Carl Chiarella and Wolfgang Runggaldier, 2001.

8. On estimation of volatility for short time series of stock prices, Nikolai Dokuchaev, 2010.

9. Recursive Bayesian Estimation: Navigation and Tracking Applications, Niclas Bergman, 1999.

10. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, Desmond J. Higham, 2001.

11. Market microstructure theory, Maureen O'Hara, Wiley-Blackwell, 1997.

12. Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation, Roger W. Lee, 2002.

13. Martingale methods in financial modeling, Marek Musiela and Marek Rutkowski.

14. The calculation of implied variances from the Black-Scholes model, S. Manaster, G. Koehler, 1982.

15. Bayesian Statistics for evaluation Research - An Introduction, William E. Pollard, 1986.

16. Nonparametric Estimation of Trends in Linear Stochastic Systems, Ian W. McKeague, Tiziano Tofoni, 1989.

17. Stochastic processes and filtering theory, Jazwinski, Andrew H., 1970.

18. An Introduction To Nonlinear Filtering, M. H. Davis, Steven I. Marcus, .

19. Calibration in a Bayesian modelling framework, Michiel J.W. Jansen, Thomas J. Hagenaars, 2006.

20. Bayesian Filtering: From Kalman Filters to Particle Filters and Beyond, Zhe Chen, 2003.

21. Maximum Likelihood Estimates Of Linear Dynamic Systems, H. E. Rauch, 1965.

22. A Bayesian Approach to Problems in Stochastic Estimation and Control, Y.C. Ho and R.C. K. Lee, 1964.

23. An application of the Kalman Filter: Pairs Trading (lecture notes), Ajay Jasra, 2011.

24. Statistical Arbitrage and High-Frequency Data with Application to Eurostoxx 50 Equities, Dunis, Giorgioni, Laws, Rudy, 2010.

25. Pairs Trading, R. J. Helliott, J. Van Der Hoek, W. P. Malcolm, 2005.

26. Adaptive Kalman Filter for Noise Identification, M.Oussalah, J. De Schutter, 2003.

27. Analysis of Kalman Filter with Correlated Noises under Different Dependence, L. Ma, H. Wang, J. Chen, .

28. Optimal State Estimation, D. Simon, 2006.

29. Optimal Filtering, B.D.O Anderson, J.B.Moore, 1979.

30. An Ornstein-Uhlenbeck Framework for Pairs Trading, S. Rampertshammer, 2007.

31. Understanding the Kalman Filter, R. J. Meinhold, N. D. Singpurwall, 1983.

32. Admissible Strategies in Semimartingale Portfolio Selection, S. Biagini, A. Cerny, 2011.

33. Statistical Inference from Sampled Data for Stochastic Processes, B. L. S. Prakasa Rao, 1988.

34. Stochastic models, estimation, and control, Volume 1, Peter S. Maybeck,

1979.