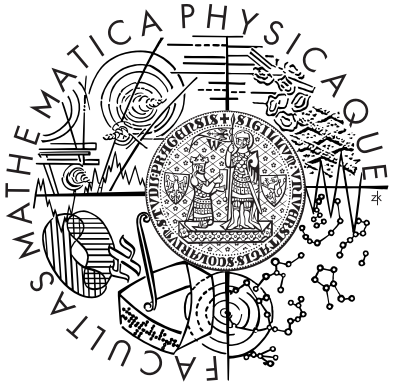
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Marek Svojík

Aplikace Kalmanova filtru

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: *Mgr. Zdeněk Hlávka, Ph.D.*

Studijní program: *Matematika*
Studijní obor: *Pravděpodobnost, matematická statistika a ekonometrie*
Studijní plán: *Matematická statistika*

Poděkování: Rád bych na tomto místě poděkoval Mgr. Zdeňku Hlávkovi, Ph.D. za cenné rady, náměty a připomínky, které mi pomohly při zpracování této diplomové práce. Poděkování patří také FEDC, SFB a Humboldt-Universität zu Berlin za poskytnutí dat.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne Marek Svojík

Contents

Název práce: *Aplikace Kalmanova filtru*
Autor: *Marek Svojík*
Katedra (ústav): *Katedra pravděpodobnosti a matematické statistiky*
Vedoucí diplomové práce: *Mgr. Zdeněk Hlávka, Ph.D.*
e-mail vedoucího: *zdenek.hlavka@mff.cuni.cz*

Abstrakt: *Cílem této diplomové práce je seznámit čtenáře s možností využití Kalmanova filtru v některých ekonomických úlohách. Kalmanův filtr (KF) je algoritmus používaný pro odhad nepozorovatelné složky nějakého stavu. V této práci se však speciálně zaměřím na možnosti použití KF pro odhad rizikově neutrální hustoty z cen kupních (CALL) opcí. V takovém případě je možné předpokládat nelineární vztah mezi stavovými a pozorovanými proměnnými a problém bude zapotřebí linearizovat pomocí Taylorova vzorce. V první kapitole představím podrobněji vlasní KF pro lineární stavový případ. Ve druhé kapitole se budeme zabývat použitím KF v případě CALL opcí. Třetí kapitolu věnuji rozšířené verzi KF a její aplikaci na nelineární stavový případ, kde bude použita linearizace Taylorovým vzorcem. V kapitole čtvrté se společně pokusíme odhadnout rizikově neutrální hustotu z cen CALL opcí. V poslední, páté, kapitole – doplněné o výstupy ze statistického prostředí* `R` *– vyzkoušíme algoritmus na realných datech. Závěrem pak shrneme všechny důležité výsledky této práce.*
Klíčová slova: *stavový model, Kalmanův filtr, nelinearní stavový model, rozšířený Kalmanův filtr, CALL opce*

Title: *Application of Kalman Filtering*
Author: *Marek Svojík*
Department: *Department of Probability and Mathematical Statistics*
Supervisor: *Mgr. Zdeněk Hlávka, Ph.D.*
Supervisor's e-mail address: *zdenek.hlavka@mff.cuni.cz*

Abstract: *The aim of this work is to discuss the use of the Kalman filter in some economical problems. Generally taken, the Kalman filter is a mathematical method (an algorithm) used for estimation of the non-observable component of a state. Especially, this approach will be applied to estimate the risk-neutral state price density of CALL options. In such case a non-linear relation between state and observed variables may be assumed, and the problem has to be linearized by Taylor expansion. In detail, the main Kalman filtering in the simple linear case will be presented in the first chapter. In the second chapter, you can find some application of that Kalman filtering in case of CALL options. The study of the extended Kalman filter and its application in case of a non-linear state model and the use of the Taylor expansion can be found in Chapter 3. In the fourth chapter, we will be talking about estimating the risk-neutral price density of a CALL option. The corresponding outputs from the program* `R` *and the most important results of this work are summarized in the last Chapter 5.*
Keywords: *state model, Kalman filter, non-linear state model, extended Kalman filter, CALL option*

Chapter 1

Space state models and the Kalman filter

1.1 Foreword and some motivation

The aim of this work is to discuss use of the Kalman filter in some economical problems. Especially, this approach will be applied to estimate the risk-neutral state price density of CALL options. In such case a non-linear relation between state and observed variables may be assumed, and the problem has to be linearized by the Taylor expansion.

In detail, the main filtering in simple linear state case will be presented in Chapter 1. The main theoretical "equipment" (mathematical definitions, theorems, etc.) was largely taken from Harvey [13], Kellerhals [15] and Welch and Bishop [21]. In the second chapter, we firstly try to build a simple linear model in case of CALL options. Later, you can find some application of the Kalman filtering to estimate the risk-neutral price density of a CALL option in Chapter 2, which is the academic value of this thesis. The study of the extended version of the Kalman filter and its application in case of a non-linear state model and the use of the Taylor expansion can be found in Chapter 3. The main mathematical equipment will also be taken from Harvey [13], Kellerhals [15] and Härdle et al. [8] in this chapter. In the first instance, we again build a model for CALL options and later we will be talking about estimating of the risk-neutral price density of a CALL option in the Chapter 4. This should be the contribution to science in the field of option pricing. In the last Chapter 5 we try to test out the designed main algorithm with real data. The most important results of this work are summarized in this chapter, which is completed by some outputs from the statistical program `R`.

Generally taken, the Kalman filter is a mathematical method (an algorithm) used to estimate the non-measurable component of a state. Widely, it faces the parallel estimation of the state vector of a non-measurable state variable depending on new observations of a related measurable variable.

It is a specific application of a particular conception of relationships between the output gap and the key (eg. physical, macro-economic, etc.) pointers of the past (already observed) data. Originally, this algorithm was designed to observe a path (trajectory) of any object, using the idea and procedure for the prediction of the object's next location. As a tool of prediction, the Kalman filter is very useful because of its recursive structure.

Mostly it features data processing (a large amount of data) acquired by any measurement or observation. Scilicet, in the first phase of this processing such data it is necessary to filter out – by mathematical methods – the "inaccuracies", caused by outer impacts during their obtaining and collection. Therefore the name "filter".

In addition to that, the Kalman filter may be used to calculate the estimation of the state, if the expected distribution of the measurement errors and the estimation of the state correspond with the normal distribution, i.e. all random errors may be represented only by the mean μ and variance σ^2.

The Kalman filter is an iterative algorithm for evaluating the estimation, which minimizes the mean square error. It has also been proved, that the Kalman filter is the best such linear algorithm (i.e. non-linear algorithms can be better). Thus, the algorithm is called a filter, because it really does filter out the input data (measurements), and therefore it removes the noise.

Let's imagine the following situation: we have a little black box which its *inner state* is being characterized by just one number. This number is **unknown**. The objective of our work will be now in every next step to estimate the value of that number as exact as possible. The change of the inner state can be described by the following *state equation*:

$$x_{t+1} = f(x_t) + \eta_t,$$

where $\eta_t \backsim \mathrm{N}(0, Q)$ is noise (error of the state) with the zero mean and variance Q. The unknown inner state x_t can – for simplicity – be imagined like a position on the axis x in step t. A measurement whose character is described as follows, is available as well:

$$y_t = h(x_t) + \epsilon_t,$$

where $\epsilon_t \backsim \mathrm{N}(0, H)$ are errors of the measurement with zero mean and variance H. If just the own state x_t is measured (e.g. the position on the axis x) in the simplest case, the function h equals the identity id. We are even able to measure a (non-trivial) transformation of the searched state in more complicated cases – and this transformation is exactly characterized by the function h as above. This is then a non-linear problem and we have to use another algorithm which reflects such non-linear dependence between the state and observed variables, e.g. the extended version of the Kalman filter will be used.

Prices (values) of an option at a specific time can be estimated (in the sense to determine their value the best way) in like manner as just presented.

1.2 Introduction

The Kalman filter was invented in the early sixties. The *Kalman filter framework* was originally developed by Kalman [14] and became a part of the astronautical guidance system of the Apollo project. However, it has also been used for applications in other scientific fields.

Recently (and this is what we will be discussing), the Kalman filter approach has been discovered as an estimation tool in continuous time finance – only in the last decade, marked by Harvey [13], Kalman filters have become important econometric tools for financial and economic estimation problems.

Firstly, the following must be said: the Kalman filter is simply an *optimal recursive data processing algorithm.*

- To the word *optimal*: Because there are many ways to define "optimal", a criterion of that optimality has to be chosen. It can be shown that, under the specific assumption, the Kalman filter is optimal with respect to virtually any criterion that makes sense. One aspect of that optimality is that the algorithm incorporates all information it can be provided with. It processes all measurements to estimate the current value of the variables of the interest, with use of:

  1. knowledge of the system and measurement device dynamics,

  2. the statistical description of the system noises, measurement errors, uncertainty in the dynamics models, and

  3. any available information about initial conditions of the variables of interest.

  Example 1.2.1. *To estimate the velocity of an aircraft, a Doppler radar can be used, or one could use the velocity indications of an inertial navigation system, or the pilot and statistic pressure and relative wind information in the air date system. Rather than ignore any of these outputs, the Kalman filter could be built to combine all of this data and knowledge to generate an overall best estimate of velocity.*

- To the word *recursive*: unlike certain data processing concepts, the Kalman filter does not require all previous data to be kept in storage and reprocessed every time a new measurement is taken. This is very useful for the filter's practical implementation.

Figure 1.1 depicts a typical situation in which the Kalman filter could be used: a system is driven by known controls and measuring devices provide the value of certain pertinent quantities. Knowledge of these system inputs and outputs is all that is explicitly available from the physical system for estimation purposes.

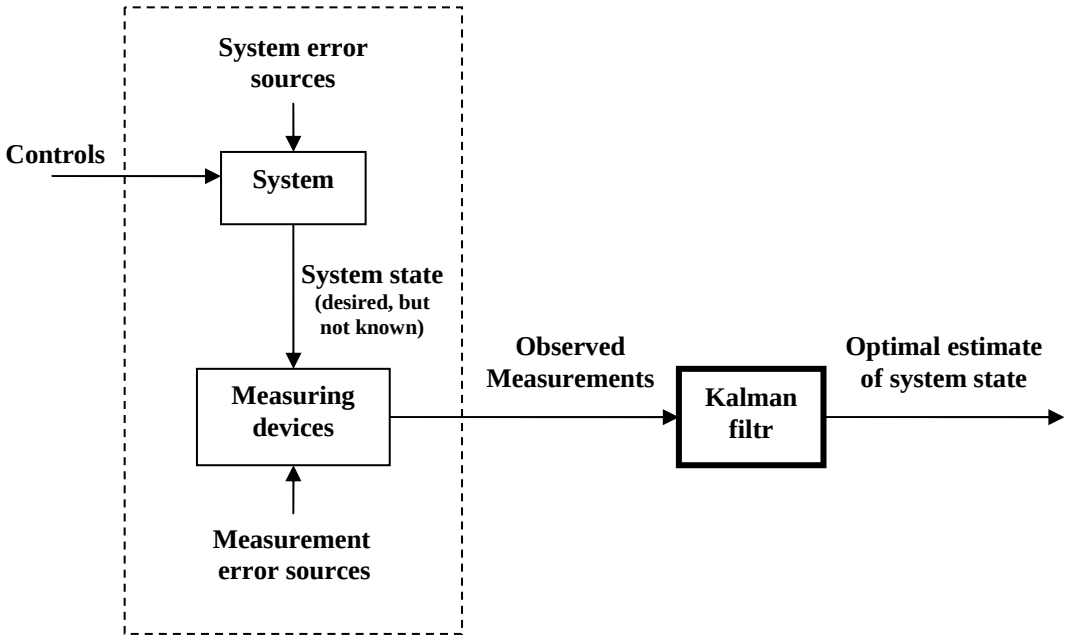


Figure 1.1: Typical Kalman filter application.

The word "need" for a filter will be now explained: often the variables of interest, some finite number of parameters to describe the *state* of the system, cannot be measured directly – some values from the available data must be generated.

1.3 State space

1.3.1 Probability space, filtration and stochastic processes

To describe the main Kalman filter we first take a look at stochastic processes in general defined on a probability space which represents the uncertainty of the stochastic economy. Therefore, the probability space, the filtration, stochastic processes and their classification (in the sense of Kellerhals [15]) may be defined now.

Definiton 1.3.1 (Probability Space)**.** *A* ***probability space*** $(\Omega, \mathcal{F}, \mathrm{P})$ *is formally a measure space. The first item,* Ω*, is a nonempty set with* $\mathrm{P}(\Omega) = 1$ *(a probability space is a measure space with total measure one). The second item,* $\mathcal{F}$*, is the* σ*-field of subsets of* Ω*, here representing the information available at time* t *The third item,* P*, is the probability measure.*

Definiton 1.3.2 (Filtration)**.** *A* ***filtration****,* $\mathbf{F} = \{\mathcal{F}_t\}_{t\in T}$*, is a non-decreasing family of sub-*σ*-fields of* $\mathcal{F}_t$*:* $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ *for* $0 \leq s < t < \infty$*, where* $T = [0, \infty)$*.*

Remark 1.3.3. *The filtration* $\mathbf{F}$ *represents the information flow evolving over time* t *and accruing to all agents in the economy.*

Definiton 1.3.4 (Stochastic Process)**.** *A scalar (n-vector)* ***stochastic process****, denoted by* $\{\mathbf{X}_t\}_{t\in T}$*, is a family of random variables (n-vectors) indexed by the parameter set* T*, where the parameter* t *will refer to time. That process is defined on a filtered probability space* $(\Omega, \mathcal{F}, \mathrm{P}, \mathbf{F})$ *and has its values in* $\mathbb{R}^n$*.*

We say that the process is ***adapted*** *to the filtration* $\mathbf{F}$ *if* $\mathbf{X}_t$ *is* $\mathcal{F}_t$*-measurable for each time* t*.*

Further, if a filtration is generated by a stochastic process, i.e. $\mathcal{F}_t = \sigma(\mathbf{X}_s; 0 \leqq s \leqq t)$*, we call* $\mathcal{F}_t$ *the* ***natural filtration*** *of the process* $\{\mathbf{X}_t\}_{t\in T}$*. Thus, a process is always adapted to its natural filtration.*

Definiton 1.3.5 (Classification of Stochastic Processes)**.** *If the random variables (vectors)* $\mathbf{X}_t$ *are discrete, we say that the stochastic process has a* ***discrete state space****. If they are continuous, the process is said to have a* ***continuous state space****. The parameter set may also be discrete (for example,* $T = \{1, 2, \ldots, n\}$*, or* $T = \{1, 2, 3, \ldots\}$*) or continuous (for example,* $T = [0, 1]$*, or* $T = [0, \infty]$*). If the parameter set* T *is discrete, the stochastic process is a* ***discrete parameter process****. If it is continuous, we say that the stochastic process is a* ***continuous parameter process****.*

A short overview of the classification of stochastic processes is shown in the following Table 1.1.

Classifying Characteristics		Parameter Set $\mathcal{T}$	
		Discrete	Continuous
State Space X	Discrete	Discrete Parameter Chain	Continuous Parameter Chain
	Continuous	Random Sequence	Stochastic Process

Table 1.1: Classification of Stochastic Processes.

1.3.2 State space model and its definition

The state space form is a very useful tool to handling a wide range of time series models and to write a dynamic model for a further analysis with the Kalman filter. Once a model has been put in state space form, the Kalman filter may be applied. That state space form is based on two important sets of system equations:

- the *measurement equation* which relates the state variables to the variables which can only be observed with measurement noise,
- the *transition* or *process equation*, describes the dynamic evolution of the unobservable variables.

The general state space form applies to a multivariate time series, $\mathbf{y}_t$, containing g observed elements, thus $\mathbf{y}_t$ is a one-dimensional vector. The vector of unknown parameters, on which the system matrices and error term specifications of the state space form depend on, will be denoted by $\boldsymbol{\psi}$. We will treat the variances of the measurement errors $\boldsymbol{\epsilon}_t$ $\big($defined below in (1.1)$\big)$ as a independent, separate state variable – not as a part of the vector $\boldsymbol{\psi}$ – as we will see further down in this thesis on the Kalman filter recursions. The unobservable variables are related to an $(k \times 1)$ vector, $\boldsymbol{\xi}_t$, known as the *state vector*, via the following *measurement equation*.

Definiton 1.3.6 (Measurement Equation)**.** *For the functional relationship of the measurable observations $\mathbf{y}_t$ with the possibly unobservable state vector $\boldsymbol{\xi}_t$, we define the* ***measurement equation*** *given by the linear form*

$$\mathbf{y}_t = \mathbf{a}_t(\boldsymbol{\psi}) + \mathbf{B}_t(\boldsymbol{\psi})\boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T, \tag{1.1}$$

with the parameters $\boldsymbol{\psi}$, an additive component $\mathbf{a}_t(\boldsymbol{\psi})$ as a $(g \times 1)$ vector, a known multiplicative $(g \times k)$ matrix $\mathbf{B}_t(\boldsymbol{\psi})$, and a $(g \times 1)$ noise term $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ of serially uncorrelated disturbances with mean zero (i.e. $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ are centered) and unknown $(g \times g)$ covariance matrix $\mathbf{H}_t(\boldsymbol{\psi})$ (i.e. the variances $\mathrm{Var}[\boldsymbol{\epsilon}_t(\boldsymbol{\psi})] = \mathrm{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_s^\top]$ are unknown for each $s = t$).

Further, we assume a normal distribution for the error term $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$, with

$$\begin{aligned} \mathrm{E}[\boldsymbol{\epsilon}_t] &= \mathbf{0}^\top, \quad \text{for all } t. \\ \mathrm{Var}[\boldsymbol{\epsilon}_t] = \mathrm{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s^\top] &= \begin{cases} \mathbf{H}_t(\boldsymbol{\psi}) & \text{for } s = t, \\ \mathbf{0} & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore we assume that

$$\boldsymbol{\epsilon}_t(\boldsymbol{\psi}) \sim \mathrm{N}_g\big[\mathbf{0}, \mathbf{H}_t(\boldsymbol{\psi})\big].$$

Example 1.3.7. *In a univariate model ($g = 1$) the measurement equation is written as*

$$y_t = a_t(\boldsymbol{\psi}) + \mathbf{b}_t^\top(\boldsymbol{\psi})\boldsymbol{\xi}_t + \epsilon_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T,$$

where

$$\begin{aligned} \mathrm{E}[\epsilon_t] &= 0 \quad \textit{for all } t, \\ \mathrm{Var}[\epsilon_t] &= h_t(\boldsymbol{\psi}). \end{aligned}$$

In general the elements of $\boldsymbol{\xi}_t$ are not observable. However, they are known to be generated by a first-order Markov process as we can see in the following definition.

Definiton 1.3.8 (Transition Equation)**.** *For the system, that describes the evolution of the $(k \times 1)$ state variables $\boldsymbol{\xi}_t$ over time, we assume the linear* ***transition equation***

$$\boldsymbol{\xi}_t = \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_{t-1} + \mathbf{c}_t(\boldsymbol{\psi}) + \boldsymbol{\eta}_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T, \tag{1.2}$$

with the known ***transition*** *$(k \times k)$* ***matrix*** *$\boldsymbol{\Phi}_t(\boldsymbol{\psi})$, an $(k \times 1)$ additive component $\mathbf{c}_t(\boldsymbol{\psi})$ and a $(k \times 1)$ Gaussian noise term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$, which is centered and has a known $(k \times k)$ covariance matrix $\mathrm{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t(\boldsymbol{\psi})$.*

It can also be written

$$\begin{aligned} \mathrm{E}[\boldsymbol{\eta}_t] &= \mathbf{0}^\top \quad \text{for all } t, \\ \mathrm{Var}[\boldsymbol{\eta}_t] = \mathrm{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_s^\top] &= \begin{cases} \mathbf{Q}_t(\boldsymbol{\psi}) & \text{for } s = t, \\ \mathbf{0} & \text{otherwise.} \end{cases} \end{aligned}$$

Thus we assume that

$$\boldsymbol{\eta}_t(\boldsymbol{\psi}) \sim \mathrm{N}_k\big[\mathbf{0}, \mathbf{Q}_t(\boldsymbol{\psi})\big].$$

Remark 1.3.9. *A state space model given by (1.1) and (1.2) includes plenty of models of one-dimensional time series, for example structural time series or* ARMA *models, and allows their further unification and expansion.*

The specification of the state space model is completed by two further assumptions.

- First, we assume a normally distributed *initial state vector* $\boldsymbol{\xi}_0$ (which describes the initial state of the model) to have a mean of $\boldsymbol{\xi}_{0|0}$ and a covariance matrix $\boldsymbol{\Sigma}_{0|0}$, i.e.

$$\begin{aligned} \mathrm{E}[\boldsymbol{\xi}_0] &= \boldsymbol{\xi}_{0|0}, \\ \mathrm{Cov}[\boldsymbol{\xi}_0] &= \boldsymbol{\Sigma}_{0|0}. \end{aligned}$$

- Further we assume independence between the error terms $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ and $\boldsymbol{\eta}_t(\boldsymbol{\psi})$. It is further assumed, that they are uncorrelated with the initial state vector $\boldsymbol{\xi}_0$. Thus, it can be written

$$\begin{aligned} \mathrm{E}[\boldsymbol{\epsilon}_t \boldsymbol{\eta}_s^\top] &= \mathbf{0} \quad \text{for all } s,t, \\ \mathrm{E}[\boldsymbol{\xi}_0 \boldsymbol{\epsilon}_t^\top] &= \mathbf{0} \quad \text{for all } t, \\ \mathrm{E}[\boldsymbol{\xi}_0 \boldsymbol{\eta}_t^\top] &= \mathbf{0} \quad \text{for all } t. \end{aligned} \tag{1.3}$$

Remark 1.3.10. *The assumption in (1.3) may be relaxed. More can be found, for example, in Harvey [13].*

Example 1.3.11. *Discuss the following 2-dimensional state model:*

$$\begin{aligned} \mathbf{y}_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t, \\ \boldsymbol{\xi}_t &= \begin{bmatrix} 0.8 & 0.2 \\ 1 & 0 \end{bmatrix} \boldsymbol{\xi}_{t-1} + \begin{bmatrix} \eta_t \\ 0 \end{bmatrix}. \end{aligned}$$

Please note that this model is a time series model corresponding to an AR(2) *process with additive Gaussian noise.*

Some remarks

1. The model order is given by the dimension of $\boldsymbol{\xi}_t$, i.e. it is of k-th order.

2. Some literature about the Kalman filter (Harvey [13], for example) assumes existence of a multiplicative $(k \times k)$ matrix $\mathbf{R}_t(\boldsymbol{\psi})$ in front of the error term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$, so the more general transition equation can be written as:

$$\boldsymbol{\xi}_t = \mathbf{c}_t(\boldsymbol{\psi}) + \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_{t-1} + \mathbf{R}_t(\boldsymbol{\psi})\boldsymbol{\eta}_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T. \tag{1.4}$$

   This inclusion of that matrix $\mathbf{R}_t(\boldsymbol{\psi})$ in front of the error term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ is arbitrary. The noise term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ could always be redefined so as to have a covariance matrix

$$\mathbf{R}_t(\boldsymbol{\psi})\mathbf{Q}_t(\boldsymbol{\psi})\mathbf{R}_t^{\top}(\boldsymbol{\psi}).$$

   Nevertheless, the representation in (1.4) is often more natural when $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ is identified with a particular set of disturbances in the model.

3. The transition equation in (1.2) is sometimes shifted forward one period so as to become

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_t + \mathbf{c}_t(\boldsymbol{\psi}) + \boldsymbol{\eta}_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T,$$

   or – more generally – to become

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_t + \mathbf{c}_t(\boldsymbol{\psi}) + \mathbf{R}_t(\boldsymbol{\psi})\boldsymbol{\eta}_t(\boldsymbol{\psi}), \qquad t = 1, \ldots, T. \tag{1.5}$$

   From the practical point of view, it makes very little differences whether (1.2) or (1.5) is used in conjunction with the measurement equation (1.1). However, it is necessary to be careful in certain circumstances, for example when dealing with correlated measurement and transition equation noise (see Harvey [13]). Throughout this thesis the emphasis will be on (1.2), though other works (Harvey [13], for example) prefer (1.5).

4. These two defined equations, the measurement equation (1.1) and the transition equation (1.2), constitute a *linear first-order Gauss-Markov state space representation* for the dynamic behavior of the observable variables $\mathbf{y}_t$. They allow a particularly tractable solution for the further analysis of the Kalman filtering.

Generally, techniques to construct a model from measured data typically contain two steps:

1. **Theory** and the *modeling step*: a family of candidate models is defined. In practice (financial applications) this step heavily draws on economic theory and mathematics, especially the field of stochastic theory.
2. **Practice** and the *parameter estimation*: we look for the particular member of this family that optimally describes the information content revealed by the data. In our applications this step is in fact a parameter estimation problem in which we maximize a likelihood function based on the prediction error decomposition. See more in Kellerhals [15].

Thus we look for a model identification as a link between the mathematical model world (theoretical step 1) and the real world of data (practical step 2).

1.3.3 System matrices, a time-invariant state model and hyperparameters

- **System matrices**

  The matrices $\mathbf{B}_t(\boldsymbol{\psi})$, $\mathbf{a}_t(\boldsymbol{\psi})$ and $\mathbf{H}_t(\boldsymbol{\psi})$ in the measurement equation (1.1) and the matrices $\boldsymbol{\Phi}_t(\boldsymbol{\psi})$, $\mathbf{c}_t(\boldsymbol{\psi})$, $\mathbf{Q}_t(\boldsymbol{\psi})$ $\big($or $\mathbf{R}_t(\boldsymbol{\psi})$ in (1.4), eventually$\big)$ in the transition equation (1.2) will be referred to as the *system matrices*. It will be further assumed that they are non-stochastic. However, they may change with time. As a result the system is linear and for any value t, the value $\mathbf{y}_t$ can be expressed as a linear combination of present and past $\boldsymbol{\epsilon}$'s and $\boldsymbol{\eta}$'s and the initial state vector, $\boldsymbol{\xi}_0$. The situation in which the system is non-linear will be discussed later in Chapters 3 and 4. More about the system matrices can be found in Harvey [13].

- **A time-invariant state model**

  If the system matrices $\mathbf{B}_t(\boldsymbol{\psi})$, $\mathbf{a}_t(\boldsymbol{\psi})$, $\mathbf{H}_t(\boldsymbol{\psi})$ in (1.1), and $\boldsymbol{\Phi}_t(\boldsymbol{\psi})$, $\mathbf{c}_t(\boldsymbol{\psi})$, $\mathbf{Q}_t(\boldsymbol{\psi})$ and $\mathbf{R}_t(\boldsymbol{\psi})$ in (1.2) and (1.4) does not change over time, the state space model is said to be *time-invariant* or *time-homogenous*. We also can say, the state model (1.1) and (1.2) is *time-invariant*, if the distributions of the random errors $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ and $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ do not change over time, i.e.

$$\begin{aligned} \operatorname{Var}\big[\boldsymbol{\epsilon}_t(\boldsymbol{\psi})\big] &= \mathbf{H}_t(\boldsymbol{\psi}) = \mathbf{H}(\boldsymbol{\psi}) \quad \text{for all } t, \text{ and} \\ \operatorname{Var}\big[\boldsymbol{\eta}_t(\boldsymbol{\psi})\big] &= \mathbf{Q}_t(\boldsymbol{\psi}) = \mathbf{Q}(\boldsymbol{\psi}) \quad \text{for all } t. \end{aligned}$$

  Later in this thesis, if we talk about time-invariant state space models, the first definition of a time-invariant state space model is considered, i.e. the system matrices in (1.1) and (1.2) – or in (1.4), respectively – does not change over time.

Although the class of time-invariant model is much broader than the class of stationary models, many time-invariant models have a stationary form which can be obtained by applying a transformation such as differencing.

- **Hyperparameter $\boldsymbol{\psi}$**

The following Example 1.3.12 leads us to the mathematical problem of estimation of the hyperparameter $\boldsymbol{\psi}$.

Example 1.3.12. *The random walk plus noise model, defined as follows:*

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \\ \mu_t &= \mu_{t-1} + \eta_t, \text{ where} \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(\epsilon_t) &= \sigma_\epsilon^2, \\ \mathrm{Var}(\eta_t) &= \sigma_\eta^2, \end{aligned}$$

is a time-invariant state space model with the state μ_t. It can be shown (see in Harvey [13]) that $\Delta \mathbf{y}_t$ is stationary.

It has already been discussed that the system matrices $\mathbf{B}_t(\boldsymbol{\psi})$, $\mathbf{H}_t(\boldsymbol{\psi})$, $\boldsymbol{\Phi}_t(\boldsymbol{\psi})$, $\mathbf{Q}_t(\boldsymbol{\psi})$ in both definition (1.1) and (1.2) – or even $\mathbf{R}_t(\boldsymbol{\psi})$ in (1.4) – may depend on a set of unknown parameters (denoted by $\boldsymbol{\psi}$). Often, one of the main statistical tasks is to estimate these parameters. For example in the random walk plus noise model $\big($Example 1.3.12$\big)$ the parameters $\sigma_{\boldsymbol{\epsilon}}^2$ and σ_η^2 will – in general – be unknown.

These unknown parameters will be denoted by an $(n \times 1)$ vector $\boldsymbol{\psi}$ and referred to as hyperparameters in order to distinguish them from parameters which may enter into the model via $\mathbf{a}_t(\boldsymbol{\psi})$ in the measurement equation and $\mathbf{c}_t(\boldsymbol{\psi})$ in the transition equation.

These hyperparameters $\boldsymbol{\psi}$ determine the stochastic properties of the model, whereas the parameters appearing in $\mathbf{a}_t(\boldsymbol{\psi})$ and $\mathbf{c}_t(\boldsymbol{\psi})$ only affect the expected value of the state and the observations in a deterministic way. It should, however, be noted that this distinction can become blurred, for example if $\mathbf{a}_t(\boldsymbol{\psi})$ is a function of a lagged value of $\mathbf{y}_t$.

If $\mathbf{a}_t(\boldsymbol{\psi})$ or $\mathbf{c}_t(\boldsymbol{\psi})$ is a linear function of unknown parameters, hyperparameters $\boldsymbol{\psi}$ can be treated as state variables.

This approach will be followed in our data analysis, i.e. both parameters, the variance of the measurement errors $\boldsymbol{\epsilon}_t$ from (1.1) and the variance of the transition (process) errors $\boldsymbol{\eta}_t$ from (1.2), will be treated as two independent, separate state variables (not as a part of the vector $\boldsymbol{\psi}$). These are to be estimated as well.

1.3.4 Uniqueness and identifiability of a state space model

- **Problems with uniqueness**

  The definition of the state vector $\boldsymbol{\xi}_t$ of any particular statistical model is determined by construction. Its elements may or may not be identifiable with components which have a substantive interpretation, for example as a *trend* or a *seasonal*.

  From the technical point of view, the aim of the state space formulation it to set up $\boldsymbol{\xi}_t$ in such a way that it contains all the relevant information on the system at time t and that it does so by having as small a number of elements as possible.

  A state space form which minimizes the length of the state vector is said to be a *minimal realization*. This minimal realization is a basic criterion for a good state space representation. However, it does not imply that there is necessarily a unique representation for any particular problem.

  In fact, a unique representation is the exception rather than the rule. There could be for example two (different) possible state space representations for the second-order autoregressive process AR(2). See more about the non-uniqueness in Harvey [13].

- **Problems with identifiability**

  A model is said to be *not identifiable*, if there are observationally equivalent structures of model parameters $\boldsymbol{\psi}$ that imply the same distribution for the observable random outcomes $\mathbf{y}_t$ – there exist different parameter vectors that lead to the same likelihood function $L(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1; \boldsymbol{\psi})$.

  More theoretically, a *concept of identification* should be defined. Please see more in Kellerhals [15].

Definiton 1.3.13 (Concept of Identification)**.** *With denoting $\boldsymbol{F}$ as parameterized cumulative distribution function, $\boldsymbol{\psi}$ the parameter vector contained in the admissible parameter space $\boldsymbol{\Psi}$, we call the set $\left\{\boldsymbol{F}(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1|\boldsymbol{\psi})\right\}_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}$ a **model** and an element $\boldsymbol{F}(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1|\boldsymbol{\psi})$ of this set a **structure**. A model is called **globally identifiable** at particular parameter values $\boldsymbol{\psi}^0$ if*

$$\begin{array}{rcll} \boldsymbol{\psi} & \neq & \boldsymbol{\psi}^0 & \Rightarrow \\ \boldsymbol{F}\big(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1|\boldsymbol{\psi}\big) & \neq & \boldsymbol{F}\big(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1|\boldsymbol{\psi}^0\big), & \end{array}$$

i.e. there exist no other parameter values besides the vector $\boldsymbol{\psi}^0$ that give rise to the same structure.

A model is said to be ***locally identified*** *at* $\boldsymbol{\psi}^0$ *if there exists a* $\delta > 0$ *such that for any value of* $\boldsymbol{\psi}$ *satisfying*

$$(\boldsymbol{\psi} - \boldsymbol{\psi}^0)^\top (\boldsymbol{\psi} - \boldsymbol{\psi}^0) < \delta,$$

there exist realizations $\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1$ *for which we obtain two different structures.*

Remark 1.3.14. *For an unidentified model example in the case of maximum likelihood estimation use the Kalman filter, see Kellerhals [15].*

Now, for the derivation of the main Kalman filter, we just focus on the general system characterized by equations (1.1) and (1.2).

1.4 The Kalman filter

Once a model has been put in a state space form, the way is opened for the application of the Kalman filter. We have already discussed that the Kalman filter is a recursive procedure for computing the optimal estimator of the state vector $\boldsymbol{\xi}_t$ at time t, based on the information available at time t. This information consists of the observations up to and including $\mathbf{y}_t$. According to the assumed state space form of equations (1.1) and (1.2), we specify stochastic processes of explaining factors $\boldsymbol{\xi}_t$, which describe the state of the system.

Instead of being able to observe the factors directly, we can only observe some noisy function $\mathbf{y}_t$ of $\boldsymbol{\xi}_t$. The problem of determining the state of the system from noisy measurements $\mathbf{y}_t$ is called *estimation*. The special estimation problem of *filtering* has the object of obtaining an expression for the optimal estimate of $\boldsymbol{\xi}_t$ given the observations up to time t. The system matrices together with the initial values $\mathrm{E}[\boldsymbol{\xi}_0] = \boldsymbol{\xi}_{0|0}$ and $\mathrm{Cov}[\boldsymbol{\xi}_0] = \boldsymbol{\Sigma}_{0|0}$ (where $\boldsymbol{\xi}_0$ is the initial state vector) are both assumed to be known in all time periods and thus do not need to be explicitly included in the information set.

The Kalman filter is important because of its on-line estimation: the current value of the state vector is of prime interest and the Kalman filter enables the estimate of the state vector to be continually updated as new observations become available. From the economical point of view, the value of such a procedure in economic applications would appear to be limited. New observations tend to appear at rather less frequent intervals and the emphasis is on making predictions of future observations based on a given sample. The state vector does not always have an economic interpretation, and – in cases where it does – it is more appropriate to estimate its value at a particular point in time using all the information in the sample, not just a part of it. These two problems are known as *prediction* and *smoothing*, respectively. The Kalman filter provides the basis for the solution of both of them. Therefore, the estimation problem can be classified upon the available and processed information into three different problems according to the following definition.

Definiton 1.4.1 (Estimation)**.** *Considering the problem of estimation $\boldsymbol{\xi}_t$ using information up to time s, denoted by the information set $\mathcal{F}_t = \{\mathbf{y}_s, \ldots, \mathbf{y}_2, \mathbf{y}_1\}$, we differentiate between the three cases of a*

$$\begin{aligned} \textbf{\textit{prediction problem}} &: \textit{for} \quad t > s, \\ \textbf{\textit{filtering problem}} &: \textit{for} \quad t = s, \quad \textit{and} \\ \textbf{\textit{smoothing problem}} &: \textit{for} \quad t < s, \end{aligned}$$

depending on which information set we use in an estimation.

In our data analysis:

- ***prediction*** means estimating $\boldsymbol{\xi}_t$ in terms of $(\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$,
- ***filtering*** means estimating $\boldsymbol{\xi}_t$ in terms of $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$, while
- ***smoothing*** means estimating $\boldsymbol{\xi}_t$ in terms of all observations $(\mathbf{y}_1, \ldots, \mathbf{y}_T)$.

Another central function of the Kalman filter is the likelihood estimation: when the disturbances and the initial state vector $\boldsymbol{\xi}_0$ are normally distributed, it allows to calculate the likelihood and this opens the way for the estimation of any unknown parameters in the model. This leads to the statistical testing and model specification. You can read more about the likelihood estimation in Kellerhals [15].

The derivation of the Kalman filter given later in this section (and in the rest of this thesis) is based on the assumption that the disturbances and the initial state vector $\boldsymbol{\xi}_0$ are normally distributed. A standard result on the multivariate normal distribution is then used to show how it is possible to calculate **recursively** the distribution of the state vector $\boldsymbol{\xi}_t$, conditional on the information set at time t, for all t from 1 to T. These conditional distributions are themselves also normal and hence are completely specified by their means and covariances matrices – it is these quantities which the Kalman filter computes.

After the derivation of the Kalman filter, it is shown that the mean of the conditional distribution of the state vector $\boldsymbol{\xi}_t$ is an **optimal estimator** of $\boldsymbol{\xi}_t$ in the sense that it minimizes the mean square error (MSE).

When the normality assumption is dropped, there is no longer any guarantee that the Kalman filter will give the conditional mean of the state vector $\boldsymbol{\xi}_t$. However, it is still an optimal estimator (within the class of all linear estimators) in the sense that is minimizes the mean square error.

Furthermore, we need a clearer concept of what constitutes a statistically optimal estimate in that context. The statistical criterion of mean square error and the other statistical terms may be defined now.

Definiton 1.4.2 (Optimality Criterion)**.** *Let $\boldsymbol{\xi}^*_{t|t-1}$ denote an estimate of $\boldsymbol{\xi}_t$ based on the information set $\mathcal{F}_{t-1}$. In order to choose the optimal of various possible forecasts, we need to specify a criterion of what "optimal" means. Very convenient results are obtained from assuming a quadratic loss function $\mathbf{L}(\mathbf{e_t})$ on the estimation error $\mathbf{e_t} = \boldsymbol{\xi}_t - \boldsymbol{\xi}^*_{t|t-1}$. Choosing the estimator as to minimize the mean square error*

$$\text{MSE}(\boldsymbol{\xi}^*_{t|t-1}) \equiv \text{E}[\boldsymbol{\xi}_t - \boldsymbol{\xi}^*_{t|t-1}]^2$$

*results in the minimum mean square estimator (*MMSE*) as the best or **optimal estimator** with respect to any quadratic function of the estimation error.*

Theorem 1.4.3 (Optimal Estimator)**.** *Given the availability of the information set $\mathcal{F}_{t-1}$, the optimal estimator $\boldsymbol{\xi}_{t|t-1}$ of $\boldsymbol{\xi}_t$ among all estimators $\boldsymbol{\xi}^*_{t|t-1}$ is the expected value of $\boldsymbol{\xi}_{t-1}$ conditional on the information at time $t-1$. Thus the* MMSE *is given by $\boldsymbol{\xi}_{t|t-1} = \text{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}]$. The corresponding* MSE *is given by*

$$\text{MSE}(\boldsymbol{\xi}_{t|t-1}) = \text{E}\big[\boldsymbol{\xi}_t - \text{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}]\big]^2 = \text{Cov}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}],$$

i.e. in this case the mean square error matrix is equal to the variance-covariance matrix.

PROOF. See Hamilton [11], page 72. □

Remark 1.4.4 (Conditional Expectation and Variance)**.** *In considering the problem of estimating $\boldsymbol{\xi}_t$ using information up to time s, i.e. the information set $\mathcal{F}_s = \{\mathbf{y}_s, \ldots, \mathbf{y}_2, \mathbf{y}_1\}$, we denote the conditional expectation of $\boldsymbol{\xi}_t$ given $\mathcal{F}_s$ in the further analysis for convenience by $\text{E}[\boldsymbol{\xi}_t|\mathcal{F}_s] \equiv \boldsymbol{\xi}_{t|s}$. For the second conditional moment – the conditional variance-covariance matrix of $\boldsymbol{\xi}_t$ given $\mathcal{F}_s$ – we will further use $\text{Cov}[\boldsymbol{\xi}_t|\mathcal{F}_s] \equiv \boldsymbol{\Sigma}_{t|s}$.*

1.4.1 General form of the Kalman filter

The Kalman filter is a **set of equations** which allows an estimator to be updated once new observational information becomes available. This process is carried out in two distinct parts:

1. **Prediction Step**: The first step consists of forming an optimal predictor of the next observation, given all the information available up to time $t-1$. We extrapolate the state vector $\boldsymbol{\xi}_t$ by means of conditional expectation utilizing the information set $\mathcal{F}_{t-1}$ and calculate a so-called *a priori estimate* for time t.

2. **Updating Step**: The a priori state estimate is updated with the new information arriving at time t that is combined with the already available information from time $t-1$. The result of this step is called the *filtered estimate* or the *a posteriori estimate*. Then the *Kalman gain* matrix in estimation is realized.

We will find that the Kalman filter really provides an **optimal** solution to the presented problem of prediction and updating. In deriving the Kalman filter we can choose among several approaches available in literature. In this thesis we present the elementary derivation of the Kalman filter under the normality assumption which yields the interpretation of the Kalman filter as the optimal filter in the sense of a *minimum mean square estimator* (MMSE).

We will also obtain the Kalman filter for the general case of non-normality by exploiting the given linear relationships of the observations and dynamics stated respectively in the state space equations (1.1) and (1.2). In this case we speak of the Kalman filter as the optimal filter in the sense of a *minimum mean square linear estimator* (MMSLE).

1.4.2 Determining of the Kalman filter, MMSE

In the case of a minimum mean squared error estimator (MMSE), it will be assumed that the additive error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are independently and normally distributed. Furthermore, the error terms are treated as independent of the initial state vector $\boldsymbol{\xi}_0$, which is assumed to be normally distributed with $\mathrm{E}[\boldsymbol{\xi}_0] = \boldsymbol{\xi}_{0|0}$ and $\mathrm{Cov}[\boldsymbol{\xi}_0] = \boldsymbol{\Sigma}_{0|0}$. Therefore, since the transition equation (1.2) is linear in $\boldsymbol{\xi}_{t-1}$ and the error term $\boldsymbol{\eta}_t$ is normally distributed, the state vector $\boldsymbol{\xi}_t$ – as the sum of $\boldsymbol{\xi}_{t-1}$ and $\boldsymbol{\eta}_t$ – is also normally distributed. Furthermore, for the measurement equation (1.1) we also have a normally distributed $\mathbf{y}_t$, since $\boldsymbol{\xi}_t$ and $\boldsymbol{\epsilon}_t$ are both normal.

1. Prediction Step: First, the state space model of (1.1) and (1.2) is considered. Let $\boldsymbol{\xi}_{t-1|t-1}$ (often also $\widehat{\boldsymbol{\xi}}_{t-1}$ or $\widehat{\boldsymbol{\xi}}_{t-1|t-1}$) denote the optimal estimator of the state vector $\boldsymbol{\xi}_{t-1}$ based on the observations up to and including $\mathbf{y}_{t-1}$, i.e. the history of the observations up to time $t-1$ denoted by $\mathbf{Y}_{t-1} = (\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$ is considered.

In the *prediction step*, we forecast the state vector by calculating the conditional mean of the state variables on both sides of equation (1.2) given the information up to time $t-1$:

$$\begin{aligned} \boldsymbol{\xi}_{t|t-1} &\equiv \mathrm{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}] \\ &= \mathrm{E}[\mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t|\mathcal{F}_{t-1}] \\ &= \mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1|t-1}. \end{aligned} \tag{1.6}$$

Hence, given $\boldsymbol{\xi}_{t-1|t-1}$ we have the optimal estimator $\boldsymbol{\xi}_{t|t-1}$ of the state vector $\boldsymbol{\xi}_t$ (still based just on the observations up to time $t-1$).

Once more, we consider $\boldsymbol{\xi}_{t-1|t-1}$ as the optimal estimator of the state vector $\boldsymbol{\xi}_{t-1}$. The *a priori estimate error* is:

$$\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}.$$

Let $\boldsymbol{\Sigma}_{t-1|t-1}$ denote the $(k \times k)$ covariance matrix of this estimation error, i.e.

$$\boldsymbol{\Sigma}_{t-1|t-1} = \mathrm{E}\left[\left(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}\right)\left(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}\right)^{\top} \middle| \mathcal{F}_{t-1}\right].$$

Given this covariance matrix $\boldsymbol{\Sigma}_{t-1|t-1}$, the corresponding variance-covariance matrix for the state variables is given by

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t-1} &\equiv \mathrm{E}\left[\left(\boldsymbol{\xi}_t - \mathrm{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}]\right)\left(\boldsymbol{\xi}_t - \mathrm{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}]\right)^{\top} \middle| \mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\left[\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)^{\top} \middle| \mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\Big[\left(\mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t - \mathbf{c}_t - \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1|t-1}\right) \\
&\qquad \times \left(\mathbf{c}_t^{\top} + \boldsymbol{\xi}_{t-1}^{\top}\boldsymbol{\Phi}_t^{\top} + \boldsymbol{\eta}_t^{\top} - \mathbf{c}_t^{\top} - \boldsymbol{\xi}_{t-1|t-1}^{\top}\boldsymbol{\Phi}_t^{\top}\right) \Big| \mathcal{F}_{t-1}\Big] \\
&= \boldsymbol{\Phi}_t\,\mathrm{E}\left[\boldsymbol{\xi}_{t-1|t-1}\boldsymbol{\xi}_{t-1|t-1}^{\top} \middle| \mathcal{F}_{t-1}\right]\boldsymbol{\Phi}_t^{\top} + \mathrm{E}\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{\top} \middle| \mathcal{F}_{t-1}\right] \\
&= \boldsymbol{\Phi}_t\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_t^{\top} + \mathbf{Q}_t. \qquad (1.7)
\end{aligned}$$

The matrix $\boldsymbol{\Sigma}_{t|t-1}$ is known as the *a priori estimate error covariance matrix*. Equations (1.6) and (1.7) are known as the *prediction equations* of the Kalman filter.

Now the **prediction error** denoted by $\mathbf{I}_t$ can be defined and using the given data for the measurable observations $\mathbf{y}_t$ can be – as an intermediate result – obtained:

$$\begin{aligned}
\mathbf{I}_t &= \mathbf{y}_t - \mathbf{y}_{t|t-1} \\
&= \mathbf{y}_t - \mathrm{E}[\mathbf{y}_t|\mathcal{F}_{t-1}] \\
&= \mathbf{y}_t - \mathbf{a}_t - \mathbf{B}_t\boldsymbol{\xi}_{t|t-1}.
\end{aligned}$$

This prediction error $\mathbf{I}_t$, i.e. the difference $\mathbf{y}_t - \mathbf{a}_t - \mathbf{B}_t\boldsymbol{\xi}_{t|t-1}$, is called the measurement *innovation* (or just the *residual*), since it represents the new information in the latest observation.

As can be seen further in the updating step, they play a key role in updating the estimator of the state vector $\boldsymbol{\xi}_t$. This residual reflects the discrepancy between the predicted measurement $\mathbf{B}_t\boldsymbol{\xi}_{t|t-1}$ and the actual measurement $\mathbf{y}_t$. A residual of zero means that the two are in complete agreement – therefore, the prediction error $\mathbf{I}_t$ is equal to zero. The further $\mathbf{I}_t$ is from a null vector, the greater is the "correction" in the estimator of $\boldsymbol{\xi}_t$.

Furthermore, the variance-covariance matrix of this prediction error $\mathbf{I}_t$ can be calculated as follows:

$$\begin{aligned}
\mathbf{F}_{t|t-1} &= \mathrm{Cov}\left[\mathbf{I}_t|\mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\left[\left(\mathbf{y}_t - \mathrm{E}[\mathbf{y}_t|\mathcal{F}_{t-1}]\right)\left(\mathbf{y}_t - \mathrm{E}[\mathbf{y}_t|\mathcal{F}_{t-1}]\right)^\top \Big| \mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\Big[\Big(\mathbf{a}_t + \mathbf{B}_t\boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t - \mathbf{a}_t - \mathbf{B}_t\boldsymbol{\xi}_{t|t-1}\Big) \\
&\qquad \times \Big(\mathbf{a}_t^\top + \boldsymbol{\xi}_t^\top\mathbf{B}_t^\top + \boldsymbol{\epsilon}_t^\top - \mathbf{a}_t^\top - \boldsymbol{\xi}_{t|t-1}^\top\mathbf{B}_t^\top\Big) \Big| \mathcal{F}_{t-1}\Big] \\
&= \mathbf{B}_t\,\mathrm{E}\left[\boldsymbol{\xi}_{t|t-1}\boldsymbol{\xi}_{t|t-1}^\top \big| \mathcal{F}_{t-1}\right]\mathbf{B}_t^\top + \mathrm{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top \big| \mathcal{F}_{t-1}\right] \\
&= \mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top + \mathbf{H}_t. \qquad (1.8)
\end{aligned}$$

The innovations $\mathbf{I}_t$ are centered (the mean of $\mathbf{I}_t$ is a vector of zeros), uncorrelated random variables $\big($i.e. $\mathrm{E}[\mathbf{I}_t\mathbf{I}_s^\top] = 0^\top$ for $t \neq s$, $t, s = 1, \ldots, T\big)$ and variables $\mathbf{F}_{t|t-1}$ are their variances, i.e. $\mathrm{Var}[\mathbf{I}_t] = \mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top + \mathbf{H}_t$. This holds even in the absence of normality assumption.

In a Gaussian model, it can also be shown $\big($in Kellerhals [15], for example$\big)$ that the joint density of the observations can be decomposed in terms of the innovations, which are independently and normally distributed. Hence,

$$\mathbf{I}_t \sim \mathrm{N}[\mathbf{0}, \mathbf{F}_{t|t-1}].$$

2. Updating Step: Once the new observation $\mathbf{y}_t$ becomes available, the estimator $\boldsymbol{\xi}_{t|t-1}$ of $\boldsymbol{\xi}_t$ can be updated. In the *updating step* we also update the inference on $\boldsymbol{\xi}_{t|t-1}$ by including the newly available information at time t; this results in the filtered estimate $\boldsymbol{\xi}_{t|t}$. In this way, the so-called *Kalman gain* $\mathbf{K}_t$ will be realized and defined.

To obtain the *updating equations* of the Kalman filter, we first consider the joint distribution of $\boldsymbol{\xi}_t$ and $\mathbf{y}_t$. As assumed, both variables are normally distributed, i.e. the joint distribution given the information $\mathcal{F}_{t-1}$ is

$$\begin{bmatrix} \boldsymbol{\xi}_t \\ \mathbf{y}_t \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\xi}_{t|t-1} \\ \mathbf{y}_{t|t-1} \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top \\ \mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1} & \mathbf{F}_{t|t-1} \end{bmatrix}\right) \qquad (1.9)$$

with the following conditional means:

$$\begin{aligned}
\boldsymbol{\xi}_{t|t-1} &\equiv \mathrm{E}[\boldsymbol{\xi}_t|\mathcal{F}_{t-1}] \\
&= \mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1|t-1}, \text{ and} \\
\\
\mathbf{y}_{t|t-1} &\equiv \mathrm{E}[\mathbf{y}_t|\mathcal{F}_{t-1}] \\
&= \mathbf{a}_t + \mathbf{B}_t\boldsymbol{\xi}_{t|t-1}.
\end{aligned}$$

The conditional variance-covariance matrices are stated in equations (1.7) and (1.8). We further compute:

$$\begin{aligned} \mathrm{E}\left[\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)\left(\mathbf{y}_t - \mathbf{y}_{t|t-1}\right)^\top \middle| \mathcal{F}_{t-1}\right] &= \mathrm{E}\left[\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)^\top \middle| \mathcal{F}_{t-1}\right]\mathbf{B}_t^\top \\ &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top, \\ \mathrm{E}\left[\left(\mathbf{y}_t - \mathbf{y}_{t|t-1}\right)\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}\right)^\top \middle| \mathcal{F}_{t-1}\right] &= \mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}. \end{aligned}$$

For the further derivation, one may take advantage of the following result for normal variables.

Lemma 1.4.5 (Conditional Normal Distribution)**.** *Let z_1 and z_2 denote $(n_1 \times 1)$ and $(n_2 \times 1)$ vectors of random variables with the following joint normal distribution:*

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^\top & \Omega_{22} \end{bmatrix}\right).$$

Then the distribution of z_1 conditional on z_2 is $\mathrm{N}[m, \Sigma]$, where

$$\begin{aligned} m &= \mu_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(z_2 - \mu_2), \quad \textit{and} \\ \Sigma &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}^\top. \end{aligned}$$

Thus the optimal forecast (in the sense that it minimizes the mean square error) of z_1 conditional on having observed z_2 is given by

$$\mathrm{E}[z_1|z_2] = m$$

with Σ characterizing the mean square error of this forecast:

$$\mathrm{E}\left[(z_1 - m)(z_1 - m)^\top \middle| z_2\right] = \Sigma.$$

PROOF. The proof can be found, for example, in Härdle and Simar [9]. □

Applying this result to equation (1.9) $\left(\text{the joint distribution of the vector } [\boldsymbol{\xi}_t, \mathbf{y}_t]^\top\right)$ yields the following distribution for $\boldsymbol{\xi}_t$ given the information $\mathcal{F}_t$ of the observable data $\mathbf{y}_t$:

$$\boldsymbol{\xi}_t \sim \mathrm{N}\left[\boldsymbol{\xi}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{F}_{t|t-1}^{-1}(\mathbf{y}_t - \mathbf{y}_{t|t-1}),\ \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{F}_{t|t-1}^{-1}\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\right].$$

Comparing this distribution with the result for the conditional moments of $\boldsymbol{\xi}_t$ given the information $\mathcal{F}_t$,

$$\boldsymbol{\xi}_t \sim \mathrm{N}\left[\boldsymbol{\xi}_{t|t},\ \boldsymbol{\Sigma}_{t|t}\right],$$

the following relationships can be obtained, where $\mathbf{I}_t$ denotes the innovations (prediction error) defined above:

$$\begin{aligned}
\boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top \mathbf{F}_{t|t-1}^{-1}\mathbf{I}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top \mathbf{F}_{t|t-1}^{-1}\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}.
\end{aligned}$$

Now, the so-called *Kalman gain* $\mathbf{K}_t$ can be defined by substituting:

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top \mathbf{F}_{t|t-1}^{-1}. \tag{1.10}$$

These steps finally result in the **update** for the state vector

$$\begin{aligned}
\boldsymbol{\xi}_{t|t} &= \mathrm{E}[\boldsymbol{\xi}_t|\mathcal{F}_t] \\
&= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t,
\end{aligned} \tag{1.11}$$

and its variance-covariance matrix

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t} &= \mathrm{E}\left[\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t}\right)\left(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t}\right)^\top \middle| \mathcal{F}_t\right] \\
&= \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1} \\
&= \left(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t\right)\boldsymbol{\Sigma}_{t|t-1}.
\end{aligned} \tag{1.12}$$

The matrix $\boldsymbol{\Sigma}_{t|t}$ is known as the *a posteriori estimate error covariance matrix*. Equations (1.11) and (1.12) are known as the *updating equations* of the Kalman filter.

Summary Taken together, the prediction equations (1.6) with (1.7) and the updating equations (1.11) with (1.12), i.e. set of equations

$$\begin{aligned}
\boldsymbol{\xi}_{t|t-1} &= \mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1|t-1}, \\
\boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Phi}_t\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_t^\top + \mathbf{Q}_t,
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, \\
\boldsymbol{\Sigma}_{t|t} &= \left(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t\right)\boldsymbol{\Sigma}_{t|t-1},
\end{aligned}$$

where $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top \mathbf{F}_{t|t-1}^{-1}$ is the Kalman gain, make up the Kalman filter. If desired they can be written as a single set of recursions (going directly from $\boldsymbol{\xi}_{t-1|t-1}$ to $\boldsymbol{\xi}_{t|t-1}$, etc.), which will be discussed in the following subsection.

1.4.3 Kalman filter and its recursions

Having derived the Kalman filter algorithm yielding the MMSE (or eventually the MMSLE in the next subsection) described in the previous subsections, the way is opened to apply the sequence of filtering equations **recursively** as each new observation $\mathbf{y}_t$ becomes available.

To be able to understand the Kalman filter well, we shall once more repeat what we have already discussed in the last paragraphs.

The Kalman filter estimates a process by using a form of feedback control: the filter estimates the process state at some time t and then obtains feedback in the form of (noisy) measurements. As such, the equations for the Kalman filter fall into two groups, as descried: the *prediction equations* (or *time update equations*) and the *updating equations* (or *measurement update equations*).

- The first group of these equations, the ***time*** *update equations*, is responsible for projecting forward (in the sense of ***time***) the current state vector $\boldsymbol{\xi}_t$ and error covariance estimates $\boldsymbol{\Sigma}_t$ to obtain the *a priori estimates* for the next time step $t \rightarrow t+1$.

- The second group, the ***measurement*** *update equations*, is responsible for the feedback – i.e. for incorporating a new ***measurement*** $\mathbf{y}_t$ into the a priori estimate to obtain an improved *a posteriori estimate*.

The time update equations can also be thought of as *predictor equations*, while the measurement update equations can be thought of as *corrector equations*. Please follow Figure 1.2.

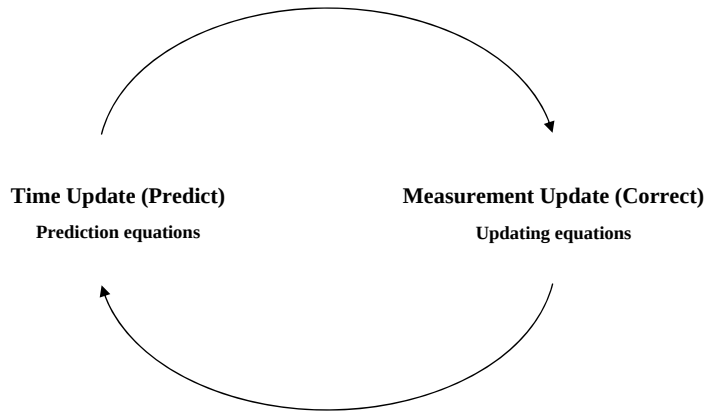


Figure 1.2: The ongoing discrete Kalman filter cycle. The ***time*** *update* projects the current state estimate ahead in ***time***. The ***measurement*** *update* adjusts the projected estimate by an actual ***measurement*** at that time.

Firstly, we shall take a look at the core of the recursive algorithm. For now, we do not consider any dependence on the hyperparameter $\boldsymbol{\psi}$ described in the third part of Subsection 1.3.3. The final estimation algorithm – or the core of the Kalman filter just as a "prediction tool", if you like – is shown below in Figure 1.3.

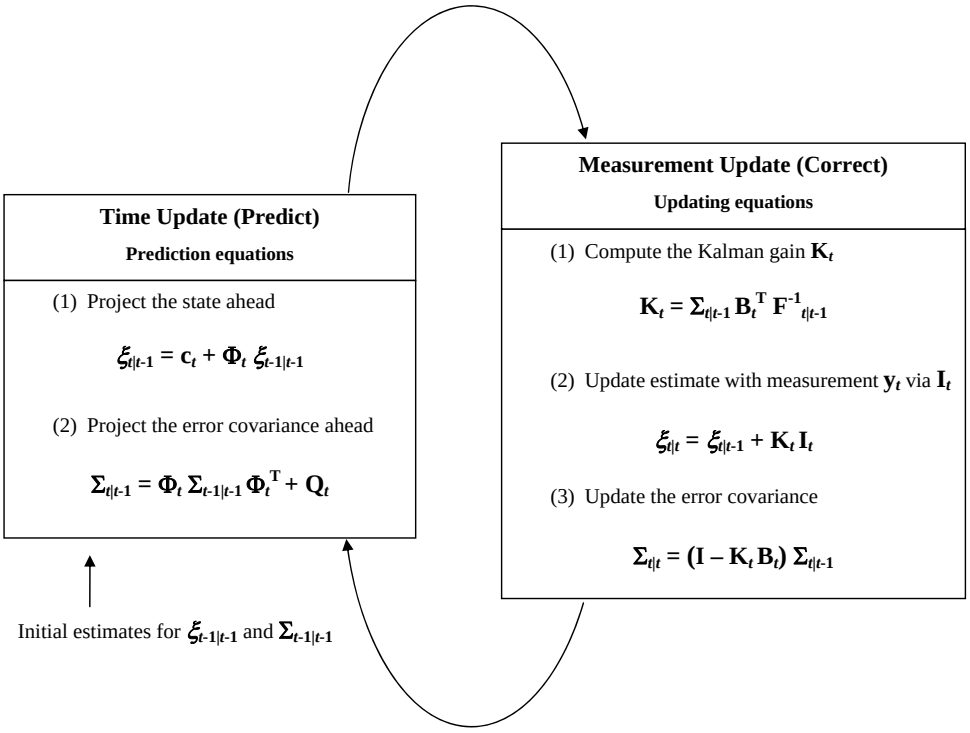


Figure 1.3: The core operations of the Kalman filter, using the measurement and prediction equations.

To be more precise and able to describe the Kalman filter as a whole, any dependence on the hyperparameter $\boldsymbol{\psi}$ (please see more in Subsection 1.3.3) should be considered. During the complete Kalman filter, this hyperparameter $\boldsymbol{\psi}$ is being estimated as well.

Please follow the flowchart of Figure 1.4 for a better, graphical illustration of the Kalman filter algorithm. We try to explain how the algorithm works "step-by-step". The recursive version of the Kalman filter works fully in the following steps:

1. **Initialization step:** The recursive algorithm starts with a feasible choice of the parameter vector $\boldsymbol{\psi}$ and corresponding initial state vector $\boldsymbol{\xi}_0$ with the initial values: the mean of $\boldsymbol{\xi}_{0|0}$ and the variance-covariance matrix of $\boldsymbol{\Sigma}_{0|0}$.

2. **Prediction step, a priori estimates:** Thereupon, we use the prediction equations to calculate the a priori estimates

$$\begin{aligned} \boldsymbol{\xi}_{t|t-1} &= \mathbf{c}_t(\boldsymbol{\psi}) + \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_{t-1|t-1}, \text{ and} \\ \boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_t(\boldsymbol{\psi})^\top + \mathbf{Q}_t(\boldsymbol{\psi}). \end{aligned}$$

3. **The prediction error, its MSE:** Using the current market information on observable variables $\mathbf{y}_t$ at time $t = 1$, the prediction error and its corresponding MSE will be derived as follows

$$\begin{aligned} \mathbf{I}_t &= \mathbf{y}_t - \mathbf{a}_t(\boldsymbol{\psi}) - \mathbf{B}_t(\boldsymbol{\psi})\boldsymbol{\xi}_{t|t-1}, \text{ and} \\ \mathbf{F}_{t|t-1} &= \mathbf{B}_t(\boldsymbol{\psi})\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t(\boldsymbol{\psi})^\top + \mathbf{H}_t(\boldsymbol{\psi}). \end{aligned}$$

4. **Updating step, the filtered estimates:** The first task during the measurement update is to compute the Kalman gain $\mathbf{K}_t$ defined in (1.10), i.e.

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t(\boldsymbol{\psi})^\top \mathbf{F}_{t|t-1}^{-1}.$$

   Given this matrix $\mathbf{K}_t$ and the intermediate results from the previous step we then update the a priori estimates yielding the filtered estimates:

$$\begin{aligned} \boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, \\ \boldsymbol{\Sigma}_{t|t} &= \big(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t(\boldsymbol{\psi})\big)\boldsymbol{\Sigma}_{t|t-1}, \end{aligned}$$

   as optimal in the sense of MMSE or MMSLE.

   In detail: the next task of this step is to actually measure the process to obtain $\mathbf{y}_t$, and then to generate an *a posteriori state estimate* by incorporating the measurement as in $\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t$. The final task is to obtain an *a posteriori error covariance estimate* via $\boldsymbol{\Sigma}_{t|t} = \big(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t(\boldsymbol{\psi})\big)\boldsymbol{\Sigma}_{t|t-1}$.

5. **Recursive step:** Now for each discrete time step $t \to t+1$, we recursively feed the results for the state vector $\boldsymbol{\xi}_{t|t}$ and its variance-covariance matrix $\boldsymbol{\Sigma}_{t|t}$ from the updating equations into the prediction equations until we reach the last observations $\mathbf{y}_T$.[1]

[1] Please note, these are then formally indexed with $t-1$, since we denote the actual information we are working with by the time index t.

After each time t and measurement update pair, the process is repeated with the previous *a posteriori* estimates used to project or predict the new *a priori* estimates.

This **recursive** nature is one of the very appealing features of the Kalman filter – it makes practical implementations much more feasible than an implementation of other filter algorithms (for example the Wiener filter, see more in Brown and Hwang [3]) which are most designed to operate on **all** of the data **directly** for each estimate. The Kalman filter instead recursively conditions the current estimate on all of the past measurements.

6. **Likelihood:** The conditional likelihood function is evaluated at the current parameter values using the results of the Kalman recursions.

7. **Criterion met?** Finally, with the obtained time-series of the state variables $\boldsymbol{\xi}_t$ we can evaluate the likelihood function (more in Kellerhals [15]) to choose other more appropriate values for the parameters $\boldsymbol{\psi}$. The Kalman algorithm is run until we reach the predefined abortion criterion.

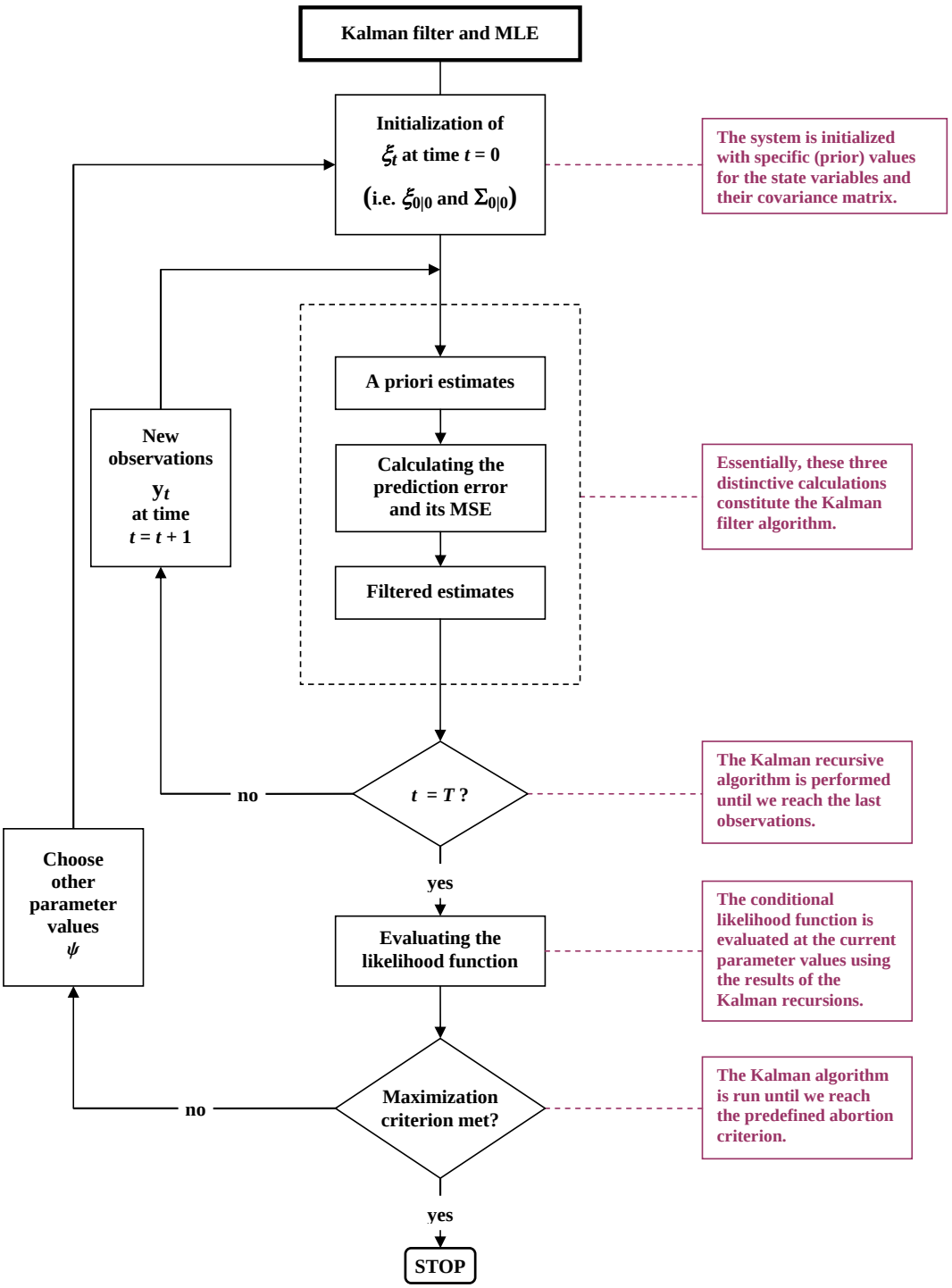


Figure 1.4: Flowchart of the Kalman filter and MLE.

1.4.4 Kalman filter, MMSLE

The Kalman filter as the the minimum mean square linear estimator (MMSLE) without requiring the normality assumption for the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ will now be derived.

1. Prediction Step: First, the state space model of (1.1) and (1.2) is considered. The *prediction equations* (1.11) and (1.12) are obtained similarly to the previous derivation by taking the conditional expectation and variance given the information up to time $t-1$.

2. Updating Step: To obtain the *updating equations*, we start with stating a general linear relationship of the updated estimator $\boldsymbol{\xi}_{t|t}$ and the information of the present sample $(\mathbf{y}_t, \mathbf{a}_t, \mathbf{c}_t)$ and the past information $\mathcal{F}_{t-1}$. The following general linear form will be assumed

$$\boldsymbol{\xi}_{t|t} = \mathbf{K}_t\mathbf{y}_t + \mathbf{L}\mathbf{a}_t + \mathbf{M}\mathbf{c}_t + \mathbf{N}\boldsymbol{\xi}_{t-1|t-1},$$

with arbitrary matrices $\mathbf{K}_t, \mathbf{L}, \mathbf{M}$ and $\mathbf{N}$. These four matrices are in the next step to be chosen so that $\boldsymbol{\xi}_{t|t}$ is the MMSLE, which is wanted.

Therefore, we define the (*a posteriori*) *estimate error* as follows:

$$\begin{aligned}
\mathbf{e}_t &= \boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t} \\
&= \mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t - \mathbf{K}_t\Big[\mathbf{a}_t + \mathbf{B}_t\left(\mathbf{c}_t + \boldsymbol{\Phi}_t\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t\right) + \boldsymbol{\epsilon}_t\Big] \\
&\quad - \mathbf{L}\mathbf{a}_t - \mathbf{M}\mathbf{c}_t - \mathbf{N}(\boldsymbol{\xi}_{t-1} - \mathbf{e}_{t-1}) \\
&= (\boldsymbol{\Phi}_t - \mathbf{N} - \mathbf{K}_t\mathbf{B}_t\boldsymbol{\Phi}_t)\,\boldsymbol{\xi}_{t-1} - (\mathbf{L} + \mathbf{K}_t)\mathbf{a}_t + (\boldsymbol{\mathcal{I}} - \mathbf{M} - \mathbf{K}_t\mathbf{B}_t)\mathbf{c}_t \\
&\quad + \mathbf{N}\mathbf{e}_{t-1} + (\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)\boldsymbol{\eta}_t - \mathbf{K}_t\boldsymbol{\epsilon}_t. \qquad (1.13)
\end{aligned}$$

We are looking for $\boldsymbol{\xi}_{t|t}$ as the MMSLE, i.e. the estimation error $\mathbf{e}_t$, defined in (1.13), needs:

1. to be **unconditionally unbiased**, i.e. the estimation error $\mathbf{e}_t$ has zero expectation, which follows to the conditions

$$\begin{aligned}
\boldsymbol{\Phi}_t - \mathbf{N} - \mathbf{K}_t\mathbf{B}_t\boldsymbol{\Phi}_t &= 0, \\
\mathbf{L} + \mathbf{K}_t &= 0, \\
\boldsymbol{\mathcal{I}} - \mathbf{M} - \mathbf{K}_t\mathbf{B}_t &= 0,
\end{aligned}$$

for equation (1.13). With the substitution $\mathbf{N} = (\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)\boldsymbol{\Phi}_t$, the estimation error can be written as

$$\mathbf{e}_t = (\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)(\boldsymbol{\Phi}_t\mathbf{e}_{t-1} - \boldsymbol{\eta}_t) - \mathbf{K}_t\boldsymbol{\epsilon}_t.$$

2. to have **minimum variance**, i.e. we need to minimize the variance-covariance matrix of the estimation error:

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t} &= \mathrm{E}[\mathbf{e}_t\mathbf{e}_t^\top] \\
&= \mathrm{E}\Bigg[\Big((\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)(\boldsymbol{\Phi}_t\mathbf{e}_{t-1} - \boldsymbol{\eta}_t) - \mathbf{K}_t\boldsymbol{\epsilon}_t\Big) \\
&\qquad \times \Big((\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)(\boldsymbol{\Phi}_t\mathbf{e}_{t-1} - \boldsymbol{\eta}_t) - \mathbf{K}_t\boldsymbol{\epsilon}_t\Big)^\top\Bigg] \\
&= (\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)\Big(\boldsymbol{\Phi}_t\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_t^\top + \mathbf{Q}_t\Big)(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)^\top + \mathbf{K}_t\mathbf{H}_t\mathbf{K}_t^\top \\
&= (\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)\boldsymbol{\Sigma}_{t|t-1}(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_t)^\top + \mathbf{K}_t\mathbf{H}_t\mathbf{K}_t^\top
\end{aligned}$$

with respect to $\mathbf{K}_t$.

For the partial derivative we further get:

$$\begin{aligned}
\frac{\partial\boldsymbol{\Sigma}_{t|t}}{\partial\mathbf{K}_t} &= \frac{\partial}{\partial\mathbf{K}_t}\Big(\boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{K}_t^\top + \\
&\qquad + \mathbf{K}_t\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{K}_t^\top + \mathbf{K}_t\mathbf{H}_t\mathbf{K}_t^\top\Big) \\
&= -(\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1})^\top - \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top + 2\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{K}_t + 2\mathbf{H}_t\mathbf{K}_t \\
&= -2\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top + 2\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\mathbf{K}_t + 2\mathbf{H}_t\mathbf{K}_t.
\end{aligned}$$

Setting the result equal to zero according to the necessary minimization condition, i.e.

$$\frac{\partial\boldsymbol{\Sigma}_{t|t}}{\partial\mathbf{K}_t} \stackrel{!}{=} \mathbf{0},$$

for the Kalman gain matrix can be solved

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top\Big(\mathbf{B}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_t^\top + \mathbf{H}_t\Big)^{-1}, \tag{1.14}$$

which is equivalent to the expression derived in equation (1.12) for the case of the Kalman filter resulting in the MMSE in the previous subsection.

Remark 1.4.6. *The Kalman gain matrix* $\mathbf{K}_t$ *defined in (1.10) or given by (1.14) is therefore chosen to be the gain or blending factor that minimizes the a posteriori error covariance (1.12).*

Looking at (1.14), we see that as the measurement error covariance $\mathbf{H}_t$ *approaches zero, the Kalman gain* $\mathbf{K}_t$ *weights the residual more heavily. Specifically,*

$$\lim_{\mathbf{H}_t \to \mathbf{0}} \mathbf{K}_t = \mathbf{B}_t^{-1}.$$

On the other hand, as the a priori estimate error covariance $\mathbf{\Sigma}_{t|t-1}$ *approaches zero, the gain* $\mathbf{K}_t$ *weights the residual less heavily. Specifically,*

$$\lim_{\mathbf{\Sigma}_{t|t-1} \to \mathbf{0}} \mathbf{K}_t = \mathbf{0}.$$

For more details please see Welch and Bishop [21].

Chapter 2

Kalman filter and CALL options

Having derived the *Kalman filter algorithm*, described its determination in the previous chapter and shown how this recursive algorithm works, the way is opened to use the theory in practice – we try to solve an economical problem in a real case of the CALL options, especially in a case of the state price densities. In detail, we will discuss approaches based on the Kalman filtering, which help us how to estimate the risk-neutral price density of the CALL options. The main algorithm will be tested out with some real data, and the corresponding outputs from the statistical program `R` can be found in the last chapter.

First, we shall remind some useful properties of CALL options (see more, for example, in Franke et al. [6]).

2.1 CALL options

- The *spot price* S_t is the price that is quoted for immediate (spot) settlement (payment and delivery at time t). Therefore S_T is the spot price at maturity T.

- The *strike price* K is a key variable in a derivatives contract between two parties. Where the contract requires delivery of the underlying instrument, the trade will be at the strike price, regardless of the spot price (market price) of the underlying instrument at that time.

- A *CALL option* is a financial contract between two parties (the buyer and the seller) that gives the holder (buyer) the right, but not the obligation, to buy an agreed quantity (usually 100 shares) of a particular commodity or financial instrument (the underlying instrument) from the seller of the option at a certain time (the expiration date T) for a certain price (the strike price K). The seller (or "writer") is obligated to sell the commodity or financial instrument should the buyer so decide. The buyer pays a fee (called a *premium*) for that right.

Remark 2.1.1. *Spot settlement is normally one or two business days from trade date. This is in contrast with the forward price established in a forward contract or futures contract, where contract terms (price) are set now, but delivery and payment will occur at a future date.*

The buyer of a CALL option wants the price of the underlying instrument to rise in the future; the seller either expects that it will not, or is willing to give up some of the upside (profit) from a price rise in return for

1. the premium (paid immediately), plus
2. retaining the opportunity to make a gain up to the strike price K.

CALL options are most profitable for the buyer when the underlying instrument is moving up, making the price of the underlying instrument closer to the strike price K. When the price of the underlying instrument surpasses the strike price K, the option is said to be "in-the-money".

The initial transaction in this context (buying/selling a CALL option) is not the supplying of a physical or financial asset (the underlying instrument). Rather it is the granting of the right to buy the underlying asset, in exchange for a fee – the option price or premium.

Exact specifications may differ depending on option style. In this thesis, we just study a *European CALL option* which allows the holder to exercise the option (i.e., to buy) **only** on the option expiration date T.[1]

From the above, it becomes clear that a CALL option has positive monetary value when the underlying instrument has a spot price S **above** the strike price K. Since the option will not be exercised unless it is "in-the-money", the *payoff* for a CALL option is given by

$$(S_T - K)_+ = \max(S_T - K, 0).$$

The following holds:

- Buying a CALL option means that the buyer
  - expects that the price may go up,
  - pays a premium he never will get back,
  - has the right to exercise the option at the strike price K.

[1] An *American CALL option* allows exercise at **any time** during the life of the option.

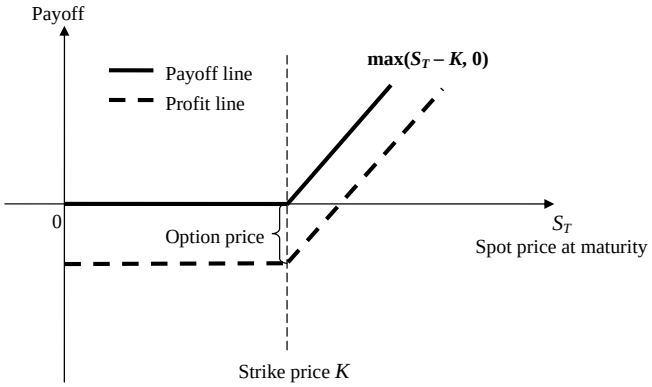


Figure 2.1: Buying a CALL option – the graphical interpretation of the payoffs and profits generated by a CALL option as seen by the buyer. A higher stock price S means a higher profit. Eventually, the price of the underlying security will be high enough to fully compensate for the price of the option.

- Writing a CALL option means that
  - the writer receives the premium,
  - if buyer decides to exercise the option, writer has to sell the stock at the strike price K.

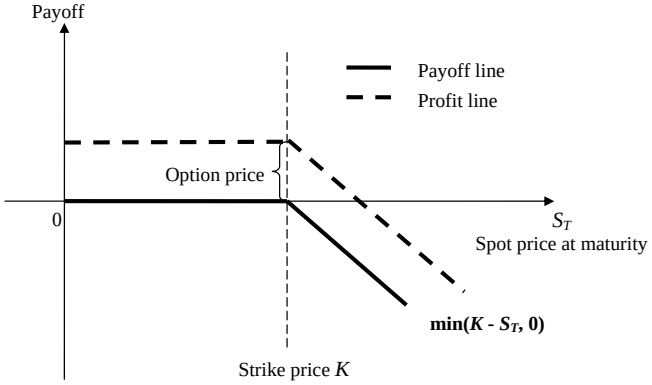


Figure 2.2: Writing a CALL option – the graphical interpretation of the payoffs and profits generated by a CALL option as seen by the writer of the option. Profit is maximized when the strike price K exceeds the price of the underlying security, because the option expires worthless and the writer keeps the premium.

The option value, and therefore the price of a CALL option, varies with the underlying price and with time t. Hence, we shall take a look at the so-called *state price densities* (SPDs) for CALL options, which describe the behavior of the CALL option prices very well, and therefore open way to study the price of a European CALL option at all.

2.2 State price densities

Before proceeding, some of the relevant financial theory will briefly be reviewed – the existence and characterization of *state price densities*.

- In economics and finance, *arbitrage* is the practice of taking advantage of a price differential between two or more markets: a combination of matching deals are struck that capitalize upon the imbalance, the profit being the difference between the market prices.
- When used by academics, an arbitrage is a transaction that involves no negative cash flow at any probabilistic or temporal state and a positive cash flow in at least one state. In simple terms, a *risk-free profit*. A person who engages in arbitrage is called an *arbitrageur*. The term is mainly applied to trading in financial instruments, such as bonds, stocks, derivatives and currencies.

Remark 2.2.1. *If the market prices do not allow for profitable arbitrage, the prices are said to constitute an* ***arbitrage free market*** *or* ***arbitrage equilibrium****.*

SPDs are an important element in applied quantitative finance. SPD has been studied by Black and Scholes [1], Merton [18], Rubinstein [20] and Lucas [17] amongst many others. Under the assumption of no-arbitrage, the SPD is usually called the *risk neutral density* because if one assumes that all investors are *risk neutral*, then the return on all assets must equal the *risk free rate* of interest r.

2.2.1 SPD and CALL options

Although it is not the aim of this thesis, we just briefly remind that the dynamics of option prices carries information on the changing (risk neutral) implied SPDs. Fitting SPDs over time provides useful insight into the behavior of the economic agents and the time inhomogeneity of the market, i.e. the SPD bears important information on the behavior and expectations of the market and is used for pricing. Knowledge of the SPD is therefore a very strong instrument!

The most important application of SPD is that it allows to price options with complicated payoff functions simply by (numerical) integration of the payoff with respect to this density.

We estimate SPDs using real data via the Kalman filter estimator of the second derivative of the European CALL price function. This estimator will then be constrained so as to satisfy no-arbitrage constraints.

Let $\mathbf{C}_t(K,T)$ denote the prices (*CALL pricing function*) of European CALL options with strike price K observed at time t and expiring at time T. It can be shown (see, for example, Breeden and Litzenberger [2]) that the second derivative of the CALL pricing function $\mathbf{C}_t(K,T)$ with respect to the strike price K is related to the state price density in the following form:

$$f(K) = e^{r(T-t)} \frac{\partial^2 \mathbf{C}_t(K,T)}{\partial K^2}. \tag{2.1}$$

Remark 2.2.2. *Equation (2.1) is often used to estimate the state price density by the means of nonparametric regression (see more in Härdle and Yatchew [10]).*

In this thesis we try to develop a simple estimation technique in order to construct SPD estimates. In detail, we construct an estimate of the SPD based on the observed CALL option prices satisfying all shape constraints given later in the following Section 2.3.

2.3 Construction of the estimate

Let us remind that the payoff for a CALL option is given by

$$(S_T - K)_+ = \max(S_T - K, 0).$$

We consider a CALL option with this payoff. Let K be the strike price for this CALL option which will expire at time T. Let t be the current time and $\tau = T - t$ the time to expiry. Let S_T denote the price of the underlying asset (of the stock) at T, and r the risk free interest rate. Then the CALL pricing function $\mathbf{C}_t(K,T)$ at the current time t is given by:

$$\mathbf{C}_t(K,T) = e^{-r(T-t)} \int_0^{+\infty} (S_T - K)_+ f(S_T) dS_T, \tag{2.2}$$

where the function $f(.)$ is the state price density. This means the fair price of a European CALL option can be written as the discounted expected value of the payoff with respect to the SPD $f(.)$. Clearly, the CALL function $\mathbf{C}_t(K,T)$ is monotone decreasing and convex in K.

Our data set

Let us focus on data over a sufficiently brief time span so that we may take the time to maturity τ, the interest rate r, both times t and T as roughly constant. Our objective will now be to estimate the CALL function $\mathbf{C}_t(K,T)$ subject to monotonicity and convexity constraints and the constraint that the implied SPD is non-negative and integrates to one (or at least does not exceed one over the range of observed strike prices).

It will be assumed in the rest of this thesis that the discount factor $e^{-r(T-t)}$ in (2.2) is known (variables r, t and T are known and contained in our data set), so that the option prices $\mathbf{C}_t(K,T)$ on the left side of the equation in (2.2) can be divided by this factor. For the sake of simplicity of the following presentation, the notation of the option prices $\mathbf{C}_t(K,T)$ (now already divided by the discount factor) will be kept.

Our data set contains further the observed option prices $\mathbf{C}_t(K,T)$ for various strike prices K and maturities T. We will analyze the option prices $\mathbf{C}_t(K,T)$ as a function of the strike prices K for fixed date and time to expiry τ.

2.3.1 Some notations

Let us denote the i-th observation of the strike price by $\mathbf{K}_i$ and the corresponding option price, divided now by the discount factor $e^{-r(T-t)}$ from (2.2), by $\mathbf{C}_i = \mathbf{C}_{t,i}(\mathbf{K}_i,T)$. Let $\boldsymbol{\mathcal{C}} = (\mathbf{C}_1,\ldots,\mathbf{C}_n)^\top$ be the vector of the observed option prices at time t. The corresponding vector of the strike prices has the following structure:

$$\boldsymbol{\mathcal{K}} = \begin{pmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \\ \vdots \\ \mathbf{K}_n \end{pmatrix} = \begin{pmatrix} k_1 1_{n_1} \\ k_2 1_{n_2} \\ \vdots \\ k_n 1_{n_p} \end{pmatrix},$$

where $k_1 < k_2 < \cdots < k_p$, $n_j = \sum_{i=1}^{n} I(\mathbf{K}_i = k_j)$ with $I(.)$ denoting the indicator function and 1_{n_j} a vector of ones of length n, and p denoting the number of **various** strike prices $k_1,\ldots,k_p$.

2.3.2 Some assumptions and constraints

It is assumed for a fixed time t and time to maturity $\tau = T - t$, that the i-th observed option price $\mathbf{C}_i$ corresponding to strike price $\mathbf{K}_i$ $\big($i.e. $\mathbf{C}_i = \mathbf{C}_{t,i}(\mathbf{K}_i,T)\big)$ follows the model

$$\mathbf{C}_{t,i}(\mathbf{K}_i,T) = \mu(\mathbf{K}_i) + \boldsymbol{\epsilon}_i, \tag{2.3}$$

where $\boldsymbol{\epsilon}_i$ are i.i.d. $\mathrm{N}\left[\mathbf{0},\boldsymbol{\Sigma}^2\right]$ distributed variables. We even can talk about *heteroscedasticity* if it is assumed that the random errors $\boldsymbol{\epsilon}_i$ in model (2.3) are $\mathrm{N}\left[\mathbf{0},\boldsymbol{\Sigma}^2_{\mathbf{K}_i}\right]$ distributed. Please note that we reduce the model to the case of *homoscedasticity* in our data analysis – this leads to simplification of that model.

Next, we will need the absence of arbitrage: Harrison and Pliska [12] characterized this **no-arbitrage** by the existence of a **unique risk neutral** SPD $f(.)$. From formula (2.2) and general properties of a probability density it follows that the function of the true conditional means $\mu(.)$ has to satisfy the following *no-arbitrage constraints*:

1. it is positive,
2. it is decreasing in K,
3. it is convex, and
4. its second derivative exists and it is a density $\big($i.e. $\mu(.)$ is nonnegative and it integrates to one$\big)$.

These four no-arbitrage constraints will be used in the further construction of the estimate. Some properties of functions $\mu(.)$ and $\mathbf{C}_i(.)$ can be found in Härdle and Hlávka [7] or in Robertson et al. [19].

2.4 Linear model

In the following, we are able to describe the configuration of data, under constraints (a)–(c), using regression models with constraints. The time t and the expiry date T will now be fixed. We also omit these symbols from the notation. However, the expiry date T in our data analysis will change in every time step t.

In the previous subsection, it has been noted that the option prices $\mathbf{C}(.)$ are repeatedly observed for a small number p od distinct strike prices $k_1 < \cdots < k_p$. This situation is visible in our data in Figure 5.1 in the Chapter 5: we have 950 observations observed at $p = 21$ various distinct strike prices $k_1 < \cdots < k_{21}$. This makes such 21 "columns of observations".

To make the following representation even simpler, the regression coefficients β_j in the situation with just four distinct strike prices $k_1 < \cdots < k_4$, i.e. $p = 4$, will be displayed. Please follow Figure 2.3 with the dummy variables, where k_j are equidistant and the distances between the neighboring observed strike prices are equal to one. Therefore, $k_1 = 1, k_2 = 2, k_3 = 3$ and $k_4 = 4$.

Let $\mu_j = \mu(k_j) = \mathrm{E}\left[\mathbf{C}(k_j)\right]$ be the expected values of the option prices given strike price k_j. Then we can write:

$$\begin{aligned}
\mu_p &= \beta_0, \\
\mu_{p-1} &= \beta_0 + \beta_1, \\
\mu_{p-2} &= \beta_0 + 2\beta_1 + \beta_2, \\
\mu_{p-3} &= \beta_0 + 3\beta_1 + 2\beta_2 + \beta_3, \\
&\vdots \\
\mu_1 &= \beta_0 + (p-1)\beta_1 + (p-2)\beta_2 + \ldots + \beta_{p-1}.
\end{aligned}$$

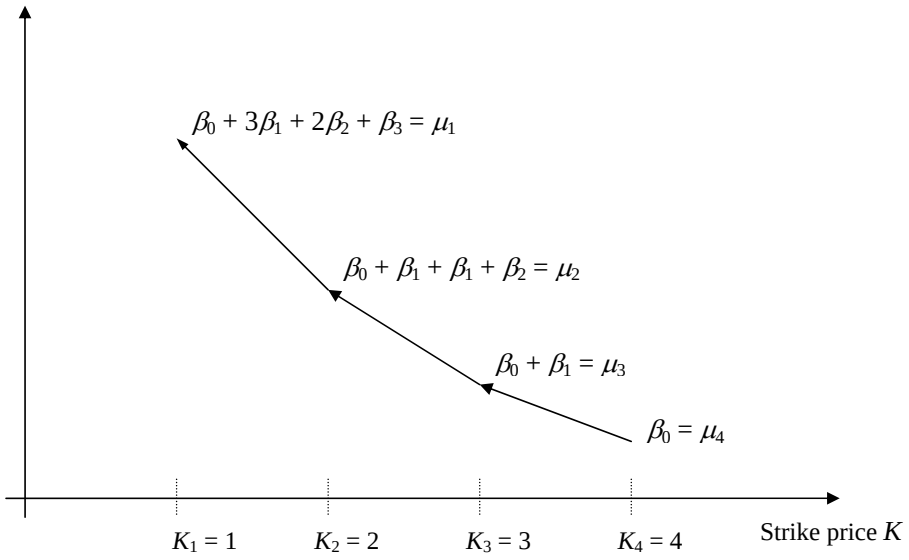


Figure 2.3: Illustration of the dummy variables for CALL options.

Thus, we fit out data using coefficients β_j for $j = 1, \ldots, p$. The conditional means μ_i for $i = 1, \ldots, p$ are replaced by the same number of parameters β_j for $j = 0, \ldots, p-1$ which allow to impose the shape constraints in a more natural way.

Some words about the interpretation

The interpretation of the coefficients β_j can be seen in Figure 2.3:

- β_0 is the mean option price at point $k_4 = 4$ (i.e. $\beta_0 = \mu_4$). Constraint 1 from Subsection 2.3.2 implies that β_0 has to be positive.
- β_1 is the difference between the mean option prices at point $k_4 = 4$ and point $k_3 = 3$ (i.e. $\beta_1 = \mu_3 - \mu_4$). Constraint 2 implies that β_1 has to be also positive.
- The next coefficient – β_2 – approximates the change in the first derivative in point $k_3 = 3$ and it can be interpreted as an approximation of the second derivative in point $k_3 = 3$. Constraint 3 implies that β_2 has to be positive as well.
- Similarly, the coefficient β_3 is an estimate of the (positive) second derivative in point $k_2 = 2$. Constraint 4 can be written as $\beta_2 + \beta_3 \leq 1$.

In practice, we start with the construction of a design matrix $\mathbf{\Delta}$ which allows us to write the above model in the following linear form. For simplicity of presentation, we set $p = 4$ again and use the introduced dummy variables:

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}. \tag{2.4}$$

Ignoring the constraints on the coefficients would lead to a simple linear regression problem. Unfortunately, this approach does not have to lead, and it usually really does not, to interpretable and stable results. Model (2.4) in the above form can reasonably be interpreted only in the dummy case, i.e. if the observed strike prices k_j are equidistant and if the distances between the neighboring observed strike prices are equal to one.

To keep the interpretation of the parameters β_j as the derivatives of the estimated function, the form $\boldsymbol{\mu} = \boldsymbol{\Delta}\boldsymbol{\beta}$ will be kept, and the following design matrix $\boldsymbol{\Delta}$ should be used:

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & \Delta_p^1 & \Delta_{p-1}^1 & \Delta_{p-2}^1 & \cdots & \Delta_3^1 & \Delta_2^1 \\ 1 & \Delta_p^2 & \Delta_{p-1}^2 & \Delta_{p-2}^2 & \cdots & \Delta_3^2 & 0 \\ \vdots & & & & & & \vdots \\ 1 & \Delta_p^{p-2} & \Delta_{p-1}^{p-2} & 0 & \cdots & 0 & 0 \\ 1 & \Delta_p^{p-1} & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{2.5}$$

where $\Delta_j^i = \max(k_j - k_i, 0)$ denotes the positive part of the distance between k_i and k_j, i.e. the i-th and the j-th ($1 \leq i \leq j \leq p$) sorted distinct observed values of the strike price.

The vector of conditional means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)^\top$ can be written in terms of the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$ using the design matrix $\boldsymbol{\Delta}$ defined in (2.5) as follows

$$\boldsymbol{\mu} = \boldsymbol{\Delta}\boldsymbol{\beta},$$

which in the matrix form is

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} 1 & \Delta_p^1 & \Delta_{p-1}^1 & \Delta_{p-2}^1 & \cdots & \Delta_3^1 & \Delta_2^1 \\ 1 & \Delta_p^2 & \Delta_{p-1}^2 & \Delta_{p-2}^2 & \cdots & \Delta_3^2 & 0 \\ \vdots & & & & & & \vdots \\ 1 & \Delta_p^{p-2} & \Delta_{p-1}^{p-2} & 0 & \cdots & 0 & 0 \\ 1 & \Delta_p^{p-1} & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}.$$

The constraints on the conditional means μ_j can now be expressed as constraints on the parameters β_j in (2.4), i.e. it suffices to request that $\beta_j > 0$ for $j = 0, \ldots, p-1$ and that $\sum_{j=2}^{p-1} \beta_j \leq 1$.

Finally, the linear model for the option prices $\mathbf{C}$ can now be written as

$$\mathbf{C}(\boldsymbol{\mathcal{K}}) = \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.6}$$

where $\boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}$ is the design matrix obtained by repeating each row of matrix $\boldsymbol{\Delta}$ n_i-times for $i = 1, \ldots, p$.

2.5 Kalman filter for a CALL option

Having defined a state space model and its *measurement equations* (1.1) and *transition equations* (1.2), derived the Kalman filter, defined a CALL option and its *state price density* (SPD), constructed the estimate and the linear model in the previous chapters and sections, we are now able to use the recursive Kalman algorithm and try to estimate the risk neutral price density using real data.

It shall be briefly reminded that – generally – a time series $\mathbf{y}_t$, $t = 1, \ldots, T$ is being observed. The unobservable variables are related to a vector $\boldsymbol{\xi}_t$, already known as the *state vector*. The measurement and the transition equations have been defined and discussed in detail earlier.

Consequently, all these equations have to be rewritten so as to be corresponding with the linear model defined in (2.6).

2.5.1 Measurement equation

The measurement equation

$$\mathbf{y}_t = \mathbf{a}_t(\boldsymbol{\psi}) + \mathbf{B}_t(\boldsymbol{\psi})\boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t(\boldsymbol{\psi})$$

has to be modified.

- The vector $\mathbf{y}_t$ (generally) containing observed elements will be replaced by the vector $\mathbf{C}_t(\boldsymbol{\mathcal{K}}) = (\mathbf{C}_1, \ldots, \mathbf{C}_T)^\top$ containing the observed option prices $\mathbf{C}_i = \mathbf{C}_{t,i}(\mathbf{K}_i, T)$, where $\mathbf{C}_i$ is the corresponding observed option price by the i-th observation of the strike price $\mathbf{K}_i$ for $i = 1, \ldots, p$.

- We do not consider any additive component $\mathbf{a}_t(\boldsymbol{\psi})$, therefore $\mathbf{a}_t(\boldsymbol{\psi})$ is a vector of zeros.

- The known multiplicative matrix $\mathbf{B}_t(\boldsymbol{\psi})$ will be replaced by the design matrix $\boldsymbol{\Delta}$ defined in (2.5), or by the matrix $\boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}$ from (2.6), respectively. In the following, we omit the symbol $\boldsymbol{\Delta}$ from the notation $\boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}$.

- The state vector $\boldsymbol{\xi}_t$ becomes the unobserved variables $\beta_0, \beta_1, \ldots, \beta_{p-1}$ related to the vector $\boldsymbol{\beta}_t = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$, which is to be estimated.

- We further assume the normal distribution for the serially uncorrelated, centered error term $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$. The variance-covariance matrix $\mathrm{Var}[\boldsymbol{\epsilon}_t] = \mathbf{H}_t(\boldsymbol{\psi})$ will now be known, assumed to be diagonal and heteroscedastic (variances depending on the strike prices $\boldsymbol{\mathcal{K}}$).

In the further Kalman filter determination for a CALL option we also omit the symbols $\boldsymbol{\psi}$ from the notation. Thus, the following can be written:

$$\begin{aligned} \mathbf{y}_t &\rightsquigarrow \mathbf{C}_t(\boldsymbol{\mathcal{K}}), \\ \mathbf{B}_t(\boldsymbol{\psi}) &\rightsquigarrow \boldsymbol{\mathcal{X}}, \\ \boldsymbol{\xi}_t &\rightsquigarrow \boldsymbol{\beta}_t, \\ \mathbf{a}_t(\boldsymbol{\psi}) &= \mathbf{0}^\top, \\ \boldsymbol{\epsilon}_t(\boldsymbol{\psi}) &\sim \mathrm{N}\big[\mathbf{0}, \mathbf{H}_t(\boldsymbol{\mathcal{K}})\big], \end{aligned}$$

with the covariance matrix

$$\mathbf{H}_t(\boldsymbol{\mathcal{K}}) = \begin{pmatrix} \boldsymbol{\Sigma}_1^2 & 0 & \cdots & & & & & \cdots & 0 \\ 0 & \ddots & & & & & & & \vdots \\ \vdots & & \boldsymbol{\Sigma}_1^2 & & & & & & \\ & & & \boldsymbol{\Sigma}_2^2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \boldsymbol{\Sigma}_2^2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \boldsymbol{\Sigma}_p^2 & \vdots \\ \vdots & & & & & & & \ddots & 0 \\ 0 & \cdots & & & & & & \cdots\ 0 & \boldsymbol{\Sigma}_p^2 \end{pmatrix}. \tag{2.7}$$

2.5.2 Transition equation

Then, the transition equation

$$\boldsymbol{\xi}_t = \boldsymbol{\Phi}_t(\boldsymbol{\psi})\boldsymbol{\xi}_{t-1} + \mathbf{c}_t(\boldsymbol{\psi}) + \boldsymbol{\eta}_t(\boldsymbol{\psi})$$

is to be by analogy modified as well.

- The known *transition matrix* $\boldsymbol{\Phi}(\boldsymbol{\psi})$ will be replaced by the identity matrix $\boldsymbol{\mathcal{I}}$.
- We do not consider any additive component $\mathbf{c}_t(\boldsymbol{\psi})$, therefore $\mathbf{c}_t(\boldsymbol{\psi})$ is a vector of zeros.
- We further assume the normal distribution for the Gaussian noise term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$, which is centered and has a known variance matrix $\mathrm{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t(\boldsymbol{\psi})$. This matrix $\mathbf{Q}_t(\boldsymbol{\psi})$ is assumed to be diagonal and on the top of that homoscedastic with a variance of σ^2, i.e. $\mathbf{Q}_t(\boldsymbol{\psi}) = \sigma^2\boldsymbol{\mathcal{I}}$.

Therefore, the following can be written:

$$\begin{aligned} \boldsymbol{\Phi_t}(\boldsymbol{\psi}) &\rightsquigarrow \boldsymbol{\mathcal{I}}, \\ \mathbf{c_t}(\boldsymbol{\psi}) &= \mathbf{0}^\top, \\ \boldsymbol{\eta_t}(\boldsymbol{\psi}) &\sim \mathrm{N}\big[\mathbf{0}, \sigma^2\boldsymbol{\mathcal{I}}\big]. \end{aligned}$$

We keep the same assumptions (about distribution, uncorrelation, etc., see more in Subsection 1.3.2) for the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ and the initial state vector $\boldsymbol{\beta}_0$.

Finally, we are able to write the measurement and the transition equations for the option prices in the following form:

$$\mathbf{C}_t(\boldsymbol{\mathcal{K}}) = \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \tag{2.8}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{2.9}$$

where $\boldsymbol{\mathcal{X}}$ is the matrix from (2.6), $\boldsymbol{\epsilon}_t \sim \mathrm{N}\big[\mathbf{0}, \mathbf{H}_t(\boldsymbol{\mathcal{K}})\big]$, with $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ defined in (2.7), and $\boldsymbol{\eta}_t \sim \mathrm{N}\big[\mathbf{0}, \sigma^2\boldsymbol{\mathcal{I}}\big]$. These two error terms are uncorrelated with each other in all time periods. Such model corresponds with linear model defined in (2.6).

In the following, the Kalman filter for a CALL option can be determined.

2.5.3 Kalman filter for a CALL option, MMSE

Even in the case of a minimum mean squared error estimator (MMSE), it will be assumed again that the additive error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are independently and normally distributed. Furthermore, the error terms are treated as independent of the initial state vector $\boldsymbol{\beta}_0$, which is assumed to be normally distributed with $\mathrm{E}[\boldsymbol{\beta}_0] = \boldsymbol{\beta}_{0|0}$ and $\mathrm{Cov}[\boldsymbol{\beta}_0] = \boldsymbol{\Sigma}_{0|0}$. Therefore, because the transition equation (2.9) is linear in $\boldsymbol{\beta}_{t-1}$ and the error term $\boldsymbol{\eta}_t$ is normally distributed, the state vector $\boldsymbol{\beta}_t$ – as the sum of $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{\eta}_t$ – is also normally distributed. Furthermore, for the measurement equation (2.8) we also have (according to Dupač and Hušková [4], Theorem 3.16 and its version for random vectors) normally distributed option prices $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$, since $\boldsymbol{\beta}_t$ and $\boldsymbol{\epsilon}_t$ are both normal.

In the following, the prediction and the updating step of the general Kalman filter can be followed, re-calculated and re-notated for the case of an European CALL option. The main results and remarks will be mentioned in the sense of a CALL option again.

1. Prediction Step: First, the CALL option model of (2.8) and (2.9) is considered. Let $\boldsymbol{\beta}_{t-1|t-1}$ denote the optimal estimator of the state vector $\boldsymbol{\beta}_{t-1}$ based on the observations up to and including $\mathbf{C}_{t-1}(\boldsymbol{\mathcal{K}})$. Therefore, the history of the observations up to time $t-1$ related to a sample vector $\mathbf{C}_{t-1}(\boldsymbol{\mathcal{K}}) = \big(\mathbf{c}_1(\boldsymbol{\mathcal{K}}), \ldots, \mathbf{c}_{t-1}(\boldsymbol{\mathcal{K}})\big)$ is also considered.

In the *prediction step*, we forecast the state vector $\boldsymbol{\beta}_t$ by calculating the conditional mean of the state variables on both sides of equation (2.9) given the information up to time $t-1$:

$$\begin{aligned}\boldsymbol{\beta}_{t|t-1} &\equiv \mathrm{E}[\boldsymbol{\beta}_t|\mathcal{F}_{t-1}] \\ &= \mathrm{E}[\boldsymbol{\beta}_{t-1}+\boldsymbol{\eta}_t|\mathcal{F}_{t-1}] \\ &= \boldsymbol{\beta}_{t-1|t-1}. \end{aligned} \tag{2.10}$$

Hence, $\boldsymbol{\beta}_t$ in the next step is the same as in the previous one. Given $\boldsymbol{\beta}_{t-1|t-1}$ we have the optimal estimator $\boldsymbol{\beta}_{t|t-1}$ of the state vector $\boldsymbol{\beta}_t$ (still based just on the observations up to time $t-1$).

Once more, we consider $\boldsymbol{\beta}_{t-1|t-1}$ as the optimal estimator of the state vector $\boldsymbol{\beta}_{t-1}$. The *a priori estimate error* is:

$$\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}_{t-1|t-1}.$$

Let $\boldsymbol{\Sigma}_{t-1|t-1}$ denote the covariance matrix of this estimation error, i.e.

$$\boldsymbol{\Sigma}_{t-1|t-1} = \mathrm{E}\left[\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}_{t-1|t-1}\right)\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}_{t-1|t-1}\right)^\top \middle| \mathcal{F}_{t-1}\right].$$

Given this covariance matrix $\boldsymbol{\Sigma}_{t-1|t-1}$, the corresponding variance-covariance matrix for the state variables is given by

$$\begin{aligned}\boldsymbol{\Sigma}_{t|t-1} &\equiv \mathrm{E}\left[\left(\boldsymbol{\beta}_t - \mathrm{E}[\boldsymbol{\beta}_t|\mathcal{F}_{t-1}]\right)\left(\boldsymbol{\beta}_t - \mathrm{E}[\boldsymbol{\beta}_t|\mathcal{F}_{t-1}]\right)^\top \middle| \mathcal{F}_{t-1}\right] \\ &= \mathrm{E}\left[\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1}\right)\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t-1}\right)^\top \middle| \mathcal{F}_{t-1}\right] \\ &= \mathrm{E}\left[\left(\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t - \boldsymbol{\beta}_{t-1|t-1}\right)\left(\boldsymbol{\beta}_{t-1}^\top + \boldsymbol{\eta}_t^\top - \boldsymbol{\beta}_{t-1|t-1}^\top\right) \middle| \mathcal{F}_{t-1}\right] \\ &= \mathrm{E}\left[\boldsymbol{\beta}_{t-1|t-1}\boldsymbol{\beta}_{t-1|t-1}^\top \middle| \mathcal{F}_{t-1}\right] + \mathrm{E}\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^\top \middle| \mathcal{F}_{t-1}\right] \\ &= \boldsymbol{\Sigma}_{t-1|t-1} + \sigma^2\boldsymbol{\mathcal{I}}. \end{aligned} \tag{2.11}$$

Thus, the *a priori estimate error covariance matrix* $\boldsymbol{\Sigma}_{t|t-1}$ and equations (2.10) and (2.11) as the *prediction equations* of the Kalman filter have been determined.

Next, an intermediate result, the innovation (or prediction error) denoted by $\mathbf{I}_t$ can be re-defined, and using the given data for the measurable observations $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ can be obtained:

$$\begin{aligned}\mathbf{I}_t &= \mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \mathbf{C}_{t|t-1}(\boldsymbol{\mathcal{K}}) \\ &= \mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \mathrm{E}\left[\mathbf{C}_t(\boldsymbol{\mathcal{K}})|\mathcal{F}_{t-1}\right] \\ &= \mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}. \end{aligned}$$

These residuals $\big($differences $\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}\big)$ represent – as explained earlier – the new information in the latest observation. We just remind, as can be seen further in the updating step, they play a key role in updating the estimator of the state vector $\boldsymbol{\beta}_t$. Residuals reflect the discrepancy between the predicted measurement $\boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}$ and the actual measurement $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$. A residual of zero means that the two are in complete agreement – therefore, the prediction error $\mathbf{I}_t$ is equal to zero. The further $\mathbf{I}_t$ is from a null vector, the greater is the "correction" in the estimator of $\boldsymbol{\beta}_t$.

Remark 2.5.1. *Unfortunately, this approach cannot be used in our data analysis: in every time step t we get just one observed CALL option price $C_t = C_{t,i}(k_i)$ with just one corresponding strike price k_i. Therefore C_t and $\mathbf{I}_t$ become one-dimensional and the corresponding equations must be modified. Please see more in Chapter 5.*

In the next step, the variance-covariance matrix of this prediction error $\mathbf{I}_t$ can be re-calculated as follows:

$$\begin{aligned}
\mathbf{F}_{t|t-1} &= \mathrm{Cov}\left[\mathbf{I}_t|\mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\left[\big(\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \mathrm{E}[\mathbf{C}_t(\boldsymbol{\mathcal{K}})|\mathcal{F}_{t-1}]\big)\big(\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \mathrm{E}[\mathbf{C}_t(\boldsymbol{\mathcal{K}})|\mathcal{F}_{t-1}]\big)^\top\Big|\mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\left[\Big(\boldsymbol{\mathcal{X}}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}\Big)\Big(\boldsymbol{\beta}_t^\top\boldsymbol{\mathcal{X}}^\top + \boldsymbol{\epsilon}_t^\top - \boldsymbol{\beta}_{t|t-1}^\top\boldsymbol{\mathcal{X}}^\top\Big)\Big|\mathcal{F}_{t-1}\right] \\
&= \boldsymbol{\mathcal{X}}\,\mathrm{E}\left[\boldsymbol{\beta}_{t|t-1}\boldsymbol{\beta}_{t|t-1}^\top|\mathcal{F}_{t-1}\right]\boldsymbol{\mathcal{X}}^\top + \mathrm{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top|\mathcal{F}_{t-1}\right] \\
&= \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top + \mathbf{H}_t(\boldsymbol{\mathcal{K}}), \qquad (2.12)
\end{aligned}$$

where the heteroscedastic matrix $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ is defined in (2.7).

The innovations $\mathbf{I}_t$ are still centered, uncorrelated random variables and variables $\mathbf{F}_{t|t-1}$ are their variances, i.e. $\mathrm{Var}[\mathbf{I}_t] = \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top + \mathbf{H}_t(\boldsymbol{\mathcal{K}})$. This holds again even in the absence of the normality assumption. In addition to that, in a Gaussian model, it can be shown, for example in Harvey [13], that the joint density of the observations can be decomposed in terms of the innovations, which are independently and normally distributed. Hence,

$$\mathbf{I}_t \sim \mathrm{N}\big[\mathbf{0}, \mathbf{F}_{t|t-1}\big].$$

2. Updating Step: According to the general Kalman filter, once a new observation $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ becomes available, the estimator $\boldsymbol{\beta}_{t|t-1}$ of $\boldsymbol{\beta}_t$ can be updated. In the *updating step* we update again the inference on $\boldsymbol{\beta}_{t|t-1}$ by including the newly available information at time t; this results in the filtered estimate $\boldsymbol{\beta}_{t|t}$. In this way, the *Kalman gain* $\mathbf{K}_t$ will be re-defined.

We do not need to follow the whole updating procedure $\big($construction the joint distribution of $\boldsymbol{\beta}_t$ and $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$, etc.$\big)$, and to obtain the *updating equations* of the Kalman filter for a CALL option, we immediately use the formulas determined earlier, which are just to be recalculated.

The following relationships can be obtained, where $\mathbf{I}_t$ denotes the innovations (prediction error) defined above:

$$\begin{aligned}\boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top \mathbf{F}_{t|t-1}^{-1}\mathbf{I}_t, \\ \boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top \mathbf{F}_{t|t-1}^{-1}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{t|t-1}.\end{aligned}$$

The *Kalman gain* $\mathbf{K}_t$ can be defined again by substituting:

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top \mathbf{F}_{t|t-1}^{-1}. \tag{2.13}$$

These steps finally result in the **update** for the state vector

$$\boldsymbol{\beta}_{t|t} = \mathrm{E}[\boldsymbol{\beta}_t|\mathcal{F}_t] = \boldsymbol{\beta}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, \tag{2.14}$$

and its variance-covariance matrix

$$\begin{aligned}\boldsymbol{\Sigma}_{t|t} &= \mathrm{E}\left[\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t}\right)\left(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t}\right)^\top \middle| \mathcal{F}_t\right] \\ &= \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{t|t-1} \\ &= \left(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\boldsymbol{\mathcal{X}}\right)\boldsymbol{\Sigma}_{t|t-1}. \end{aligned} \tag{2.15}$$

Thus, the *a posteriori estimate error covariance matrix* $\boldsymbol{\Sigma}_{t|t}$ and the *updating equations* of the Kalman filter for a CALL option have been determined.

Summary Taken together, the modified prediction equations (2.10) with (2.11) and the modified updating equations (2.14) with (2.15), i.e. equations

$$\begin{aligned}\boldsymbol{\beta}_{t|t-1} &= \boldsymbol{\beta}_{t-1|t-1}, \\ \boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Sigma}_{t-1|t-1} + \sigma^2\boldsymbol{\mathcal{I}},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, \\ \boldsymbol{\Sigma}_{t|t} &= \left(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\boldsymbol{\mathcal{X}}\right)\boldsymbol{\Sigma}_{t|t-1},\end{aligned}$$

where $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top \mathbf{F}_{t|t-1}^{-1}$ is the Kalman gain, make up the Kalman filter for an European CALL option. If desired they can again be written as a single set of recursions, which will be discussed in detail in the following subsection.

2.5.4 A CALL option, Kalman filter and its recursions

From the general version of the Kalman filter we already know that – if desired – the prediction equations (2.10) with (2.11) and the updating equations (2.14) with (2.15) can be written as a single set of recursions (now going directly from $\boldsymbol{\beta}_{t-1|t-1}$ to $\boldsymbol{\beta}_{t|t-1}$, etc.). This approach will be discussed now.

Having derived the Kalman filter algorithm for a CALL option in two different ways yielding the MMSE (or eventually the MMSLE) described in the previous subsections, the way is similarly opened to apply the sequence of filtering equations **recursively** as each new observation $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ becomes available.

Understandably, the Kalman filter for a CALL option works the same as it has been described in the general case. However, some remarks shall be mentioned again.

We discussed that the Kalman filter estimates a process by using a form of feedback control, i.e. the filter estimates the process state at some time t and then obtains feedback in the form of noisy measurements. As such, the equations for the Kalman filter fall into two groups, as described above: the *prediction equations* (or *time update equations*) and the *updating equations* (or *measurement update equations*).

- The first group of these equations, the ***time*** *update equations*, is responsible for projecting forward (in the sense of the ***time***) the current state vector $\boldsymbol{\beta}_t$ and error covariance estimates $\boldsymbol{\Sigma}_t$ to obtain the *a priori estimates* for the next time step $t \rightarrow t+1$.

- The second group, the ***measurement*** *update equations*, is responsible for the feedback – i.e. for incorporating a new ***measurement*** $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ into the a priori estimate to obtain an improved *a posteriori estimate*.

It has been mentioned that the time update equations can also be thought of as *predictor equations*, while the measurement update equations can be thought of as *corrector equations*. The final core estimation algorithm for a CALL option is shown in Figure 2.4.

In detail, please follow the flowchart of Figure 2.5 for a better, graphical illustration of the Kalman filter algorithm for a CALL option. We try to explain again – now in this special case – how the algorithm works "step-by-step".

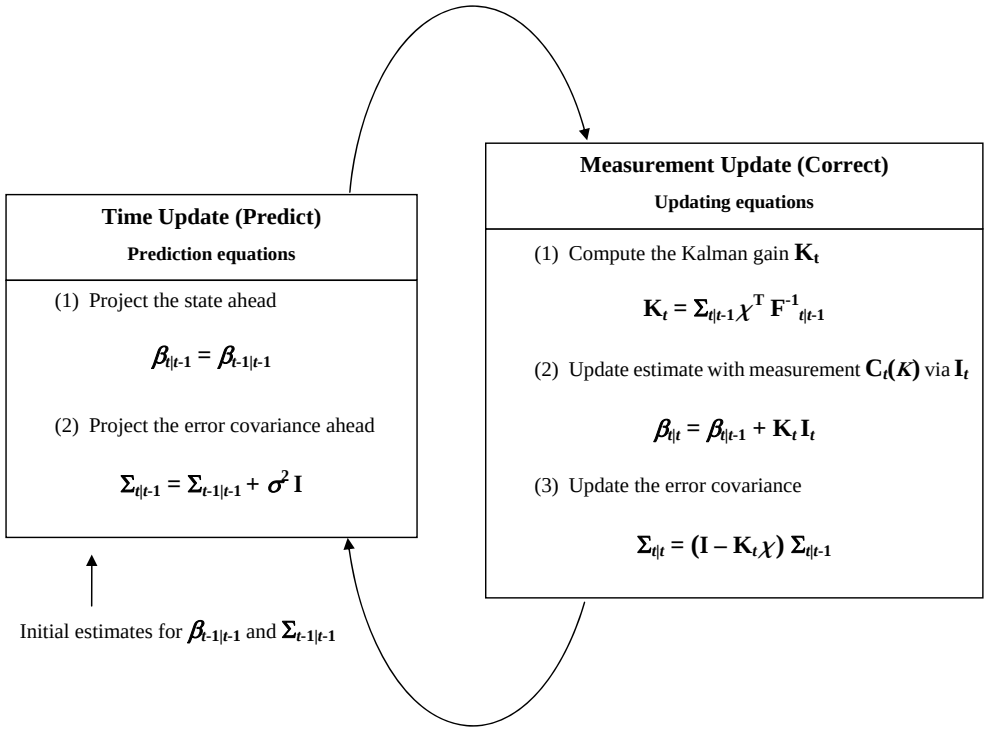


Figure 2.4: A complete picture of the operation of the Kalman filter for a CALL option, using the measurement and prediction equations.

The recursive version of the Kalman filter for a CALL option works fully in the following steps:

1. **Initialization step:** The recursive algorithm starts with a choice of the parameter $\boldsymbol{\psi}$ and corresponding initial state vector $\boldsymbol{\beta}_0$ with the initial values: the mean of $\boldsymbol{\beta}_{0|0}$ and the variance-covariance matrix of $\mathbf{\Sigma}_{0|0}$.

2. **Prediction step, a priori estimates:** Thereupon, we use the prediction equations to calculate the a priori estimates

$$\begin{aligned} \boldsymbol{\beta}_{t|t-1} &= \boldsymbol{\beta}_{t-1|t-1}, \text{ and} \\ \mathbf{\Sigma}_{t|t-1} &= \mathbf{\Sigma}_{t-1|t-1} + \sigma^2\boldsymbol{\mathcal{I}}. \end{aligned}$$

3. **The prediction error, its MSE:** Using the current market information on observable variables $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ on date $t = 1$, the prediction error and its corresponding MSE will be derived as follows

$$\begin{aligned} \mathbf{I}_t &= \mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}, \text{ and} \\ \mathbf{F}_{t|t-1} &= \boldsymbol{\mathcal{X}}\mathbf{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^{\top} + \mathbf{H}_t(\boldsymbol{\mathcal{K}}). \end{aligned}$$

4. **Updating step, the filtered estimates:** Given these intermediate results we then update the a priori estimates yielding the filtered estimates

$$\begin{aligned} \boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \mathbf{K}_t \mathbf{I}_t, \\ \boldsymbol{\Sigma}_{t|t} &= (\boldsymbol{\mathcal{I}} - \mathbf{K}_t \boldsymbol{\mathcal{X}})\, \boldsymbol{\Sigma}_{t|t-1}, \end{aligned}$$

   as optimal in the sense of MMSE or MMSLE using the Kalman gain $\mathbf{K}_t$ for a CALL option defined in (2.13) as $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{\mathcal{X}}^\top \mathbf{F}_{t|t-1}^{-1}$.

5. **Recursive step:** Now for each discrete time step $t \to t+1$, we recursively feed the results for the state vector $\boldsymbol{\beta}_{t|t}$ and its variance-covariance matrix $\boldsymbol{\Sigma}_{t|t}$ from the updating equations into the prediction equations until we reach the last observations $\mathbf{C}_T(\boldsymbol{\mathcal{K}})$ at time T.[2]

6. **Likelihood:** The conditional likelihood function is evaluated at the current parameter values using the results of the Kalman recursions.

7. **Criterion met?** Finally, with the obtained time-series of the state variables $\boldsymbol{\beta}_t$ we can evaluate the likelihood function (see more in Kellerhals [15]) to choose other more appropriate values for the parameters $\boldsymbol{\psi}$. The Kalman algorithm is run until we reach the predefined abortion criterion.

[2]Please note again, these are then formally indexed with $t-1$, since we denote the actual information we are working with by the time index t.

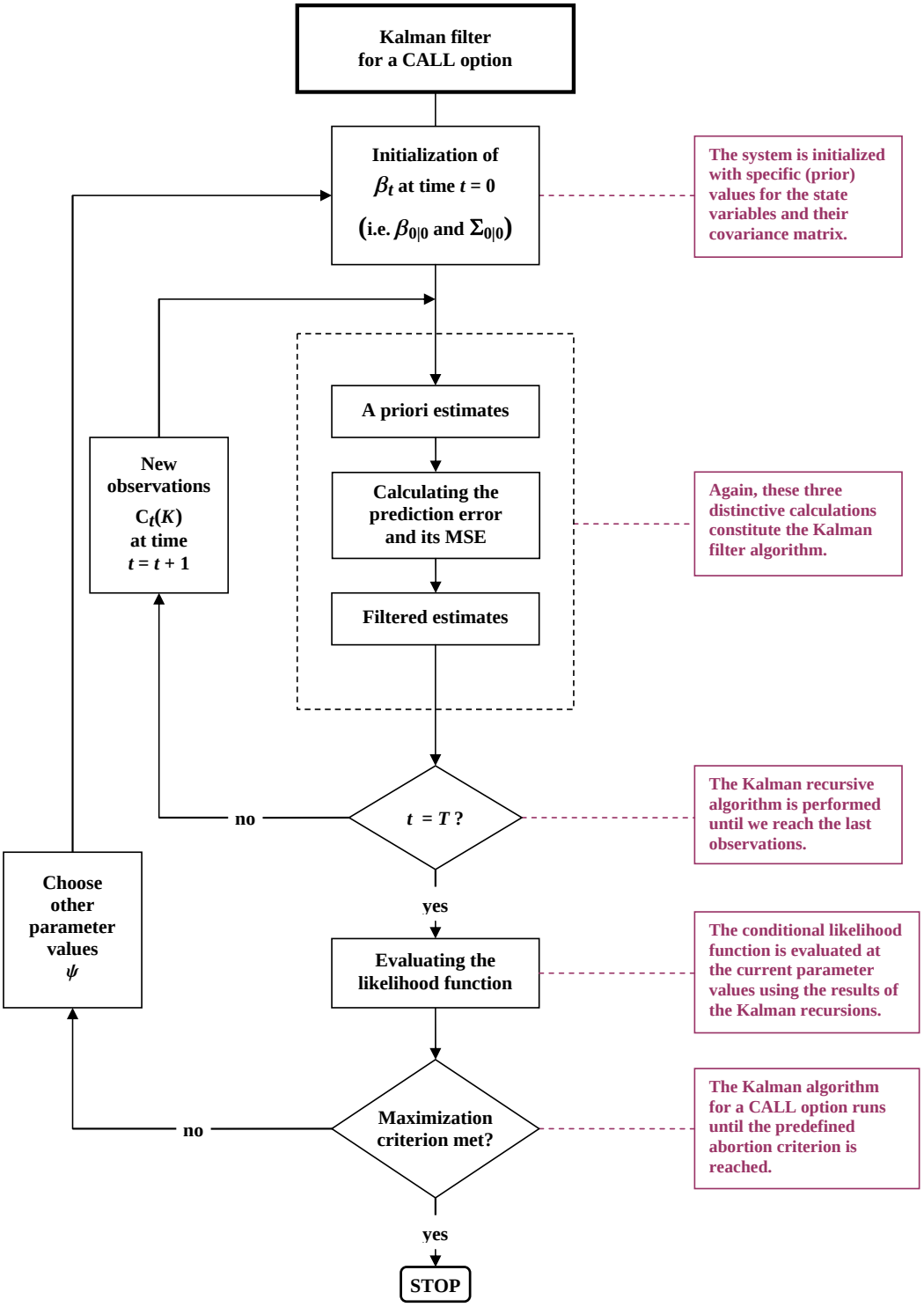


Figure 2.5: Flowchart of the Kalman filter and MLE for a CALL option.

2.5.5 Kalman filter for a CALL option, MMSLE

According to the general version of the Kalman filter as the minimum mean square linear estimator (MMSLE), the Kalman filter for a CALL option without requiring the normality assumption for the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ will in the following be derived.

1. Prediction Step: First, the state space model of (2.8) and (2.9) for a CALL option is considered. The *prediction equations* (2.10) and (2.11) are obtained similarly to the previous derivation by taking the conditional expectation and variance given the information up to time $t-1$.

2. Updating Step: To obtain the *updating equations*, we again start with stating a general linear relationship of the updated estimator $\boldsymbol{\beta}_{t|t}$ and the information of the present sample $\mathbf{C}_t(\boldsymbol{\mathcal{K}})$ and the past information $\mathcal{F}_{t-1}$. Similarly to the previous step, we do not follow the complete determining procedure (solving the minimization problem), and we just recalculate the Kalman gain $\mathbf{K}_t$ for our case.

The new Kalman gain matrix $\mathbf{K}_t$ can be written as

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top \left(\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\mathcal{X}}^\top + \mathbf{H}_t(\boldsymbol{\mathcal{K}})\right)^{-1}, \tag{2.16}$$

which is equivalent to the expression derived in equation (2.15) for the case of the Kalman filter resulting in the MMSE in the previous subsections.

Chapter 3

Extended Kalman filter

3.1 Non-linear state space model

In the previous chapter we discussed the case of linear measurement and transition equations and normally distributed error terms as specified in equations (1.1) and (1.2). However, the most realistic applications derived from financial theory exhibit **non-linear** functional relationships of non-normally distributed state variables in the measurement and transition equations. A somewhat different kind of non-linearity is easily obtained when the observations in the measurement equation are no longer a linear function of the state vector $\boldsymbol{\xi}_t$ and, in the transition equation, the state vector itself is no longer a linear function of the state vector in the previous time period.

We will refer to such models as being *functionally non-linear*. They are not – in general – conditionally Gaussian. In the absence of this property, it is necessary to resort to approximate filters and the most basic of these, the *extended Kalman filter*, is described in the following section. In such cases **explicit expressions** for the filtering algorithms cannot be derived – and some approximations are necessary for the estimation procedure.

There are two main approaches to obtain a non-linear filtering algorithm.

1. The first approach is to approximate the non-linear measurement and transition equations. These linearized non-linear functions are then applied to a modification of the linear Kalman filter algorithm as derived in the previous chapters.
2. The second types of algorithms can be summarized under the approach of approximating the underlying density functions of the state vector. Then a recursive algorithm on the densities is derived using Bayes's formula. The advantage of this second approach is that it results in asymptotically unbiased Kalman filtering estimates. However, the estimators based on the density approach require a great amount of computational burden compared with those based on the Taylor series approximations. Please see more in Kellerhals [15].

Later in this thesis, our analysis of the Kalman filter will be constrained on algorithms derived from the Taylor series expansions. We shall start with the following *general state space model* definition.

Definiton 3.1.1 (Non-linear State Space Model)**.** *We treat the non-linear filtering problem based on a state space model with the measurement and transition equations being specified as*

$$\begin{aligned} \mathbf{y}_t &= \mathbf{g}_t\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big), && (3.1)\\ \boldsymbol{\xi}_t &= \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big), && (3.2) \end{aligned}$$

with the terms $\mathbf{g}_t\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ *and* $\mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ *denoting the possible non-linear functional relationships. The error terms* $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ *and* $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ *represent the process and measurement noise and are assumed to follow the properties*

$$\begin{aligned} \mathrm{E}\begin{bmatrix} \boldsymbol{\epsilon}_t(\boldsymbol{\psi}) \\ \boldsymbol{\eta}_t(\boldsymbol{\psi}) \end{bmatrix} &= \mathbf{0}^\top, \\ \mathrm{Var}\begin{bmatrix} \boldsymbol{\epsilon}_t(\boldsymbol{\psi}) \\ \boldsymbol{\eta}_t(\boldsymbol{\psi}) \end{bmatrix} &= \begin{bmatrix} \mathbf{H}_t(\boldsymbol{\psi}) & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_t(\boldsymbol{\psi}) \end{bmatrix}, \end{aligned}$$

as normally distributed random vectors.

In this case the **non-linear** function $\mathbf{g}_t$ in the measurement equation (3.1) relates the state $\boldsymbol{\xi}_t$ to the measurement $\mathbf{y}_t$. The **non-linear** function $\mathbf{h}_t$ in the transition equation (3.2) relates the state $\boldsymbol{\xi}_{t-1}$ at the previous time step $t-1$ to the state at the current time step t.

Example 3.1.2. *The simplest models (we do not consider any dependence on* $\boldsymbol{\psi}$*) can easily be written in the following form*

$$\begin{aligned} \mathbf{y}_t &= \mathbf{g}_t\big(\boldsymbol{\xi}_t\big) + \boldsymbol{\epsilon}_t. \\ \boldsymbol{\xi}_t &= \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}\big) + \boldsymbol{\eta}_t. \end{aligned}$$

Even under the assumption that $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are normally distributed, obtaining an optimal filter for a model of this kind is not possible. In the following, we try to obtain an approximate filter by linearizing the model and then applying a modification of the usual Kalman filter.

3.2 Taylor series expansion

To estimate a process with non-linear measurement and transition relationships, we begin by writing new governing equations that linearize an estimate about (3.1) and (3.2). We briefly discuss the Taylor series expansion, which is needed for the case of the extended Kalman filter: if the non-linear functions $\mathbf{g}_t(\cdot)$ and $\mathbf{h}_t(\cdot)$ from the previous Example 3.1.2 are sufficiently smooth, they can be expanded in Taylor series around the conditional means $\boldsymbol{\xi}_{t|t-1}$ and $\boldsymbol{\xi}_{t-1|t-1}$, to give

$$\begin{aligned}\mathbf{g}_t\big(\boldsymbol{\xi}_t\big) &\approx \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}\big) + \mathbf{B}_{t|t-1}(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}),\\ \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}\big) &\approx \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}\big) + \boldsymbol{\Phi}_{t|t-1}(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}),\end{aligned}$$

with

$$\mathbf{B}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t\big(\boldsymbol{\xi}_t\big)}{\partial \boldsymbol{\xi}_t^\top}\right|_{\boldsymbol{\xi}_t=\boldsymbol{\xi}_{t|t-1}},$$

$$\boldsymbol{\Phi}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}\big)}{\partial \boldsymbol{\xi}_{t-1}^\top}\right|_{\boldsymbol{\xi}_{t-1}=\boldsymbol{\xi}_{t-1|t-1}}.$$

Thus, $\mathbf{B}_{t|t-1}$ is the Jacobian matrix of partial derivatives of $\mathbf{g}_t$ with respect to $\boldsymbol{\xi}_t$ and $\boldsymbol{\Phi}_{t|t-1}$ is the Jacobian matrix of partial derivatives of $\mathbf{h}_t$ with respect to $\boldsymbol{\xi}_{t-1}$.

In detail: if the non-linear functions $\mathbf{g}_t(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi})$ and $\mathbf{h}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi})$ in the general form in (3.1) and (3.2) are smooth enough, they are approximated around the conditional means and error terms, i.e. around the vectors $(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0})$ and $(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0})$, as follows:

1. The measurement equation: The *measurement equation* is approximated via the first order Taylor series expansions around the vectors $(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0})$, which results in the approximate expression

$$\mathbf{g}_t\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big) \approx \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) + \mathbf{B}_{t|t-1}(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}) + \mathbf{R}_{t|t-1}\boldsymbol{\epsilon}_t,$$

with the corresponding Jacobian matrices

$$\mathbf{B}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\xi}_t^\top}\right|_{(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi})},$$

$$\mathbf{R}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\epsilon}_t^\top}\right|_{(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi})}.$$

2. The transition equation: The *transition equation* is approximated around $(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0})$ which results in:

$$\mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big) \approx \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) + \boldsymbol{\Phi}_{t|t-1}(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}) + \mathbf{S}_{t|t-1}\boldsymbol{\eta}_t,$$

with the Jacobian matrices

$$\boldsymbol{\Phi}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\xi}_{t-1}^\top}\right|_{(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi})},$$

$$\mathbf{S}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\eta}_t^\top}\right|_{(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi})}.$$

3.3 Extended Kalman filter

According to the previous section, substituting the Taylor series expansions of the functions $\mathbf{g}_t\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ and $\mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ in both general state space model equations (3.1) and (3.2), and assuming knowledge of the estimators $\boldsymbol{\xi}_{t|t-1}$ and $\boldsymbol{\xi}_{t-1|t-1}$ leads us to approximate the original non-linear model by

$$\mathbf{y}_t \approx \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) + \mathbf{B}_{t|t-1}(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|t-1}) + \mathbf{R}_{t|t-1}\boldsymbol{\epsilon}_t, \tag{3.3}$$

with

$$\mathbf{B}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\xi}_t^\top}\right|_{(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi})},$$

$$\mathbf{R}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\epsilon}_t^\top}\right|_{(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi})}$$

for the *measurement equation.*

The *transition equation* is approximated via $\mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ which results in:

$$\boldsymbol{\xi}_t \approx \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) + \boldsymbol{\Phi}_{t|t-1}(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|t-1}) + \mathbf{S}_{t|t-1}\boldsymbol{\eta}_t, \tag{3.4}$$

with

$$\boldsymbol{\Phi}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\xi}_{t-1}^\top}\right|_{(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi})},$$

$$\mathbf{S}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi})}{\partial \boldsymbol{\eta}_t^\top}\right|_{(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t, \boldsymbol{\psi}) = (\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi})}.$$

The *non-linear state space model* of equations (3.3) and (3.4) can further be stated as

$$\mathbf{y}_t \approx \mathbf{a}_t + \mathbf{B}_{t|t-1}\boldsymbol{\xi}_t + \mathbf{R}_{t|t-1}\boldsymbol{\epsilon}_t, \tag{3.5}$$
$$\boldsymbol{\xi}_t \approx \mathbf{c}_t + \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\xi}_{t-1} + \mathbf{S}_{t|t-1}\boldsymbol{\eta}_t, \tag{3.6}$$

where $\mathbf{a}_t$ and $\mathbf{c}_t$ are defined by

$$\mathbf{a}_t = \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) - \mathbf{B}_{t|t-1}\boldsymbol{\xi}_{t|t-1}, \text{ and} \tag{3.7}$$
$$\mathbf{c}_t = \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi}\big) - \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\xi}_{t-1|t-1}, \tag{3.8}$$

respectively. It follows the treatment of the linear state space model of equations (1.1) and (1.2).

We again consider the simpler case of the non-linear state space model from the Example 3.1.2, i.e. the model

$$\mathbf{y}_t = \mathbf{g}_t\big(\boldsymbol{\xi}_t\big) + \boldsymbol{\epsilon}_t,$$
$$\boldsymbol{\xi}_t = \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}\big) + \boldsymbol{\eta}_t.$$

As above, such non-linear state space model can further be stated as

$$\mathbf{y}_t \approx \mathbf{a}_t + \mathbf{B}_{t|t-1}\boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t, \tag{3.9}$$
$$\boldsymbol{\xi}_t \approx \mathbf{c}_t + \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t, \tag{3.10}$$

with

$$\mathbf{B}_{t|t-1} = \left.\frac{\partial \mathbf{g}_t\big(\boldsymbol{\xi}_t\big)}{\partial \boldsymbol{\xi}_t^\top}\right|_{\boldsymbol{\xi}_t = \boldsymbol{\xi}_{t|t-1}},$$

$$\boldsymbol{\Phi}_{t|t-1} = \left.\frac{\partial \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}\big)}{\partial \boldsymbol{\xi}_{t-1}^\top}\right|_{\boldsymbol{\xi}_{t-1} = \boldsymbol{\xi}_{t-1|t-1}}.$$

The terms $\mathbf{a}_t$ and $\mathbf{c}_t$ in (3.9) and (3.10) are defined by

$$\begin{aligned} \mathbf{a}_t &= \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}\big) - \mathbf{B}_{t|t-1}\boldsymbol{\xi}_{t|t-1}, \text{ and} && (3.11)\\ \mathbf{c}_t &= \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}\big) - \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\xi}_{t-1|t-1}, && (3.12) \end{aligned}$$

respectively.

The quantities $\boldsymbol{\xi}_{t|t}$ and $\boldsymbol{\xi}_{t|t-1}$ are calculated by applying the **usual Kalman filter** to (3.9)–(3.12) with the modification so that the state *prediction equations* (1.6) and (1.7) become the *extended prediction equations*

$$\begin{aligned} \boldsymbol{\xi}_{t|t-1} &= \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}\big), && (3.13)\\ \boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_{t|t-1}^{\top} + \mathbf{Q}_t, && (3.14) \end{aligned}$$

where $\boldsymbol{\Sigma}_{t|t-1}$ is the *a priori estimate error covariance matrix*.

By analogy, the state *updating equations* (1.11) and (1.12) become the *extended updating equations*

$$\begin{aligned} \boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, && (3.15)\\ \boldsymbol{\Sigma}_{t|t} &= \big(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_{t|t-1}\big)\boldsymbol{\Sigma}_{t|t-1}, && (3.16) \end{aligned}$$

with

$$\begin{aligned} \mathbf{I}_t &= \mathbf{y}_t - \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}\big), && (3.17)\\ \mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top}\mathbf{F}_{t|t-1}^{-1}, \text{ and} && (3.18)\\ \mathbf{F}_{t|t-1} &= \mathbf{B}_{t|t-1}\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top} + \mathbf{H}_t. && (3.19) \end{aligned}$$

Thus, we directly can write

$$\begin{aligned} \boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top}\mathbf{F}_{t|t-1}^{-1}\Big[\mathbf{y}_t - \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}\big)\Big], \text{ and}\\ \boldsymbol{\Sigma}_{t|t} &= \Big(\boldsymbol{\mathcal{I}} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top}\mathbf{F}_{t|t-1}^{-1}\mathbf{B}_{t|t-1}\Big)\,\boldsymbol{\Sigma}_{t|t-1}. \end{aligned}$$

These recursions in (3.13)–(3.19) are known as the *extended Kalman filter*. Please note that $\mathbf{a}_t$ in (3.7) and $\mathbf{c}_t$ in (3.8) are never actually computed.

Generally: for the approximated state space model in the general form (3.5)–(3.8) we are able to derive the following modification of the linear Kalman filter algorithm containing the modified version of the innovations $\mathbf{I}_t$ and the Kalman gain $\mathbf{K}_t$. Therefore, the complete set of extended Kalman filter equations can be written as follows:

$$\begin{aligned}
\boldsymbol{\xi}_{t|t-1} &= \mathbf{h}_t\big(\boldsymbol{\xi}_{t-1|t-1}, \mathbf{0}, \boldsymbol{\psi}\big), && (3.20)\\
\boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Phi}_{t|t-1}\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{\Phi}_{t|t-1}^{\top} + \mathbf{S}_{t|t-1}\mathbf{Q}_t\mathbf{S}_{t|t-1}^{\top}, && (3.21)
\end{aligned}$$

$$\begin{aligned}
\mathbf{I}_t &= \mathbf{y}_t - \mathbf{g}_t\big(\boldsymbol{\xi}_{t|t-1}, \mathbf{0}, \boldsymbol{\psi}\big), && (3.22)\\
\mathbf{F}_{t|t-1} &= \mathbf{B}_{t|t-1}\boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top} + \mathbf{R}_{t|t-1}\mathbf{H}_t\mathbf{R}_{t|t-1}^{\top},
\end{aligned}$$

$$\begin{aligned}
\mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{B}_{t|t-1}^{\top}\mathbf{F}_{t|t-1}^{-1}, && (3.23)\\
\boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{K}_t\mathbf{I}_t, && (3.24)\\
\boldsymbol{\Sigma}_{t|t} &= \big(\boldsymbol{\mathcal{I}} - \mathbf{K}_t\mathbf{B}_{t|t-1}\big)\boldsymbol{\Sigma}_{t|t-1}. && (3.25)
\end{aligned}$$

These recursions make up the complete *extended Kalman filter*.

- The first two relationships in (3.20) and (3.21) are known as *extended Kalman filter time update equations*. As with the basic discrete Kalman filter, these time update equations **project** the state and covariance estimates from the previous time step $t-1$ to the current time step t.

  Again $\mathbf{h}_t$ in (3.20) comes from (3.2), the matrices $\boldsymbol{\Phi}_{t|t-1}$ and $\mathbf{S}_{t|t-1}$ are the transition Jacobians at time step t, and $\mathbf{Q}_t$ is the transition noise covariance $\big(\text{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t\big)$ at the current time step t.

- The last three relationships in (3.23)–(3.25) are known as *extended Kalman filter measurement update equations*. As with the basic discrete Kalman filter, these measurement update equations **correct** the state and covariance estimates with the measurement $\mathbf{y}_t$.

  Again $\mathbf{g}_t$ in (3.22) comes from (3.1), the matrices $\mathbf{B}_{t|t-1}$ and $\mathbf{R}_{t|t-1}$ are the measurement Jacobians at time step t, and $\mathbf{H}_t$ is the measurement noise covariance $\big(\text{Var}[\boldsymbol{\epsilon}_t] = \mathbf{H}_t\big)$ at the current time step t.

This system of equations can recursively be implemented in analogy to the algorithm in the linear case presented in the previous chapters.

The basic operation of the extended Kalman filter is the same as the linear discrete Kalman filter described in the previous chapters, i.e. as shown in Figure 1.2. For the ongoing extended Kalman filter cycle holds: The *time update* projects the current state estimate ahead in time. The *measurement update* adjusts the projected estimate by an actual measurement at that time.

Figure 3.1 below offers a complete picture of the operation of the extended Kalman filter, combining the high-level diagram of Figure 1.2 with the extended time update and measurement update equations from (3.20)–(3.21) and (3.23)–(3.25).

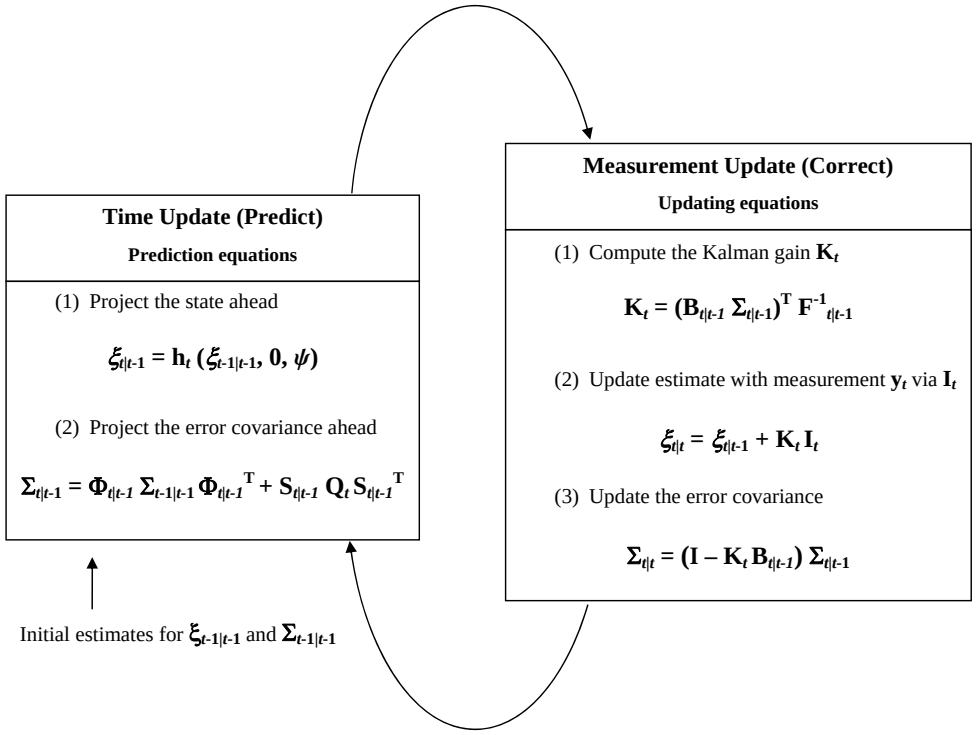


Figure 3.1: A complete picture of the operation of the extended Kalman filter.

We complete this chapter by two practical examples.

Example 3.3.1. *Consider the following non-linear model:*

$$\begin{aligned} \mathbf{y}_t &= \log \boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t, \\ \boldsymbol{\xi}_t &= \boldsymbol{\xi}_{t-1}^2 + \boldsymbol{\eta}_t. \end{aligned}$$

The linear approximation to this non-linear model (in which $\boldsymbol{\xi}_t$ is a scalar) is

$$\begin{aligned} \mathbf{y}_t &\approx \boldsymbol{\xi}_{t|t-1}^{-1}\boldsymbol{\xi}_t + \left[\log \boldsymbol{\xi}_{t|t-1} - 1\right] + \boldsymbol{\epsilon}_t, \\ \boldsymbol{\xi}_t &\approx 2\boldsymbol{\xi}_{t-1|t-1}\boldsymbol{\xi}_{t-1} + \left[\boldsymbol{\xi}_{t-1|t-1}^2 - 2\boldsymbol{\xi}_{t-1|t-1}\right] + \boldsymbol{\eta}_t. \end{aligned}$$

In the following Example 3.3.2 we show how some constraints (mentioned in Section 2.4 in Chapter 2) on the state vector $\boldsymbol{\beta}_t$ (or $\boldsymbol{\xi}_t$ generally) – which will be further needed in our option pricing analysis – can easily be implemented.

Example 3.3.2. *Suppose that the (one-dimensional) level component in a model is know to lie between zero and one. One way of incorporating a constraint of this kind into a model is by means of a logistic transformation. Thus*

$$\begin{aligned} y_t &= \frac{1}{1+\exp(-\beta_t)} + \epsilon_t, \\ \beta_t &= \beta_{t-1} + \eta_t. \end{aligned}$$

The range $[-\infty, +\infty]$ *for* β_t *translates into a range of* $[0,1]$ *for the component*

$$g_t(\beta_t) = \frac{1}{1+\exp(-\beta_t)}. \tag{3.26}$$

The extended Kalman filter for this problem takes the form of the usual Kalman filter with

$$\begin{aligned} B_{t|t-1} &= \left.\frac{\partial g_t(\beta_t)}{\partial \beta_t}\right|_{\beta_t=\beta_{t|t-1}} \\ &= \frac{\exp(-\beta_{t|t-1})}{\left[1+\exp(-\beta_{t|t-1})\right]^2}, \end{aligned}$$

with the innovations

$$I_t = y_t - \frac{1}{1+\exp(-\beta_{t|t-1})}.$$

The above methods can be extended to handle other kinds of non-linearity in the state space model. The quality of such approximations depends on the degree of non-linearity and the accuracy of the conditional means $\boldsymbol{\xi}_{t|t-1}$ and $\boldsymbol{\xi}_{t-1|t-1}$ as the optimal estimators of $\boldsymbol{\xi}_t$ and $\boldsymbol{\xi}_{t-1}$ respectively.

Further discussion on the extended Kalman filter and other non-linear filtering techniques can be found in Harvey [13].

Chapter 4

Extended Kalman filter and CALL options

In the following, please consider again the same case of a CALL option exactly as described in Chapter 2. We again omit the symbols $\boldsymbol{\psi}$ from the notation. Some constraints on the state vector $\boldsymbol{\beta}_t$ may be defined.

4.1 Implementing the constraints

According to Section 2.4 and in order to impose mentioned constraints on state vector $\boldsymbol{\beta}_t = (\beta_0, \ldots, \beta_{p-1})$ resulting from the constraints (1)–(4) in Subsection 2.3.2 $\big($requested was that $\beta_j > 0$ for $j = 0, \ldots, p-1$ and that $\sum_{j=2}^{p-1} \beta_j \leq 1\big)$, we propose the following reparametrization – transformation via a smooth function $\mathbf{g}_t(\cdot) = \big(g_0(\cdot), \ldots, g_{p-1}(\cdot)\big)$ – of the model in terms of parameters $\widetilde{\beta}_j$, $j = 0, \ldots, p-1$. These are related to the vector $\widetilde{\boldsymbol{\beta}}_t = (\widetilde{\beta}_0, \ldots, \widetilde{\beta}_{p-1})$, which is useful for calculating the estimates. We set:

$$
\begin{aligned}
\widetilde{\beta}_0 &= g_0(\boldsymbol{\beta}_t) \\
&= \exp(\beta_0), \\
\\
\widetilde{\beta}_1 &= g_1(\boldsymbol{\beta}_t) \\
&= \frac{\exp(\beta_1)}{\sum_{j=1}^{p-1} \exp(\beta_j)}, \\
&\vdots \\
\widetilde{\beta}_{p-1} &= g_{p-1}(\boldsymbol{\beta}_t) \\
&= \frac{\exp(\beta_{p-1})}{\sum_{j=1}^{p-1} \exp(\beta_j)}.
\end{aligned}
$$

Therefore, we directly have

$$\sum_{j=1}^{p-1} \widetilde{\beta}_j = 1.$$

Clearly, the introduced parameters $\widetilde{\beta}_j$, $j = 0, \ldots, p-1$ also satisfy the constraints

$$\widetilde{\beta}_j > 0, \quad j = 0, \ldots, p-1.$$

This means that the parameters $\widetilde{\beta}_1, \ldots, \widetilde{\beta}_{p-1}$ can be considered as points estimates of the state price density: these estimates are positive and integrate to one.

The following projection $\mathbf{g}_t(\cdot)$ has also been derived:

$$\underbrace{\big(g_0(\cdot), g_1(\cdot), \ldots, g_{p-1}(\cdot)\big)}_{\mathbf{g}_t(\cdot)} = \underbrace{\big(\widetilde{\beta}_0, \widetilde{\beta}_1, \ldots, \widetilde{\beta}_{p-1}\big)}_{\widetilde{\boldsymbol{\beta}}_t}, \tag{4.1}$$

which concretely is:

$$\mathbf{g}_t(\boldsymbol{\beta}_t) = \Big(\exp(\beta_0), \frac{\exp(\beta_1)}{\sum_{j=1}^{p-1} \exp(\beta_j)}, \ldots, \frac{\exp(\beta_{p-1})}{\sum_{j=1}^{p-1} \exp(\beta_j)} \Big). \tag{4.2}$$

The linear model for the option prices $\mathbf{C}$ derived in Section 2.4, i.e. the model

$$\mathbf{C}(\boldsymbol{\mathcal{K}}) = \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{4.3}$$

where $\boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}$ is the corresponding design matrix, must be rewritten in terms of parameters $\widetilde{\beta}_j$, $j = 0, \ldots, p-1$. This logically leads to a non-linear state space model which will be estimated using extended Kalman filter approaches.

4.2 Measurement and transition equations

According to Subsections 2.5.1 and 2.5.2, the extended measurement and transition equations must be modified. Because we are looking for the SPD, the range $[-\infty, +\infty]$ for $\boldsymbol{\beta}_t$ in general case will be translated into a range of $[0, 1]$ for the component $\mathbf{g}_t(\boldsymbol{\beta}_t)$.

As the simpler case of the non-linear state space model from Example 3.1.2, the model for the CALL option prices $\mathbf{C}$ has the following, corresponding form

$$\begin{aligned} \mathbf{C}_t(\boldsymbol{\mathcal{K}}) &= \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}} \mathbf{g}_t(\boldsymbol{\beta}_t) + \boldsymbol{\epsilon}_t, && (4.4) \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, && (4.5) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\epsilon}_t &\sim \mathrm{N}\big[\mathbf{0}, \mathbf{H}_t(\boldsymbol{\mathcal{K}})\big], \\ \boldsymbol{\eta}_t &\sim \mathrm{N}\big[\mathbf{0}, \sigma^2 \boldsymbol{\mathcal{I}}\big], \end{aligned}$$

with $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ defined in (2.7). These two error terms stay uncorrelated with each other in all time periods.

We required[1] in (4.5) that $\mathbf{h}_t(\cdot) = id$. The function $\mathbf{g}_t(\cdot)$ in (4.4) is the projection derived in (4.1) and (4.2), i.e.

$$\begin{aligned} \mathbf{g}_t(\boldsymbol{\beta}_t) &= \widetilde{\boldsymbol{\beta}}_t \\ &= \left(\exp(\beta_0), \frac{\exp(\beta_1)}{\sum_{j=1}^{p-1} \exp(\beta_j)}, \ldots, \frac{\exp(\beta_{p-1})}{\sum_{j=1}^{p-1} \exp(\beta_j)} \right) \\ &= \frac{1}{S} \big(S \exp(\beta_0), \exp(\beta_1), \ldots, \exp(\beta_{p-1}) \big), \end{aligned}$$

with

$$S = \sum_{j=1}^{p-1} \exp(\beta_j).$$

Now, it is directly satisfied in this designed model that all of the SPD components $\widetilde{\beta}_j$ lie between zero and one and integrate to one.

[1] The model stated in (4.4) and (4.5) corresponds to the general model for the option prices mentioned in Section 4.3.

4.3 Extended Kalman filter recursions

The extended Kalman filter for the above problem takes the form of the usual Kalman filter with the corresponding Jacobian matrix

$$\begin{aligned}
\mathbf{B}_{t|t-1} &= \left.\frac{\partial \mathbf{g}_t(\boldsymbol{\beta}_t)}{\partial \boldsymbol{\beta}_t^\top}\right|_{\boldsymbol{\beta}_t=\boldsymbol{\beta}_{t|t-1}} \\
&= \left.\frac{\partial\big(g_0(\boldsymbol{\beta}_t), g_1(\boldsymbol{\beta}_t), \ldots, g_{p-1}(\boldsymbol{\beta}_t)\big)}{\partial\big(\beta_0, \ldots, \beta_{p-1}\big)}\right|_{\boldsymbol{\beta_t}=\boldsymbol{\beta_{t|t-1}}} \\
&= \begin{pmatrix}
\frac{\partial g_0(\boldsymbol{\beta}_t)}{\partial \beta_0} & \frac{\partial g_0(\boldsymbol{\beta}_t)}{\partial \beta_1} & \frac{\partial g_0(\boldsymbol{\beta}_t)}{\partial \beta_2} & \cdots & \frac{\partial g_0(\boldsymbol{\beta}_t)}{\partial \beta_{p-1}} \\
\frac{\partial g_1(\boldsymbol{\beta}_t)}{\partial \beta_0} & \frac{\partial g_1(\boldsymbol{\beta}_t)}{\partial \beta_1} & \frac{\partial g_1(\boldsymbol{\beta}_t)}{\partial \beta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\beta}_t)}{\partial \beta_{p-1}} \\
\frac{\partial g_2(\boldsymbol{\beta}_t)}{\partial \beta_0} & \frac{\partial g_2(\boldsymbol{\beta}_t)}{\partial \beta_1} & \frac{\partial g_2(\boldsymbol{\beta}_t)}{\partial \beta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\beta}_t)}{\partial \beta_{p-1}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial g_{p-1}(\boldsymbol{\beta}_t)}{\partial \beta_0} & \frac{\partial g_{p-1}(\boldsymbol{\beta}_t)}{\partial \beta_1} & \frac{\partial g_{p-1}(\boldsymbol{\beta}_t)}{\partial \beta_2} & \cdots & \frac{\partial g_{p-1}(\boldsymbol{\beta}_t)}{\partial \beta_{p-1}}
\end{pmatrix}. \qquad (4.6)
\end{aligned}$$

The matrix (4.6) is always computed in the point $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t|t-1}$, i.e. by computing this matrix we use the previous values of the state vector $\boldsymbol{\beta}_t$ computed at the previous time step $t-1$. From (4.2) we concretely know the structure of the function $\mathbf{g}_t(\boldsymbol{\beta}_t)$. Therefore, it can immediately be computed:

$$\mathbf{B}_{t|t-1} = \frac{1}{S^2}\begin{pmatrix}
S^2 e^{\beta_0} & 0 & 0 & \cdots & 0 \\
0 & e^{\beta_1}\big(S - e^{\beta_1}\big) & -e^{\beta_1+\beta_2} & \cdots & -e^{\beta_1+\beta_{p-1}} \\
0 & -e^{\beta_2+\beta_1} & e^{\beta_2}\big(S - e^{\beta_2}\big) & \cdots & -e^{\beta_2+\beta_{p-1}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & -e^{\beta_{p-1}+\beta_1} & -e^{\beta_{p-1}+\beta_2} & \cdots & e^{\beta_{p-1}}\big(S - e^{\beta_{p-1}}\big)
\end{pmatrix},$$

where

$$S = \sum_{j=1}^{p-1} \exp(\beta_j).$$

Now, the Kalman filter algorithm for this model can be used, and equations (3.13)–(3.19) for the case of a CALL option become:

$$\begin{aligned} \boldsymbol{\beta}_{t|t-1} &= \boldsymbol{\beta}_{t-1|t-1}, && (4.7)\\ \boldsymbol{\Sigma}_{t|t-1} &= \boldsymbol{\Sigma}_{t-1|t-1} + \sigma^2 \boldsymbol{\mathcal{I}}, && (4.8) \end{aligned}$$

for the *extended prediction equations*, and

$$\begin{aligned} \boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \mathbf{K}_t \mathbf{I}_t, && (4.9)\\ \boldsymbol{\Sigma}_{t|t} &= \big(\boldsymbol{\mathcal{I}} - \mathbf{K}_t \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}} \mathbf{B}_{t|t-1}\big)\boldsymbol{\Sigma}_{t|t-1}, && (4.10) \end{aligned}$$

with

$$\begin{aligned} \mathbf{I}_t &= \mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}_{\Delta} \mathbf{g}_t(\boldsymbol{\beta}_{t|t-1}), && (4.11)\\ \mathbf{F}_{t|t-1} &= \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}} \mathbf{B}_{t|t-1} \boldsymbol{\Sigma}_{t|t-1} \mathbf{B}_{t|t-1}^{\top} \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}^{\top} + \mathbf{H}_t(\boldsymbol{\mathcal{K}}), \text{ and} && (4.12)\\ \mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1} \mathbf{B}_{t|t-1}^{\top} \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}^{\top} \mathbf{F}_{t|t-1}^{-1}. && (4.13) \end{aligned}$$

for the *extended updating equations*.

Recursive equations in (4.7)–(4.13) are the wanted *extended Kalman filter recursions for a CALL option* resulting to the state vector $\mathbf{g}_t(\boldsymbol{\beta}_t) = \widetilde{\boldsymbol{\beta}}_t$, representing the points estimates of the risk neutral state price density. These recursions will be implemented in R in the next chapter.

Chapter 5

Data analysis

Having determined both versions of the Kalman filter for a CALL option, we are now able to use these methods to estimate a model based on real data: we use the discussed approaches to estimate the risk-neutral price density of a European CALL option. We estimate SPD using EUREX option data on the DAX index via the Kalman Filter estimator of the second derivative of the European CALL price function. This estimator will be constrained so as to satisfy no-arbitrage constraints. The whole problem should briefly be mentioned: we developed a simple estimation technique in order to construct the SPD (see Subsection 2.2) estimates. The aim of the following data analysis is to construct an estimate of the SPD based on the observed CALL option prices satisfying all constraints on a general density function (the implied SPD is non-negative and integrates to one).

In the form of measurement and transition equations from (1.1) and (1.2), the following simple linear model for CALL option prices has been constructed:

$$\begin{aligned} \mathbf{C}_t(\boldsymbol{\mathcal{K}}) &= \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, & (5.1)\\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, \ldots, T, & (5.2) \end{aligned}$$

where $\boldsymbol{\mathcal{X}}$ is the matrix from (2.6), $\boldsymbol{\epsilon}_t \sim \mathrm{N}\big[\mathbf{0}, \mathbf{H}_t(\boldsymbol{\mathcal{K}})\big]$, with $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ defined in (2.7), and $\boldsymbol{\eta}_t \sim \mathrm{N}\big[\mathbf{0}, \sigma^2\boldsymbol{\mathcal{I}}\big]$. These two error terms are uncorrelated with each other in all time periods.

In the previous chapters, we derived the usual and the extended Kalman filter for this model. In the following, we try to estimate the SPD based on these recursions in a case with real data.

The model (5.1) and (5.2) was constructed in a way that the second $(p-1)$ components of the estimated state vector $\boldsymbol{\beta}_t$ can be – as estimated second derivatives of the European CALL price function – considered as points estimates of the SPD. We repeat that p is number of the various strike prices k_j.

Our data set (one set corresponds to one day) contains the observed CALL and PUT option prices for various strike prices $\mathbf{K}$ and maturities T. We have 39 data sets corresponding to 39 days (from JAN-02-2003 till FEB-25-2003, abbreviated from `EUREX1.dat` to `EUREX39.dat`). Therefore, every day such a point cloud with about 2500 data points has been observed.

We will be working just with CALL option prices $\mathbf{C}_t(\mathcal{K}, T)$, $t = 1, \ldots, T$ with the shortest time to expiry τ. Figure 5.1 below displays the observed prices of European CALL options with the shortest maturity written on the DAX for the first observed day (JAN-02-2003, abbreviated as `EUREX1.dat`). In our empirical study we will be considering this data set containing $T = 950$ data points.

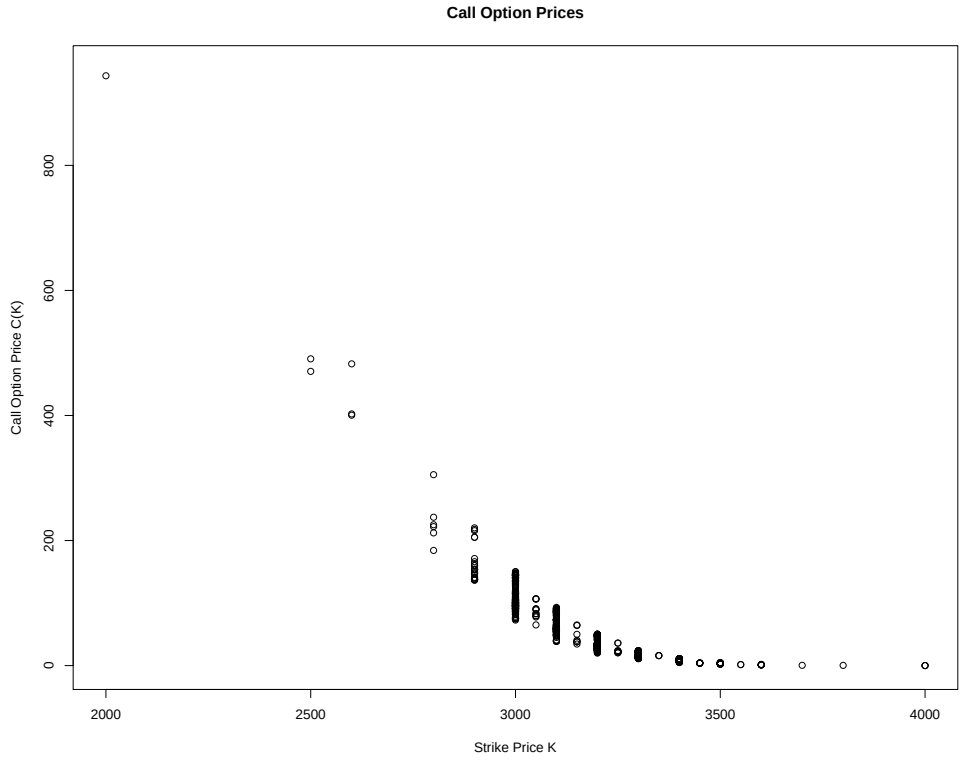


Figure 5.1: European CALL option prices with shortest time to expiry plotted against strike price K on JAN-02-2003, $T = 950$ observed prices.

Other variables are contained: the interest rate r and time of trade t (in seconds from midnight on the current day). We will analyze the option prices $\mathbf{C}_t(\mathcal{K}, T)$ as a function of the strike price K for a fixed date and a fixed time to expiry τ. One of the first tasks of our data analysis is to standardize the option prices $\mathbf{C}_t(\mathcal{K}, T)$: because we did not consider any discount factor $e^{-r(T-t)}$ above (it was assumed to be known, please see in Section 2.2), each observed option price $\mathbf{C}_t(\mathcal{K}, T)$ is divided by this discount factor $e^{-r(T-t)}$.

In the following, both Kalman filters (the usual and the extended version) are distinguished.

5.1 Kalman filter algorithm

5.1.1 Initialization steps

Model error terms

To be able to use both versions of the Kalman filter, an initialization step must be proceeded. We have to estimate variances of both (measurement and process) error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ from (5.1) and (5.2). It is clear that the key factor which determines the quality of (all) estimations in our considered model is the difference between the real and the used variance of the random error term $\boldsymbol{\epsilon}_t$. Therefore, it is in our interest to find a confident adaptive (and consistent) estimation of that variance $\sigma^2_{\boldsymbol{\epsilon}}$.

There are plenty of methods – mostly based on EM-algorithm (you can find more in Franěk [5]) – how to estimate unknown variances $\mathrm{Var}[\boldsymbol{\epsilon}_t]$ and $\mathrm{Var}[\boldsymbol{\eta}_t]$. Some of them can be found in Harvey [13] or in Kellerhals [15]. However, we choose as the simplest one just to take feasible values of $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ and $\sigma^2\boldsymbol{\mathcal{I}}$, to use these as initialization values to be able to start the Kalman filter for the first time, to compute more accurately values based on likelihood estimation within the main recursive procedure, and then to use these new values for the second run of the Kalman filter.

We discussed in Subsection 2.3.2 that the variance-covariance matrix $\mathbf{H}_t(\boldsymbol{\mathcal{K}})$ will be assumed to be homoscedastic.

Thus, for both variances we set

$$\begin{aligned} \mathrm{Var}[\boldsymbol{\epsilon}_t] &= \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}, \\ \mathrm{Var}[\boldsymbol{\eta}_t] &= \sigma^2_{\boldsymbol{\eta}}\boldsymbol{\mathcal{I}}, \end{aligned}$$

where $\sigma^2_{\boldsymbol{\epsilon}} = \mathrm{Var}[\mathbf{C}_t(\mathcal{K}, T)]$ as the simplest initialization, and on top of that we set $\sigma^2_{\boldsymbol{\eta}} = 1$ so that $\mathrm{Var}[\boldsymbol{\eta}_t] = \boldsymbol{\mathcal{I}}$.

Kalman filter initialization values

To get the Kalman filter started, initialization values $\boldsymbol{\beta}_{0|0}$ and $\boldsymbol{\Sigma}_{0|0}$ are set as follows:

$$\begin{aligned} \boldsymbol{\beta}_{0|0} &= \left(\underbrace{\mathrm{E}\left[C(k_p)\right]}_{\beta_0}, \underbrace{\frac{1}{p-1}, \ldots, \frac{1}{p-1}}_{p-1} \right), \\ \boldsymbol{\Sigma}_{0|0} &= \boldsymbol{\mathcal{I}}. \end{aligned}$$

The first component β_0 is set in the sense of the construction of the linear model in Section 2.4, where $k_1 < k_2 < \cdots < k_p$, and the second $(p-1)$ values $\frac{1}{p-1}$ define the initial distribution of the SPD – the uniform distribution in this case. This approach can be vindicated thanks to the fact, that the influences of the initial distribution $\boldsymbol{\beta}_{0|0}$ disappear in the time invariant filter exponentially fast. The error caused by the substituting of the real distribution by the uniform is absorbed by the system at an exponential speed (you can find more for example in Franěk [5] or in Künsch [16]). Therefore, it can even be set: $\boldsymbol{\beta}_{0|0} = \mathbf{0}^\top$, and $\boldsymbol{\Sigma}_{0|0} = \mathbf{0}$.

The following values are computed from the data of the first day:

$$\begin{aligned} T &= 950, \\ p &= 21, \\ \\ \sigma^2_{\boldsymbol{\epsilon}} &= \mathrm{Var}[\mathbf{C}_t(\mathcal{K}, T)] \\ &= 3558.05, \text{ and} \\ \\ \beta_0 &= \mathrm{E}\left[C(k_{21} = 4000)\right] = 0.10012. \end{aligned}$$

Now, the Kalman filter is ready to be started with the following initial values:

$$\begin{aligned} \mathrm{Var}[\boldsymbol{\epsilon}_t] &= 3558.05\, \boldsymbol{\mathcal{I}}_{T\times T}, \\ \mathrm{Var}[\boldsymbol{\eta}_t] &= \boldsymbol{\mathcal{I}}_{p\times p}, \\ \boldsymbol{\beta}_{0|0} &= \left(0.10012, \frac{1}{20}, \ldots, \frac{1}{20}\right), \text{ and} \\ \boldsymbol{\Sigma}_{0|0} &= \boldsymbol{\mathcal{I}}_{p\times p}. \end{aligned}$$

5.1.2 Kalman filter recursive procedure

Model error terms

Within the procedure, the variances of model error terms $\mathrm{Var}[\boldsymbol{\epsilon}_t]$ and $\mathrm{Var}[\boldsymbol{\eta}_t]$ are being estimated. As computing tool, we choose the estimators based on the EM-algorithm which are well described in Härdle et al. [8]. The algorithm is derived under the assumption of Gaussian error terms. Then the logarithm of the likelihood function for our model is expressed as

$$\begin{aligned} \log L &= -\frac{1}{2}\log\left|\boldsymbol{\Sigma}_{0|0}\right| - \frac{1}{2}\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_{0|0}\right)^\top \left(\boldsymbol{\Sigma}^{-1}_{0|0}\right)\left(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_{0|0}\right) \\ &\quad -\frac{T}{2}\log\left|\sigma^2_{\boldsymbol{\eta}}\boldsymbol{\mathcal{I}}\right| - \frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t-1}\right)^\top \left(\sigma^2_{\boldsymbol{\eta}}\boldsymbol{\mathcal{I}}\right)^{-1}\left(\boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t-1}\right) \\ &\quad -\frac{T}{2}\log\left|\sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}\right| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}\right)^\top \left(\sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}\right)^{-1}\left(\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}\right). \end{aligned}$$

This directly leads to the following estimators:

$$\begin{aligned}
\widehat{\sigma}^2_{\boldsymbol{\epsilon}} &= \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbf{C}_t(\boldsymbol{\mathcal{K}}) - \boldsymbol{\mathcal{X}}\boldsymbol{\beta}_{t|t-1}\Big)^2, \text{ and} \\
\widehat{\sigma}^2_{\boldsymbol{\eta}} &= \frac{1}{T}\sum_{t=1}^{T}\Big(\boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t-1}\Big)^\top\Big(\boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t-1}\Big) \\
&= \frac{1}{Tp}\sum_{t=1}^{T}\sum_{j=1}^{p}\Big(\beta^{(j)}_{t|t} - \beta^{(j)}_{t|t-1}\Big)^2,
\end{aligned}$$

where $\beta^{(j)}_{\cdot|\cdot}$ denotes the j-th component of the corresponding vector $\boldsymbol{\beta}_{\cdot|\cdot}$. From the data of the first day, the following variances can be estimated:

$$\begin{aligned}
\widehat{\sigma}^2_{\boldsymbol{\epsilon}} &= 59.22, \\
\widehat{\sigma}^2_{\boldsymbol{\eta}} &= 0.00001018.
\end{aligned}$$

These values will be used for the second run of the filter as the new initial values.

Remark 5.1.1 (Another estimators)**.** *In Franěk [5], another interesting solution of the estimation problem can be found: estimation of the innovations* $\mathbf{I}_t$ *based on estimated* $\sigma^2_{\boldsymbol{\epsilon}}$.

State price density

Above all – in every time step $t = 1, \ldots, T$ – within the recursive algorithm the state vector $\boldsymbol{\beta}_{t|t}$ is being estimated, whose second $p - 1 = 20$ components β_j represent the points estimates of the risk neutral SPD, and the matrix $\boldsymbol{\Sigma}_{t|t}$ as the variance-covariance matrix of that estimator. Estimations of $\boldsymbol{\beta}_t$ are graphically displayed in Figures 5.2–5.4. Because we have $T = 950$ observed CALL option prices with the shortest time to expiry, i.e. 950 estimations of $\boldsymbol{\beta}_t$, and 950 estimations of (20×20) matrix $\boldsymbol{\Sigma}_t$ on JAN-02-2003, just nine values of $\widehat{\boldsymbol{\beta}}_t$ are displayed: estimations of the SPD based on $t = 1, 119, 238, 356, 475, 594, 712, 831, 950$ (distances are approximately $T/8$) observations.

At time $T = 950$ we get the last estimation of $\boldsymbol{\beta}_t = \boldsymbol{\beta}_T$, which can be considered as the real risk neutral SPD:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_T = \big(&0.10001, -0.00065, 0.00497, 0.00909, 0.01355, 0.01271, 0.01090, 0.02588, \\
&0.04076, 0.04380, 0.07811, 0.04811, 0.08558, 0.08363, 0.08190, 0.07843, \\
&0.08150, 0.14573, 0.09980, 0.08574, 0.06920\big).
\end{aligned}$$

Thus

$$\sum_{j=2}^{p}\widehat{\beta}_j = 1.099.$$

The estimated values $\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\Sigma}}_T$ obtained from data of the first day can simply be used as the initial values:

$$\begin{aligned} \boldsymbol{\beta}_{0|0}^{[2]} &= \widehat{\boldsymbol{\beta}}_T, \quad \text{and} \\ \boldsymbol{\Sigma}_{0|0}^{[2]} &= \widehat{\boldsymbol{\Sigma}}_T \end{aligned}$$

for the second run of the Kalman filter with the same observed data (to get more accurate estimations) or even for estimation SPD based on observed data of the second day. These could be used as initial values for the following days, etc. So we can write:

$$\begin{aligned} \boldsymbol{\beta}_{0|0}^{[i]} &= \widehat{\boldsymbol{\beta}}_T^{[i-1]}, \quad \text{and} \\ \boldsymbol{\Sigma}_{0|0}^{[i]} &= \widehat{\boldsymbol{\Sigma}}_T^{[i-1]}. \end{aligned}$$

Please note, that the dimensions of $\widehat{\boldsymbol{\beta}}_T^{[i-1]}$ as the estimation on $(i-1)$-st day and $\boldsymbol{\beta}_{0|0}^{[i]}$ as the initial value for the i-th day must be the same, i.e. data on the following day must contain the same various strike prices p as the day before.

The second run of the algorithm with new (more accurate) initial values $\boldsymbol{\beta}_{0|0}^{[2]}$, $\boldsymbol{\Sigma}_{0|0}^{[2]}$, $\widehat{\sigma}_{\boldsymbol{\epsilon}}^2$ and $\widehat{\sigma}_{\boldsymbol{\eta}}^2$ gives us:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_T^{[2]} = \big(&0.10012, 0.00085, 0.00116, 0.01436, 0.00293, 0.02382, 0.01125, 0.02275, \\ &0.03997, 0.06430, 0.05638, 0.04811, 0.08955, 0.07152, 0.08688, 0.06711, \\ &0.09409, 0.16017, 0.08293, 0.20604, -0.04175\big), \end{aligned}$$

with

$$\sum_{j=2}^{p} \widehat{\beta}_j^{[2]} = 1.103.$$

We also have

$$\begin{aligned} \widehat{\sigma}_{\boldsymbol{\epsilon}^{[2]}}^2 &= 36.32, \quad \text{and} \\ \widehat{\sigma}_{\boldsymbol{\eta}^{[2]}}^2 &= 0.00001027. \end{aligned}$$

As you can see, some components of $\boldsymbol{\beta}_T$ or $\widehat{\boldsymbol{\beta}}_T^{[2]}$ (and of course some of other estimations $\widehat{\boldsymbol{\beta}}_t^{[i]}$ obtained during the whole i-th day) can be negative and the sum of the second 20 components can be a little bit more than one. It is caused, understandably, by non-implementing the constraints on these parameters. Therefore, the extended Kalman filter shall be used.

Remark 5.1.2 (Kalman smoothing algorithm)**.** *There is another elegant approach how to compute the initial distribution of* $\boldsymbol{\beta}_t$*, i.e. values* $\boldsymbol{\beta}_{0|0}^{[2]}$ *and* $\boldsymbol{\Sigma}_{0|0}^{[2]}$*. It is the* ***Kalman smoothing algorithm*** *which estimates* $\boldsymbol{\beta}_t$ *and* $\boldsymbol{\Sigma}_t$ *in terms of all available observations* $(\mathbf{C}_1, \ldots, \mathbf{C}_T)$*. The smoothing recursion consists of the backward recursion that uses the filtered values of* $\boldsymbol{\beta}_t$ *and* $\boldsymbol{\Sigma}_t$ *obtained in the Kalman filter. The Kalman smoothing equations for our model are:*

$$\begin{aligned}
\boldsymbol{\beta}_{t|T} &= \boldsymbol{\beta}_{t|t} + \boldsymbol{\Sigma}_t^* \left(\boldsymbol{\beta}_{t+1|T} - \boldsymbol{\beta}_{t|t} \right), \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_{t|t} - \boldsymbol{\Sigma}_t^* \left(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{t+1|t} \right) \boldsymbol{\Sigma}_t^*, \\
\boldsymbol{\Sigma}_t^* &= \boldsymbol{\Sigma}_{t|t} \boldsymbol{\Sigma}_{t+1|t}^{\top}.
\end{aligned}$$

The smoothed values of $\boldsymbol{\beta}_t$ *and* $\boldsymbol{\Sigma}_t$ *can be again used as the initial values* $\boldsymbol{\beta}_{0|0}$ *and* $\boldsymbol{\Sigma}_{0|0}$ *for the same or the following day (again, if the corresponding dimensions match).*

There can be found plenty of graphical outputs from the statistical program `R` on the next pages.

- Estimates $\widehat{\boldsymbol{\beta}}_t$, based on $t = 1, 119, 238, 356, 475, 594, 712, 831, 950$ observations, are graphically displayed in Figures 5.2–5.4.

- The situation, in which the estimates $\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\beta}}_T^{[2]}$ differ, is displayed in Figure 5.5.

- Estimates $\widehat{\boldsymbol{\beta}}_t^{[2]}$, based on $t = 1, 119, 238, 356, 475, 594, 712, 831, 950$ observations, are graphically displayed in Figures 5.6–5.8.

In the following, the same steps will be done for the non-linear case of relation between state and observed variables. Hereafter, both algorithm (the usual Kalman filter and the extended Kalman filter) will be compared.

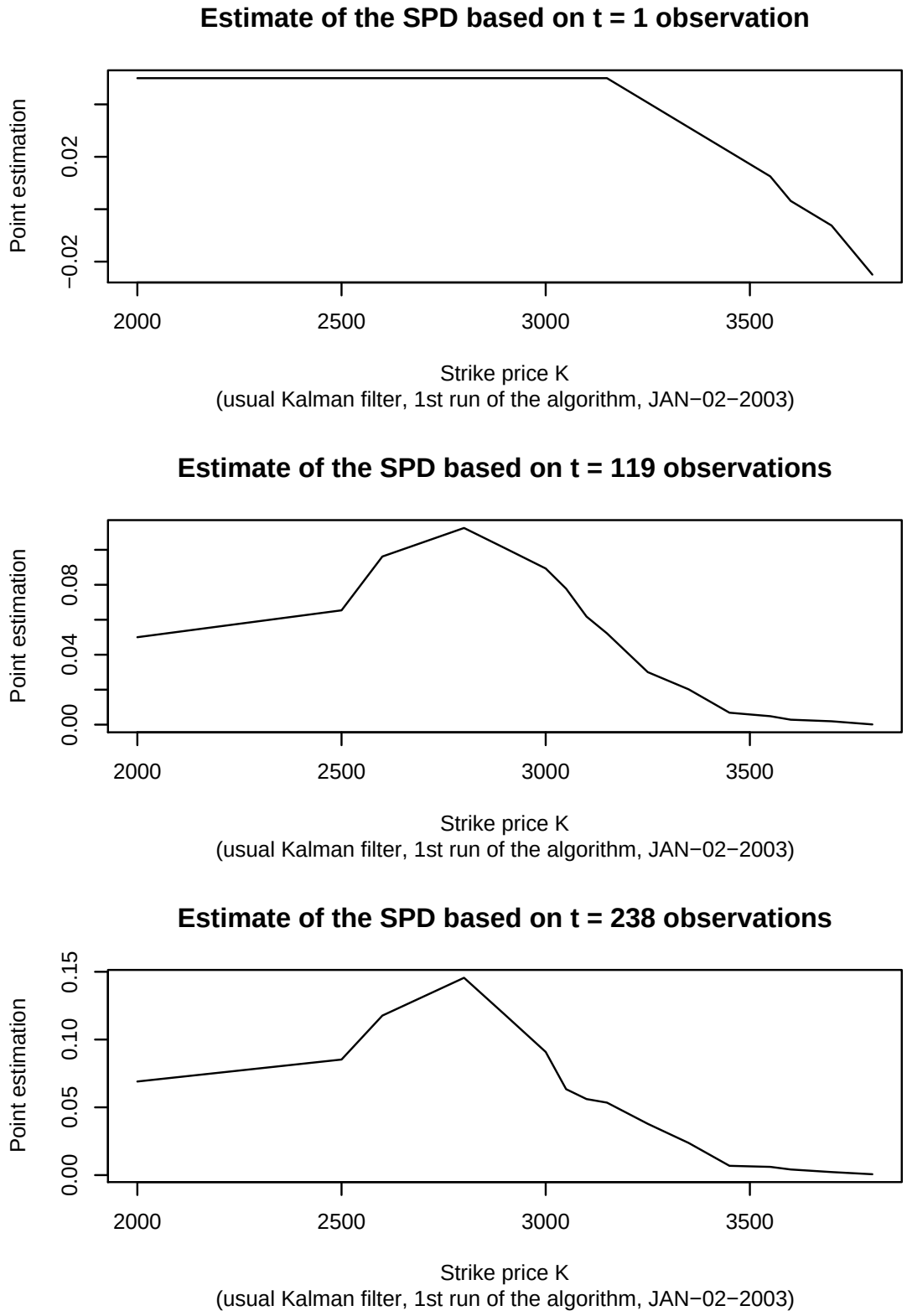


Figure 5.2: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t$, based on $t_1 = 1, t_2 = 119$ and $t_3 = 238$ observed prices.

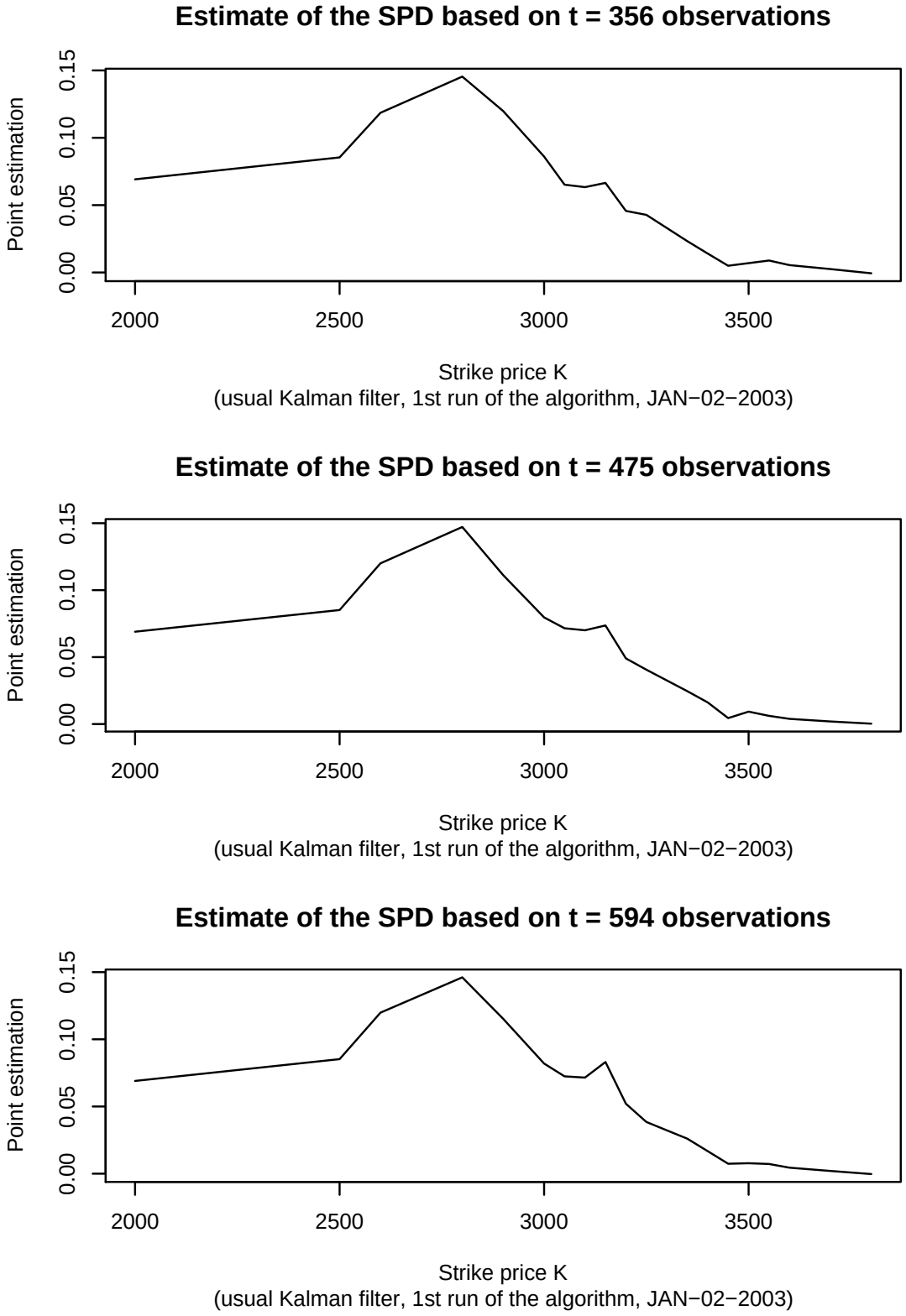


Figure 5.3: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t$, based on $t_4 = 356, t_5 = 475$ and $t_6 = 594$ observed prices.

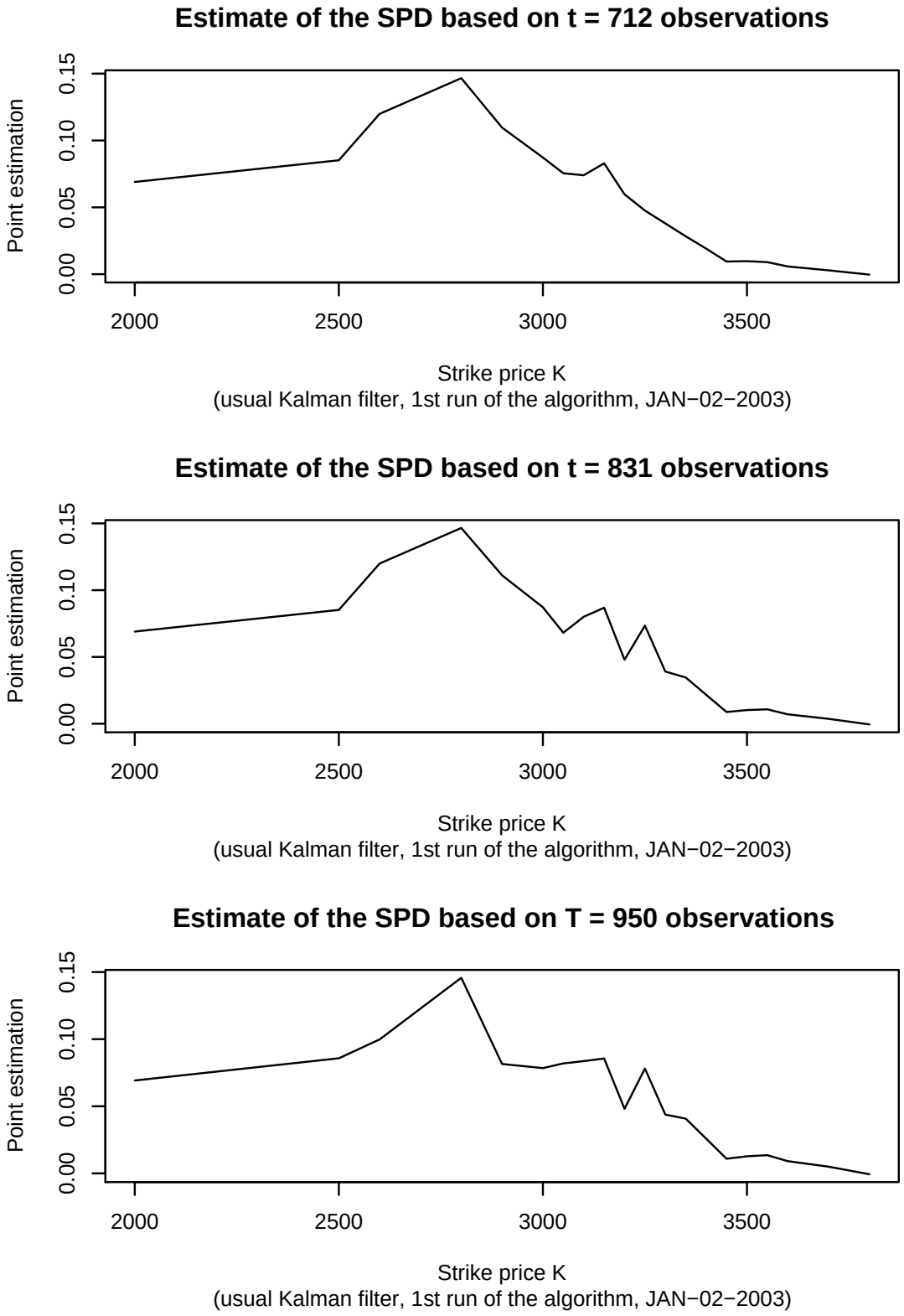


Figure 5.4: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t$, based on $t_7 = 712, t_8 = 831$ and $t_9 = T = 950$ observed prices.

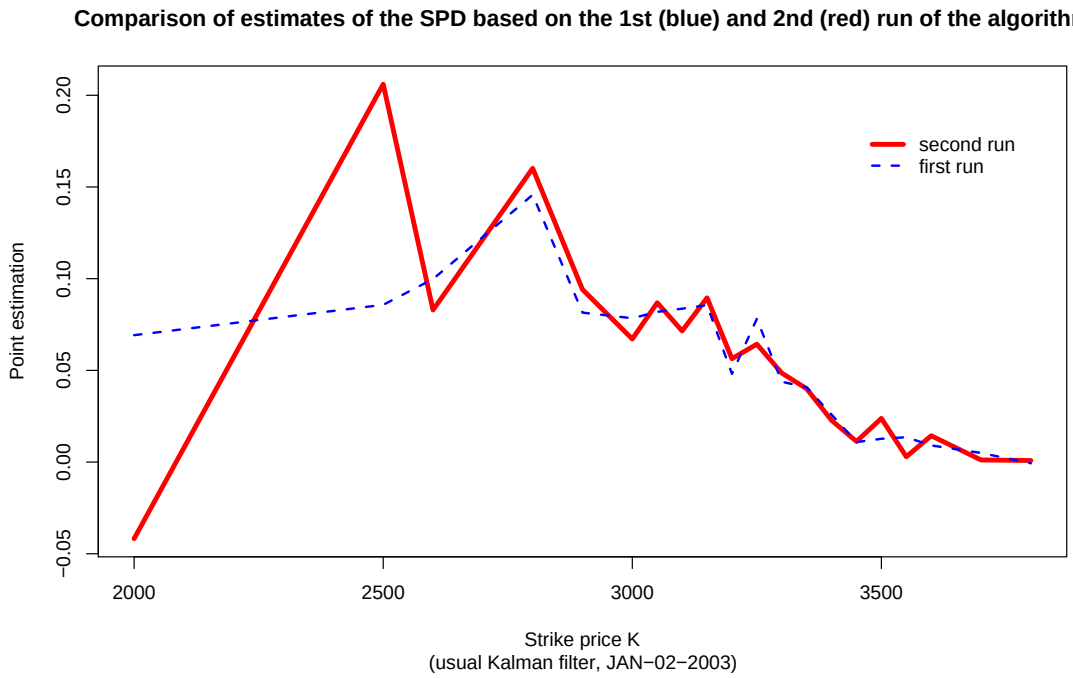


Figure 5.5: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\beta}}_T^{[2]}$, based on $T = 950$ observed prices.

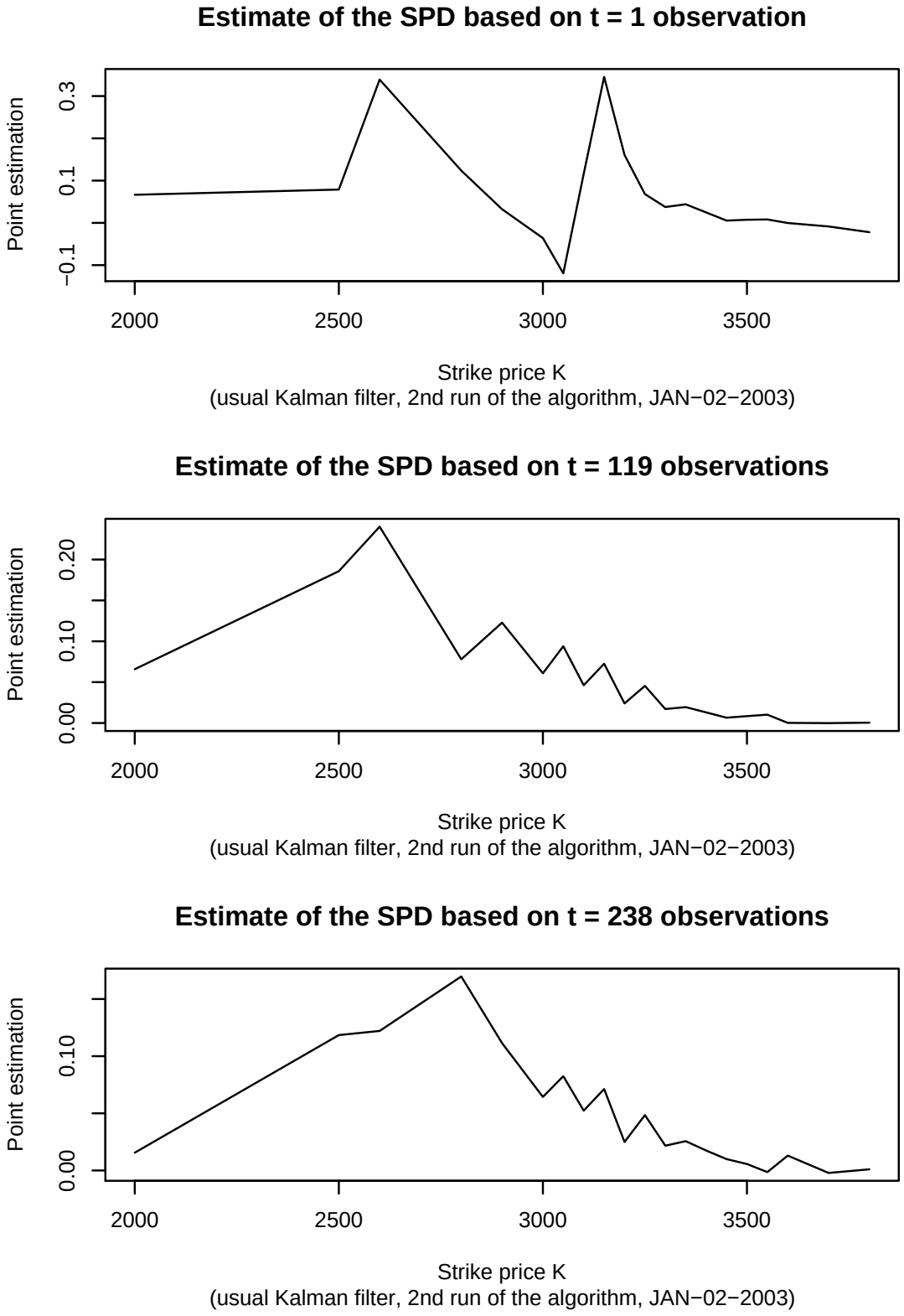


Figure 5.6: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t^{[2]}$, based on $t_1 = 1, t_2 = 119$ and $t_3 = 238$ observed prices.

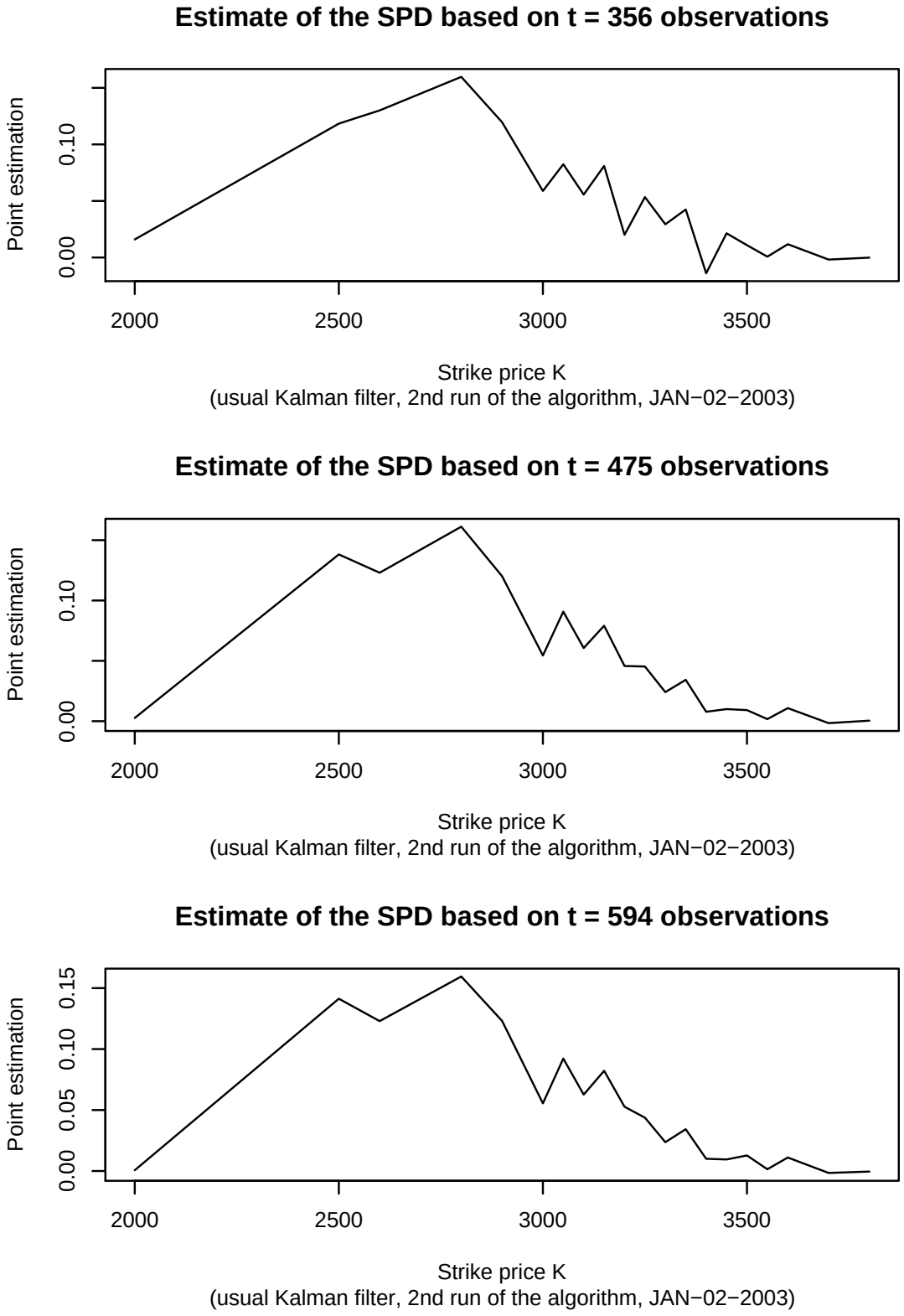


Figure 5.7: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t^{[2]}$, based on $t_4 = 356, t_5 = 475$ and $t_6 = 594$ observed prices.

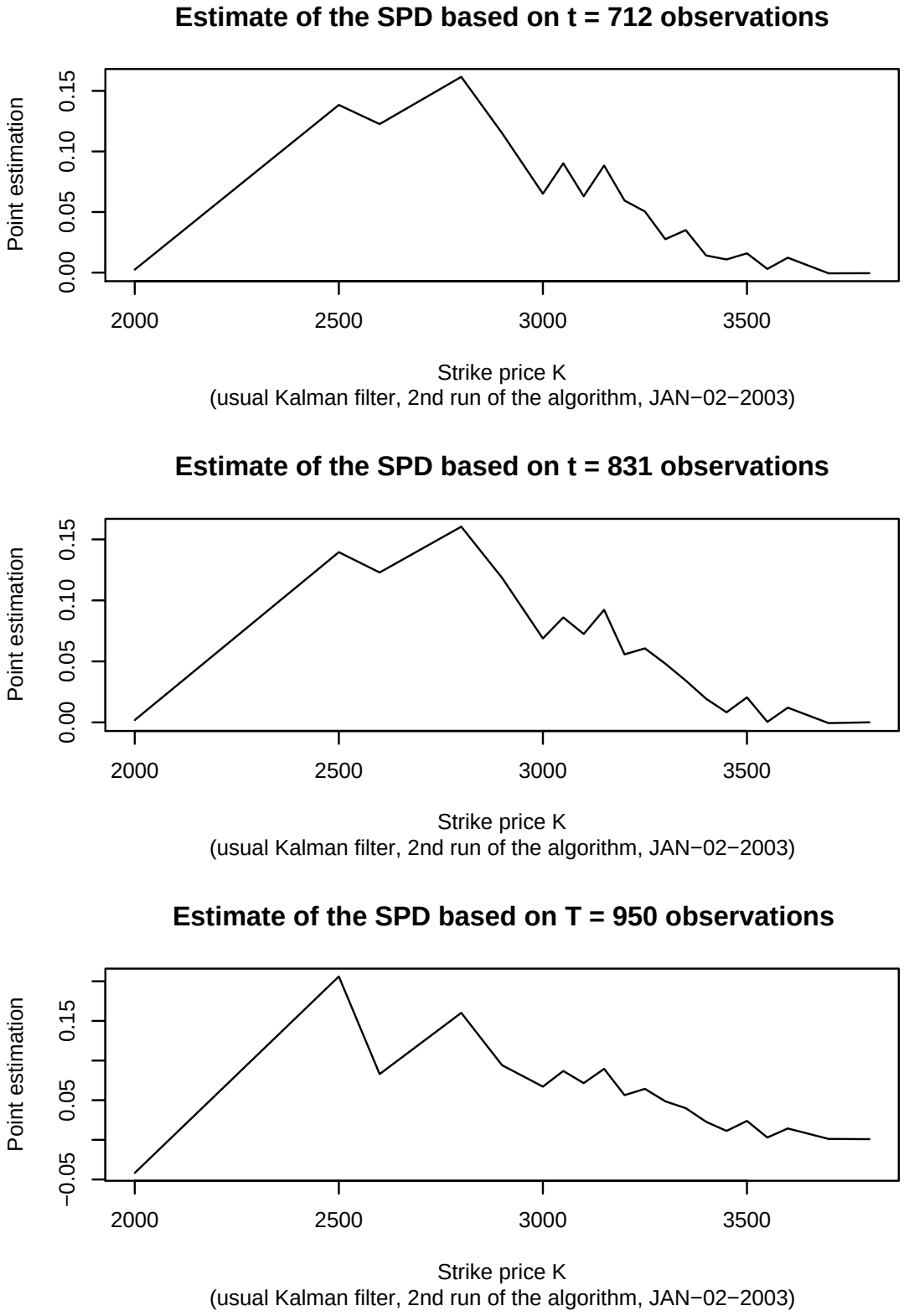


Figure 5.8: The unconstrained estimates of SPD against strike price K, $\widehat{\boldsymbol{\beta}}_t^{[2]}$, based on $t_7 = 712, t_8 = 831$ and $t_9 = T = 950$ observed prices.

5.2 Extended Kalman filter algorithm

The non-linear transformation of the state vector $\boldsymbol{\beta}_t$ is considered as derived in (4.1) and (4.2), i.e.

$$\begin{aligned}\mathbf{g}_t(\boldsymbol{\beta}_t) &= \widetilde{\boldsymbol{\beta}}_t \\ &= \left(\exp(\beta_0), \frac{\exp(\beta_1)}{\sum_{j=1}^{p-1}\exp(\beta_j)}, \ldots, \frac{\exp(\beta_{p-1})}{\sum_{j=1}^{p-1}\exp(\beta_j)}\right) \\ &= \frac{1}{S}\big(S\exp(\beta_0), \exp(\beta_1), \ldots, \exp(\beta_{p-1})\big),\end{aligned}$$

with

$$S = \sum_{j=1}^{p-1}\exp(\beta_j).$$

Therefore, the state vector $\widetilde{\boldsymbol{\beta}}_t$, whose second $p-1=20$ components $\widetilde{\beta}_j$, $j=2,\ldots,p-1$ directly satisfy the necessary constraints (they lie between zero and one and integrate to one), is to be estimated.

5.2.1 Initialization steps

Model error terms

The initialization step in case of the extended Kalman filter is the same as in Subsection 5.1.1, i.e. the measurement and process variances of error terms are assumed to be homoscedastic as follows:

$$\begin{aligned}\mathrm{Var}[\boldsymbol{\epsilon}_t] &= \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}, \\ \mathrm{Var}[\boldsymbol{\eta}_t] &= \boldsymbol{\mathcal{I}},\end{aligned}$$

where $\sigma^2_{\boldsymbol{\epsilon}} = \mathrm{Var}[\mathbf{C}_t(\mathcal{K}, T)]$.

Extended Kalman filter initialization values

The extended Kalman filter initializations values $\widetilde{\boldsymbol{\beta}}_{0|0}$ and $\boldsymbol{\Sigma}_{0|0}$ are set as follows:

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{0|0} &= \left(\beta_0, \frac{1}{p-1}, \ldots, \frac{1}{p-1}\right), \\ \boldsymbol{\Sigma}_{0|0} &= \boldsymbol{\mathcal{I}}.\end{aligned}$$

Therefore, it can be computed from the data of the first day:

$$\begin{aligned}\widetilde{\beta}_0 &= \mathrm{E}\left[C(k_{21}=4000)\right] \\ &= 0.10012, \quad \text{and}\end{aligned}$$

$$\begin{aligned}\sigma^2_{\boldsymbol{\epsilon}} &= \mathrm{Var}[\mathbf{C}_t(\mathcal{K},T)] \\ &= 3558.05.\end{aligned}$$

The extended Kalman filter starts with the following initial values:

$$\begin{aligned}\mathrm{Var}[\boldsymbol{\epsilon}_t] &= 3558.05\,\boldsymbol{\mathcal{I}}_{T\times T}, \\ \mathrm{Var}[\boldsymbol{\eta}_t] &= \boldsymbol{\mathcal{I}}_{p\times p}, \\ \widetilde{\boldsymbol{\beta}}_{0|0} &= \left(0.10012, \frac{1}{20}, \ldots, \frac{1}{20}\right), \text{ and} \\ \boldsymbol{\Sigma}_{0|0} &= \boldsymbol{\mathcal{I}}_{p\times p}.\end{aligned}$$

5.2.2 Extended Kalman filter recursive procedure

Model error terms

To estimate variances of model error terms $\mathrm{Var}[\boldsymbol{\epsilon}_t]$ and $\mathrm{Var}[\boldsymbol{\eta}_t]$ we again choose the estimators based on the EM-algorithm (under the assumption of Gaussian error terms). The logarithm of the likelihood function for this case is determined the same way as in Subsection 5.1.2, i.e. the estimations of the variances $\mathrm{Var}[\boldsymbol{\epsilon}_t]$ and $\mathrm{Var}[\boldsymbol{\eta}_t]$ are computed as follows:

$$\begin{aligned}\widehat{\sigma}^2_{\boldsymbol{\epsilon}} &= 80.56, \\ \widehat{\sigma}^2_{\boldsymbol{\eta}} &= 0.000000538.\end{aligned}$$

These values $\widehat{\sigma}^2_{\boldsymbol{\epsilon}}$ and $\widehat{\sigma}^2_{\boldsymbol{\eta}}$ will be used again for the second run of the extended filter as the new initial values.

State price density

The points estimates of the risk neutral SPD, i.e. estimations of the state vector $\widetilde{\boldsymbol{\beta}}_t$, are computed in every time step $t = 1, \ldots, T$. Estimates of the SPD, based on $t = 1, 119, 238, 356, 475, 594, 712, 831, 950$ observations, are graphically displayed in Figures 5.9–5.11.

At time $T = 950$ we get the last estimation of $\widetilde{\boldsymbol{\beta}}_t = \widetilde{\boldsymbol{\beta}}_T$, which can be considered as the real risk neutral SPD satisfying all constraints:

$$\begin{aligned}\widehat{\widetilde{\boldsymbol{\beta}}}_T &= \big(0.37171, 0.00143, 0.00217, 0.00296, 0.01025, 0.02362, 0.01399, 0.01260,\\ &\quad 0.05390, 0.01626, 0.124, 0.03027, 0.0863, 0.07146, 0.09751, 0.05914,\\ &\quad 0.10902, 0.1168, 0.0863, 0.05213, 0.02988\big).\end{aligned}$$

Thus

$$\sum_{j=2}^{p} \widehat{\widetilde{\beta}}_j = 1.00.$$

Similarly to the case of the usual Kalman filter, the estimated values $\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\Sigma}}_T$ can be used as the new initial values:

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{0|0}^{[2]} &= \widehat{\widetilde{\boldsymbol{\beta}}}_T, \quad \text{and}\\ \boldsymbol{\Sigma}_{0|0}^{[2]} &= \widehat{\boldsymbol{\Sigma}}_T.\end{aligned}$$

The second run of the extended algorithm with new (more accurate) initial values $\widetilde{\boldsymbol{\beta}}_{0|0}^{[2]}$, $\boldsymbol{\Sigma}_{0|0}^{[2]}$, $\widehat{\sigma}_{\boldsymbol{\epsilon}}^2$ and $\widehat{\sigma}_{\boldsymbol{\eta}}^2$ gives us:

$$\begin{aligned}\widehat{\widetilde{\boldsymbol{\beta}}}_T^{[2]} &= \big(0.000042, 0.0018, 0.0047, 0.02023, 0.0000012, 0.00281, 0.00062, 0.01016,\\ &\quad 0.09661, 0.03536, 0.01348, 0.08286, 0.01225, 0.1935, 0.06578, 0.05954,\\ &\quad 0.08123, 0.08308, 0.16846, 0.00000048, 0.06754\big),\end{aligned}$$

with

$$\sum_{j=2}^{p} \widehat{\widetilde{\beta}}_j^{[2]} = 1.00.$$

We also have

$$\begin{aligned}\widehat{\sigma^2}_{\boldsymbol{\epsilon}}^{[2]} &= 32.42, \quad \text{and}\\ \widehat{\sigma^2}_{\boldsymbol{\eta}}^{[2]} &= 0.0000314.\end{aligned}$$

- The situation, in which the estimates $\widehat{\widetilde{\boldsymbol{\beta}}}_T$ and $\widehat{\widetilde{\boldsymbol{\beta}}}_T^{[2]}$ differ, are displayed in Figure 5.12.

- There can be found some outputs from the second run of the extended algorithm on the next pages: estimates $\widehat{\widetilde{\boldsymbol{\beta}}}_t^{[2]}$, based on $t = 1, 119, 238, 356, 475, 594, 712, 831, 950$ observations, are graphically displayed in Figures 5.13–5.15.

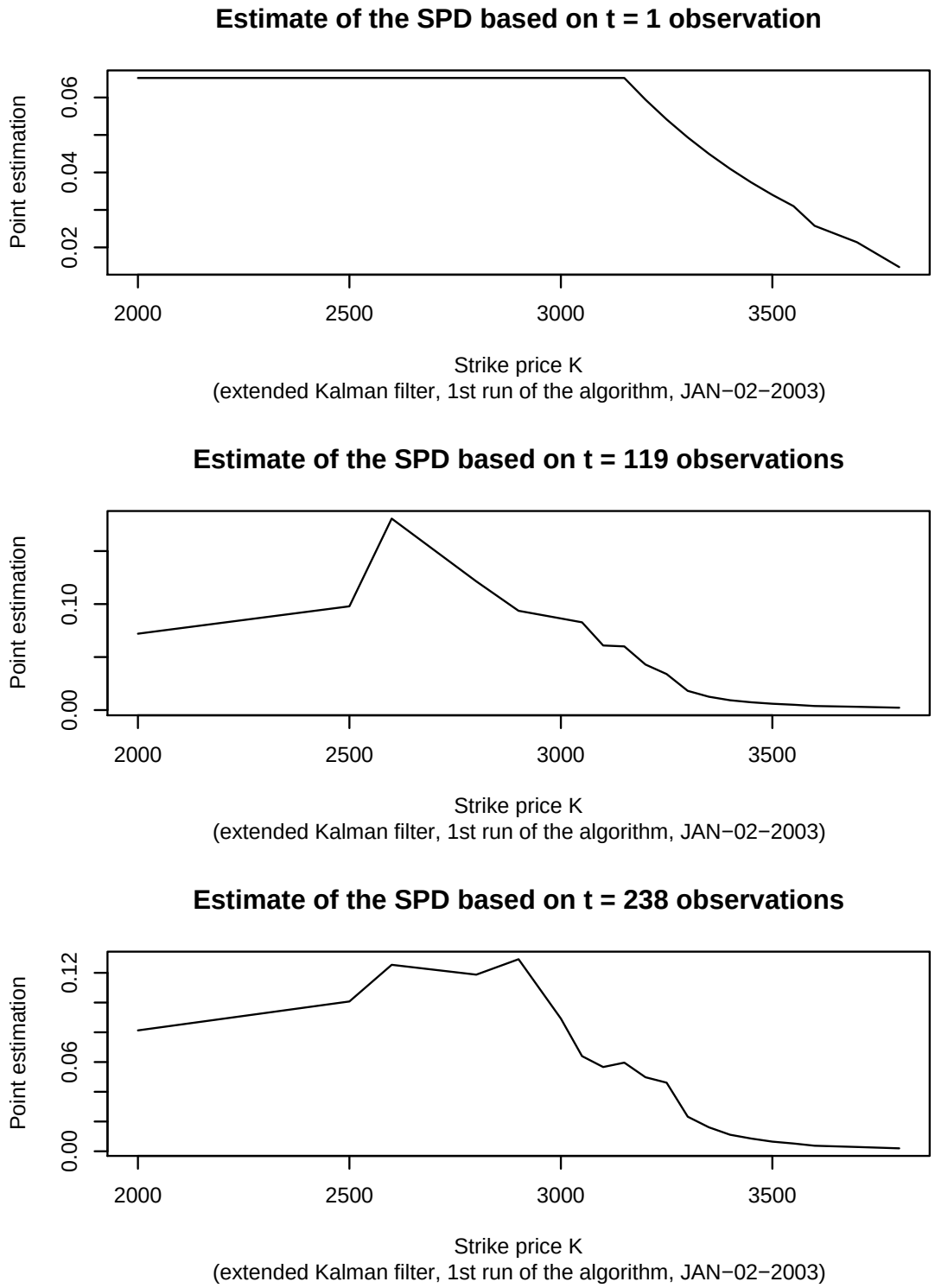


Figure 5.9: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_t$, based on $t_1 = 1, t_2 = 119$ and $t_3 = 238$ observed prices, satisfy the constraints.

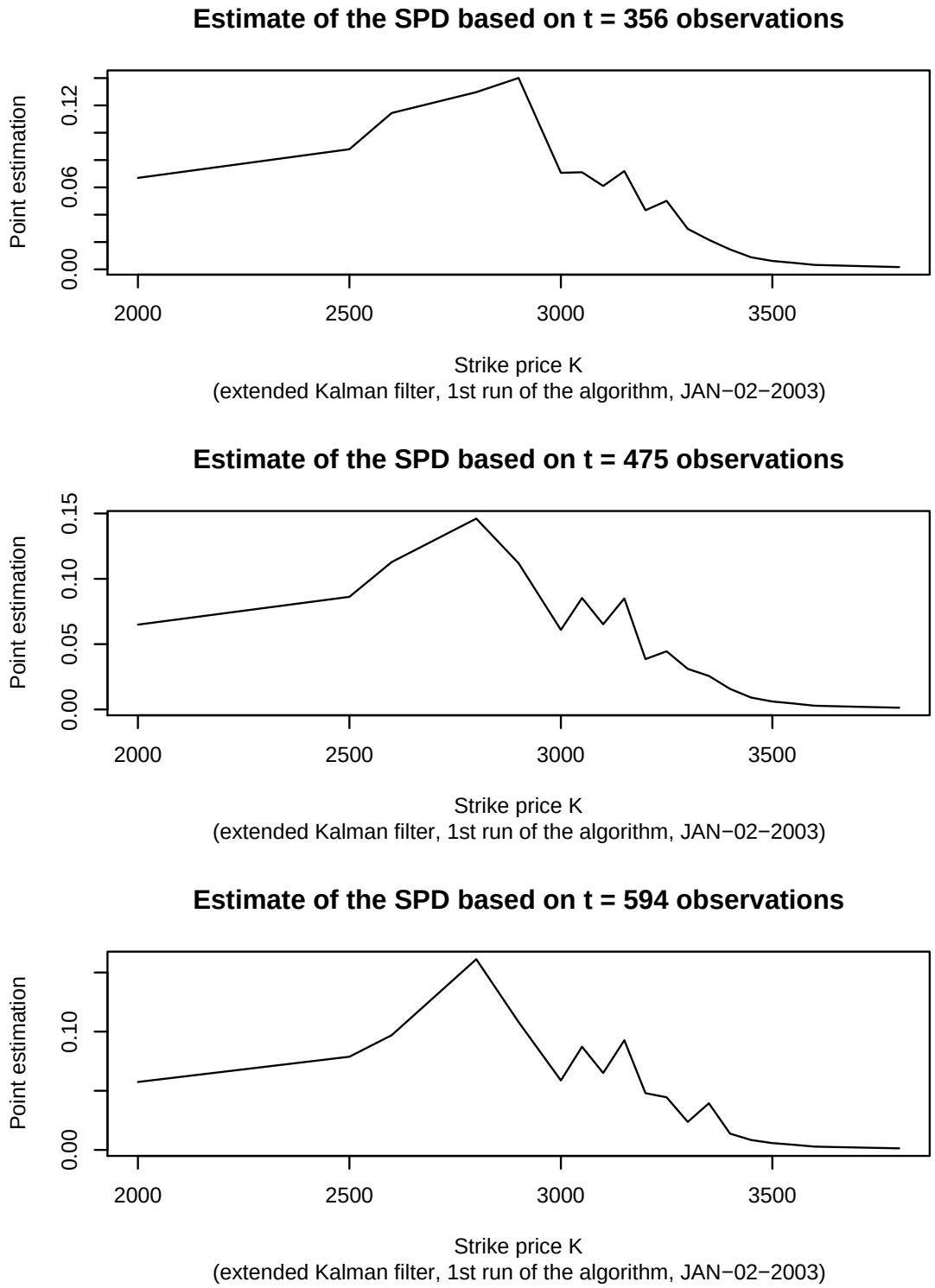


Figure 5.10: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_t$, based on $t_4 = 356, t_5 = 475$ and $t_6 = 594$ observed prices, satisfy the constraints.

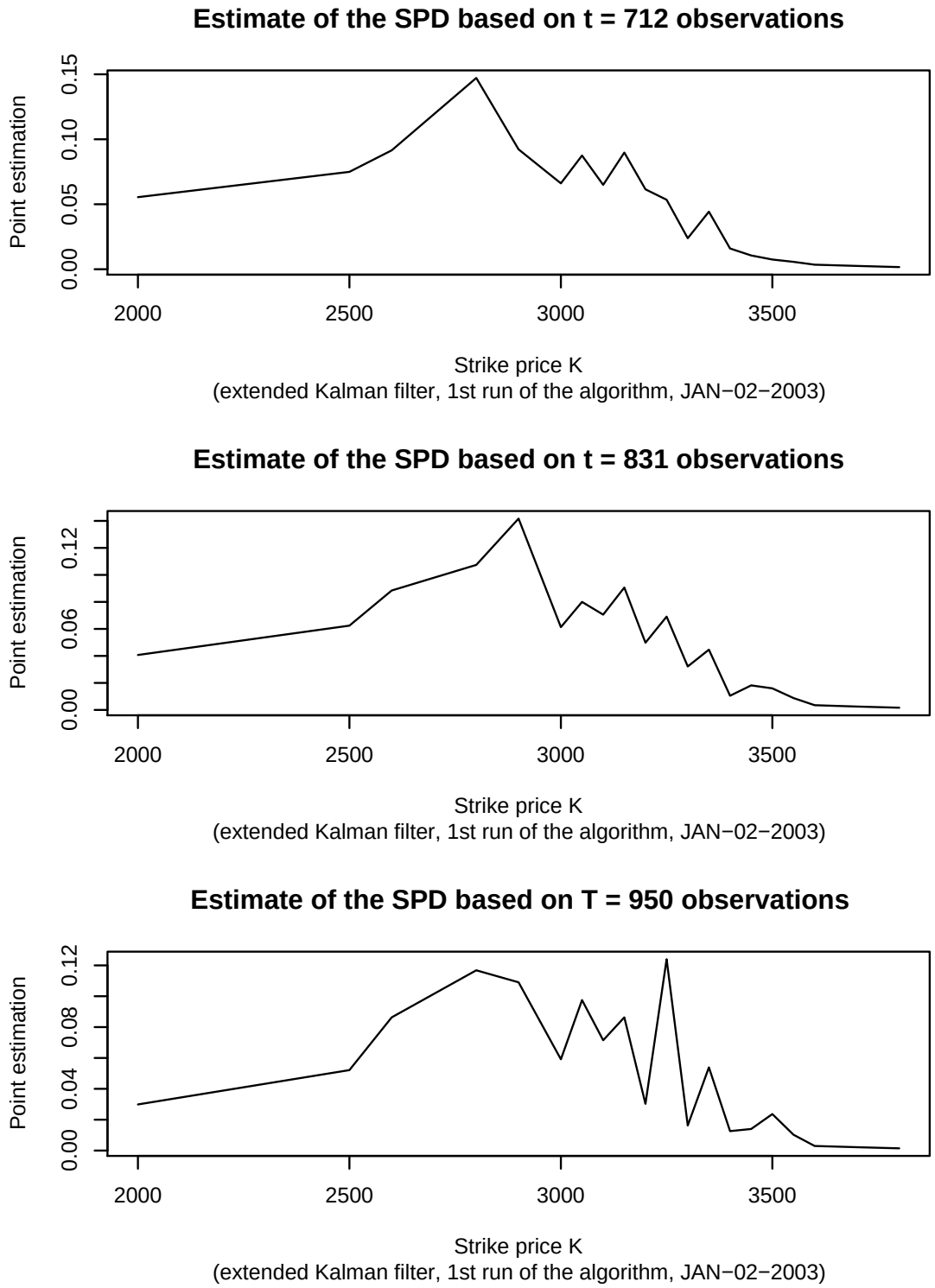


Figure 5.11: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_t$, based on $t_7 = 712, t_8 = 831$ and $t_9 = T = 950$ observed prices, satisfy the constraints.

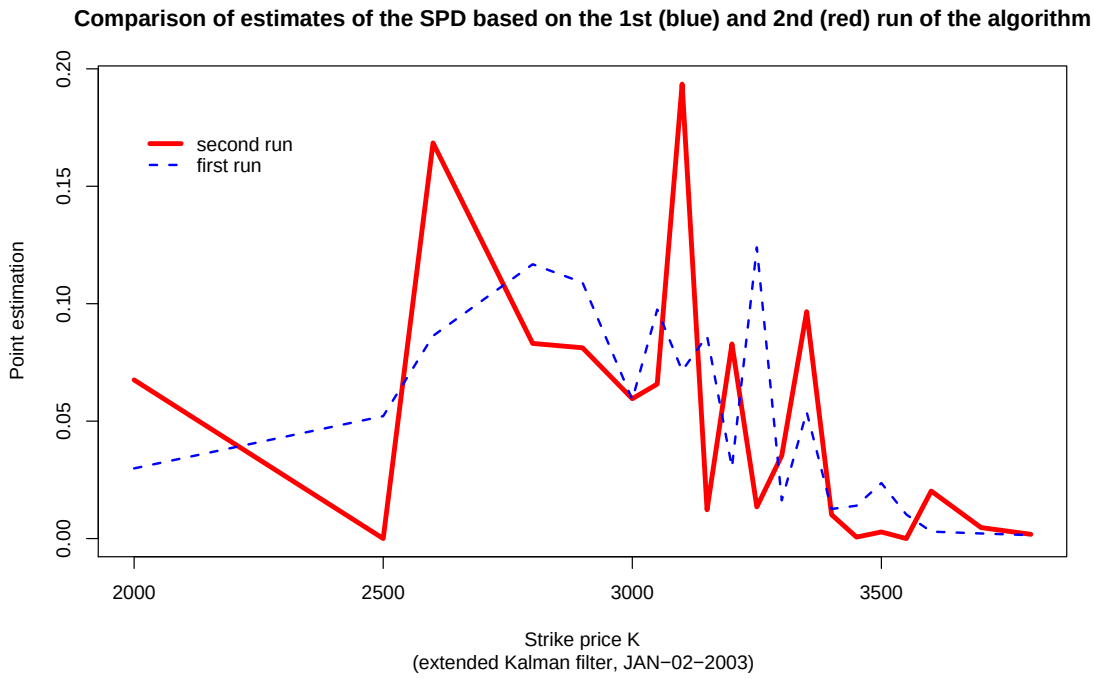


Figure 5.12: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_T$ and $\widehat{\widetilde{\boldsymbol{\beta}}}_T^{[2]}$, based on $T = 950$ observed prices, satisfy the constraints.

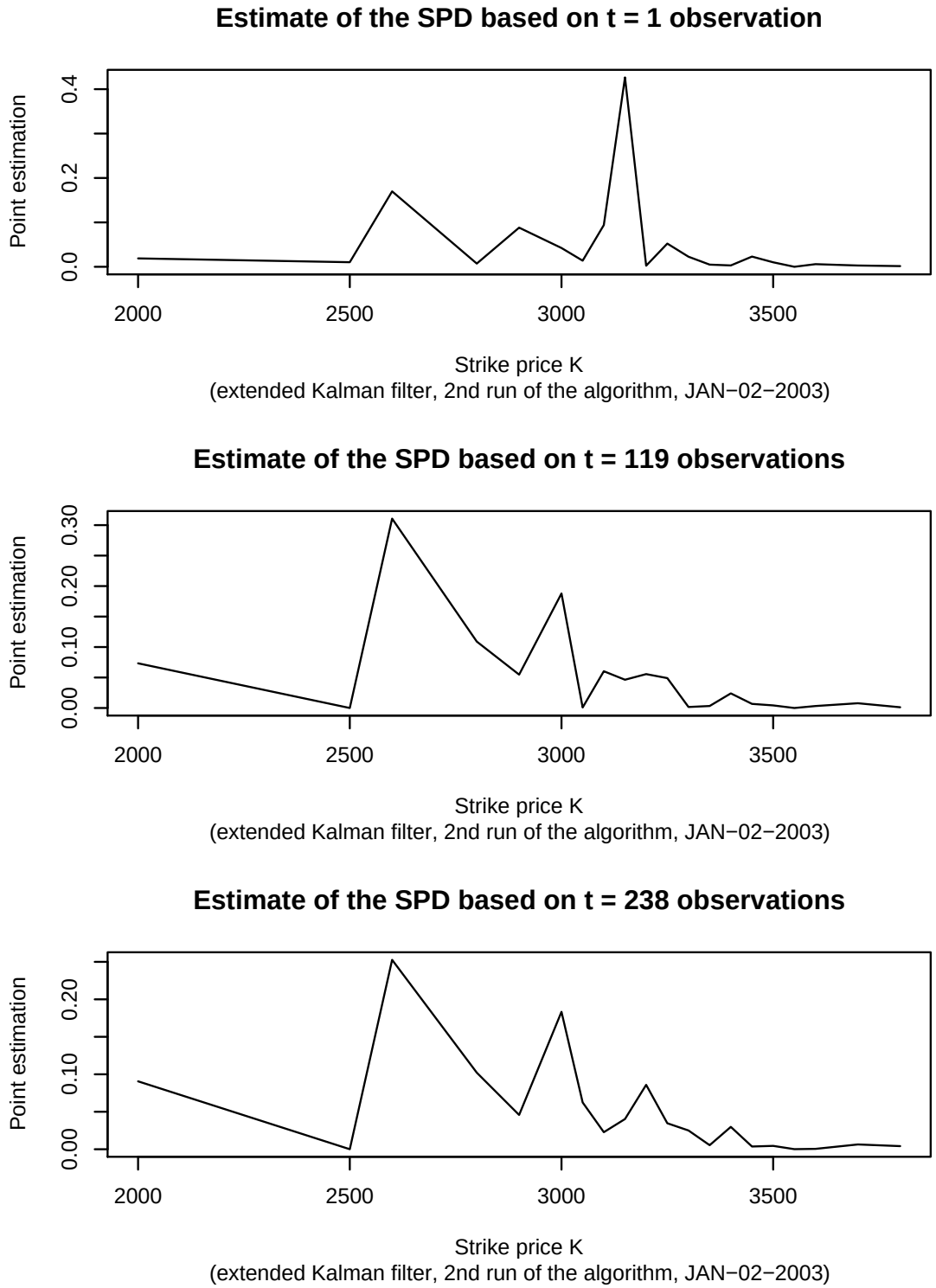


Figure 5.13: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_t^{[2]}$, based on $t_1 = 1, t_2 = 119$ and $t_3 = 238$ observed prices, satisfy the constraints.

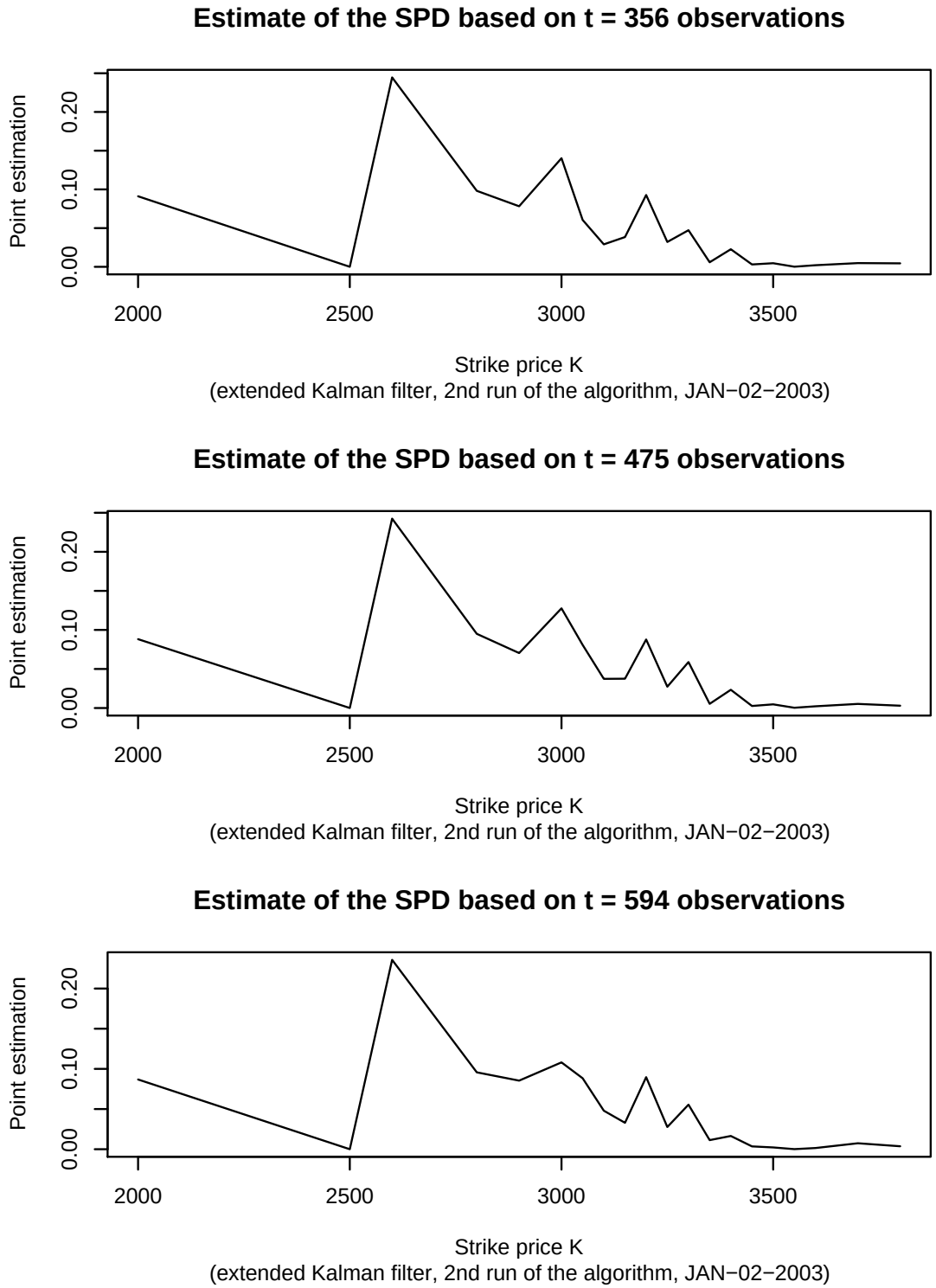


Figure 5.14: The constrained estimates of SPD against strike price K, $\widehat{\widetilde{\boldsymbol{\beta}}}_t^{[2]}$, based on $t_4 = 356, t_5 = 475$ and $t_6 = 594$ observed prices, satisfy the constraints.

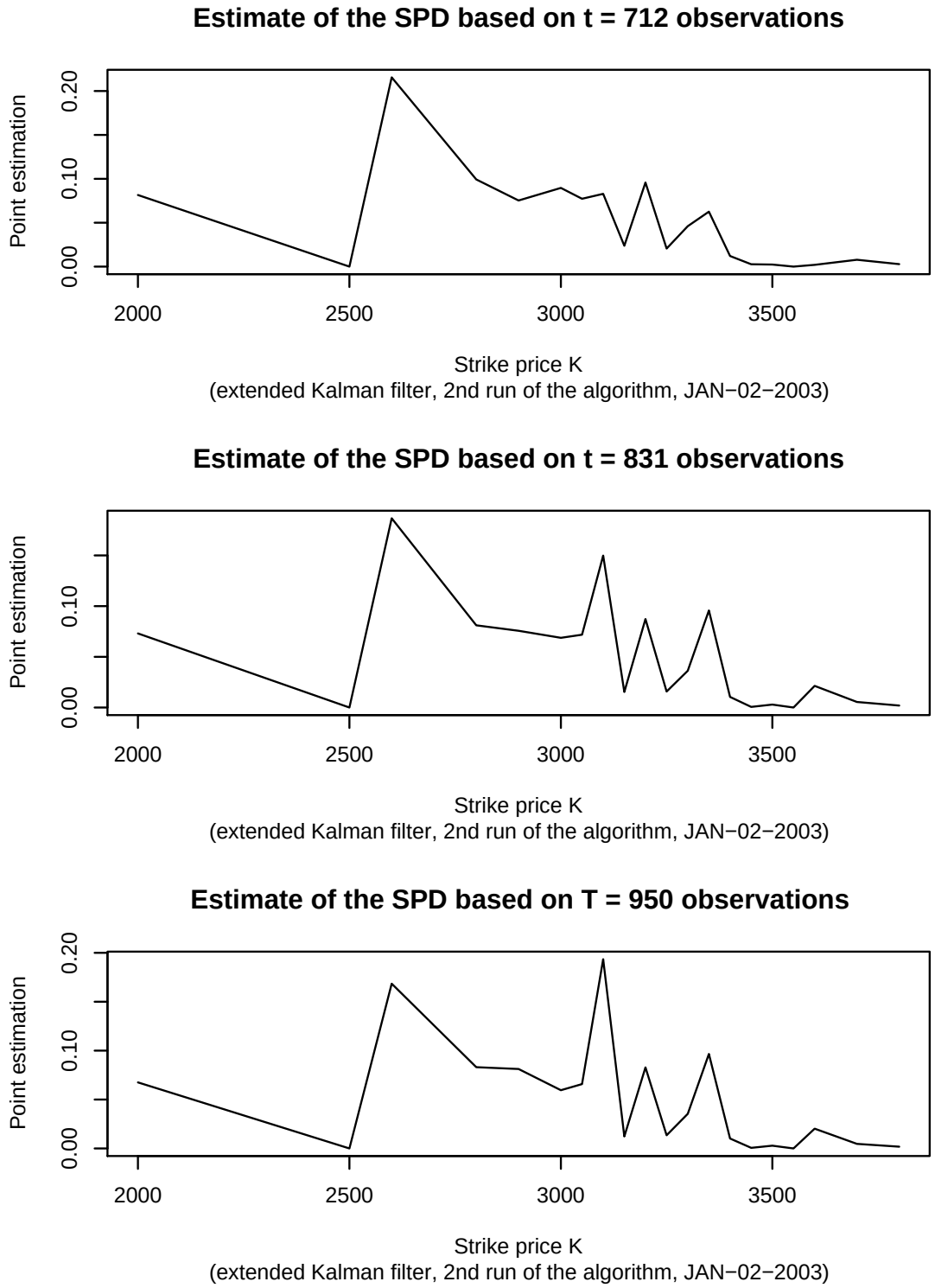


Figure 5.15: The constrained estimates of SPD against strike price K, $\widehat{\widehat{\boldsymbol{\beta}}}_t^{[2]}$, based on $t_7 = 712, t_8 = 831$ and $t_9 = T = 950$ observed prices, satisfy the constraints.

5.3 Comparison of the usual and the extended Kalman filter

One of the most interesting results probably is a comparison of the usual Kalman filter with its extended version.

- The situation, in which the estimates with and without constraints differ, is displayed in Figures 5.16 and 5.17.

Note that the difference between these two estimates of the state price density is surprisingly large.

– The unconstrained estimates, $\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\beta}}_T^{[2]}$, clearly show that the constraints are not satisfied (they include some negative components).

– The constrained versions, $\widehat{\widehat{\boldsymbol{\beta}}}_T$ and $\widehat{\widehat{\boldsymbol{\beta}}}_T^{[2]}$, satisfy the constraints and behave more reasonably. They are a little bit less stable than the unconstrained estimates.

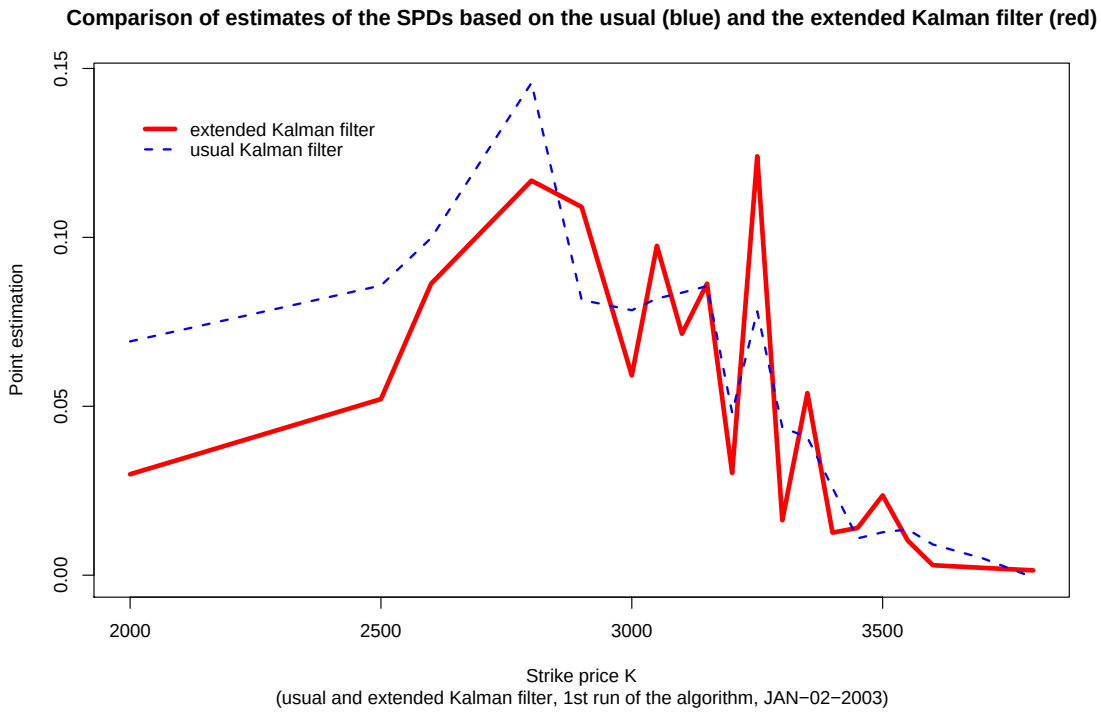


Figure 5.16: The constrained estimate of SPD, $\widehat{\widehat{\boldsymbol{\beta}}}_T$, satisfies the constraints although it is a little bit less stable than the unconstrained estimate $\widehat{\boldsymbol{\beta}}_T$.

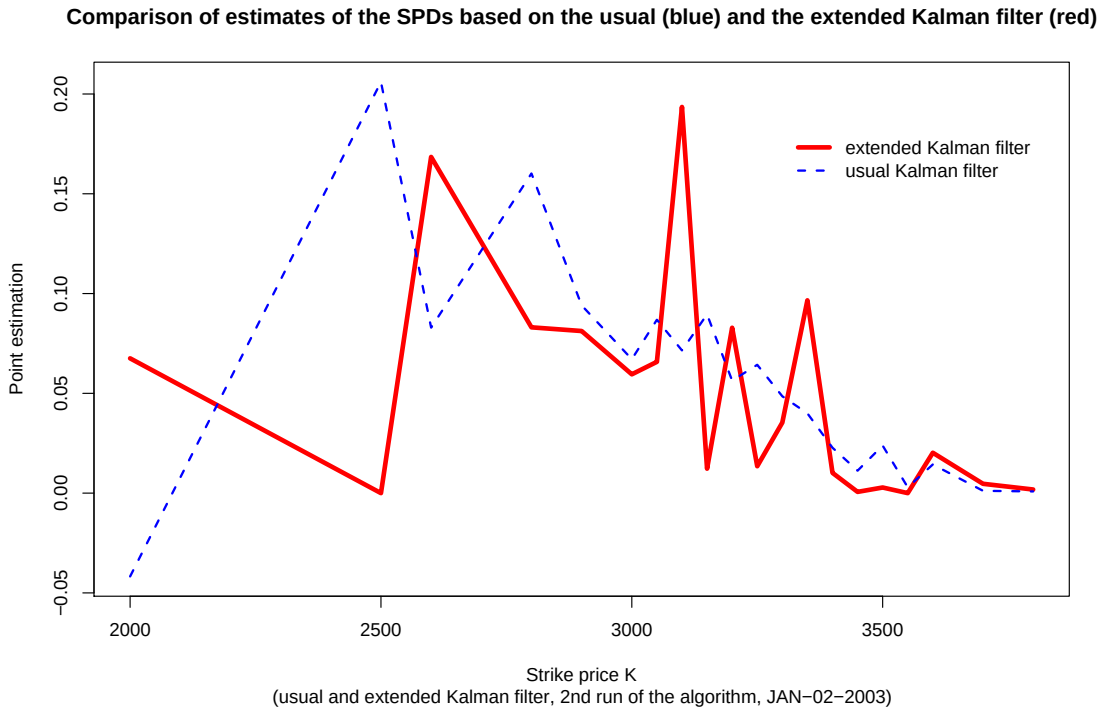


Figure 5.17: The constrained estimate of SPD, $\widehat{\widetilde{\boldsymbol{\beta}}}_T^{[2]}$, satisfies the constraints although it is a little bit less stable than the unconstrained estimate $\widehat{\boldsymbol{\beta}}_T^{[2]}$.

5.4 Confidence intervals

We present a simple methods for calculating confidence intervals for the SPD in the considered Gaussian model. These approaches are based on the use of the estimated variance-covariance matrix $\widehat{\boldsymbol{\Sigma}}_T$ computed within the usual and extended Kalman filter algorithm. The constructed confidence intervals determine the accuracy of the estimates $\widehat{\boldsymbol{\beta}}_T$ or $\widehat{\widetilde{\boldsymbol{\beta}}}_T$, respectively. The description of the x-axis in Figures 5.18 and 5.19 on the next page shows the number of observations at each of the observed strike price K.

1. **The unconstrained model:** Clearly, the confidence intervals based on the unconstrained model do not make much sense, and they do only if the constraints are – by change – satisfied. Even if this is the case, there is no guarantee that the lower bounds will be positive. In Figure 5.18, some of the lower bounds of the confidence intervals are visibly negative. We can observe large variability in regions with low number of observations.

2. **The constrained model:** In Figure 5.19, the confidence intervals based on the constrained model is displayed. We follow the intuitive approach: we construct the confidence intervals conditional on the fact that $\sum_{j=1}^{p} \exp(\beta_j) < 1$. Using maximum likelihood theory, we calculate confidence intervals for the parameters $\boldsymbol{\beta}_T$ $\big($re-scaled so that $\sum_{j=1}^{p} \exp(\beta_j) = 1\big)$. Exponentiating the limits of these confidence intervals leads to valid (non-negative) confidence intervals for parameters $\widetilde{\boldsymbol{\beta}}_T$. Another intuitive approach: the asymptotic normal distribution of $\widehat{\boldsymbol{\beta}}_T$ is assumed. This asymptotic distribution is re-calculated to the asymptotic distribution of the transformed vector $\widehat{\widetilde{\boldsymbol{\beta}}}_T$: the variance matrix $\widehat{\boldsymbol{\Sigma}}_T$ is multiplied by the Jacobian $\mathbf{B}_{t|t-1}$ from the left and by the transposed matrix $\mathbf{B}_{t|t-1}^{\top}$ from the right. We get a new variance-covariance matrix, whose diagonal elements can be used as the asymptotic variances of $\widehat{\widetilde{\boldsymbol{\beta}}}_T$. Then, the confidence intervals for $\widehat{\widetilde{\boldsymbol{\beta}}}_T$ (i.e. for the SPD) can easily be constructed the usual way. Thanks the exponential transformation via projection $\mathbf{g}_t(\boldsymbol{\beta}_t)$ derived in (4.2) cannot the valid confidence intervals be negative.

The unconstrained and the constrained methods logically lead to very different estimates. We can observe that the confidence intervals on the right hand side are much narrower for the constrained method. On the left hand side, both methods tend to provide confidence intervals that seem to be overly wide. For both methods, we observe that the length of the confidence intervals expands when the number of observation in that region is small. Please see the description of the x-axis.

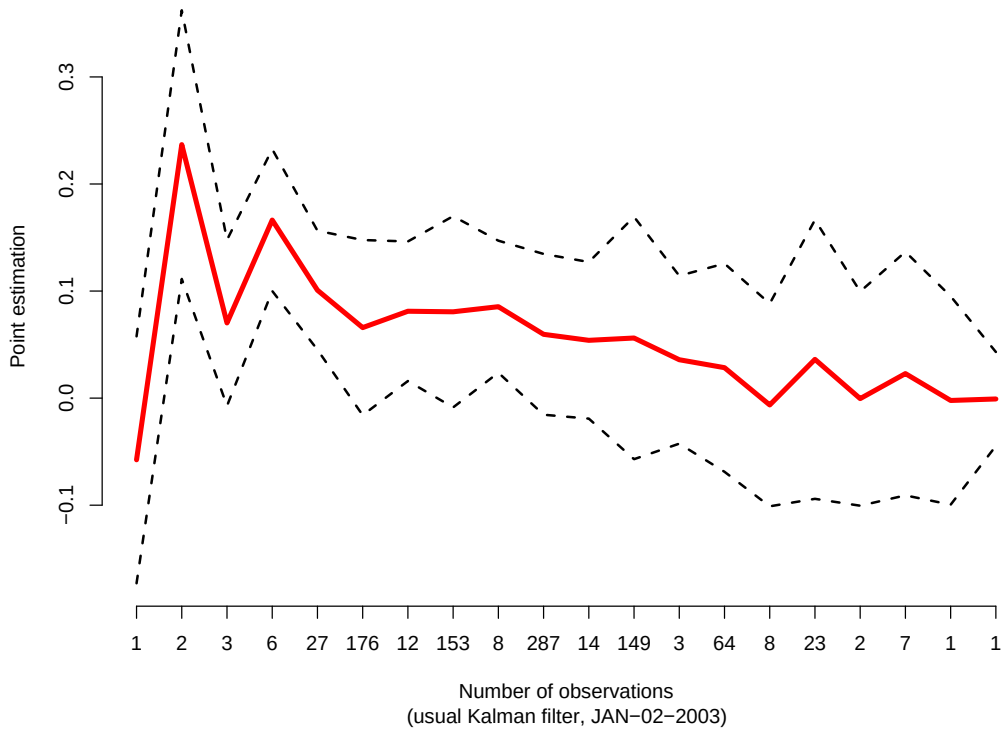


Figure 5.18: The unconstrained confidence interval for SPD.

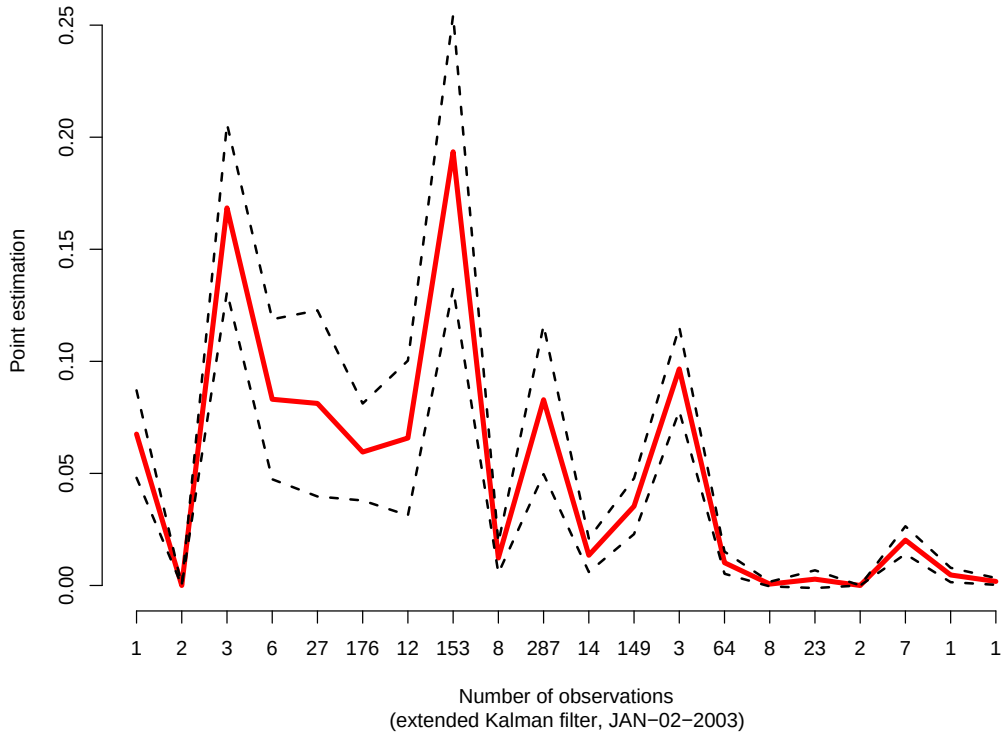


Figure 5.19: The constrained confidence interval for SPD.

5.5 Conclusion

We have proposed a simple model for the arbitrage free estimation of the SPD for a CALL option. Both versions of the Kalman filter – the usual and the extended Kalman filter – have been determined and tested out with real data. The designed estimation tools seem to be a very strong instrument to estimate the SPD. We have found a set of estimators $\widehat{\boldsymbol{\beta}}_T$, or $\widehat{\widetilde{\boldsymbol{\beta}}}_T$, respectively, as $(p \times 1)$ vectors for p various strike prices $\mathcal{K}$.

The coefficients $\widehat{\beta}_2, \ldots, \widehat{\beta}_{p-1}$, or $\widehat{\widetilde{\beta}}_2, \ldots, \widehat{\widetilde{\beta}}_{p-1}$, respectively, have been interpreted as follows: they can be described as estimates of the changes of the first derivative in that point. Therefore, these coefficients estimate probabilities associated with the corresponding strike price K. Thus, we can interpret the coefficients as a histogram-like estimator of the state price density.

The contribution to science in the field of option pricing

- Our proposed recursive procedures provide very good results. Consequently, they can be compared with other competing procedures (which are mostly mathematically much more complicated) estimating the SPD of a CALL option.
- The designed methods are stable and fast and can easily be used for estimation of the SPD of European CALL options.
- The approaches described in this thesis can even be recommended – because of their universality – for estimation of European PUT options, however another (much more complicated) experiment matrix $\boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}}$ must be defined and determined.

The results of this thesis could also be a basis for further academically interesting analyses: a residual analysis, a study of the covariance structure or a study of dynamics of SPD, respectively. Unfortunately, this all exceeds the framework of this thesis. Hence, these approaches (including study of PUT options) could be a "rewarding" theme for another master or even a doctoral thesis.

The designed algorithms have been implemented in the statistical program `R`.

Appendix

Usual Kalman filter algorithm

```
# Source code from the statistical program R
# Master thesis "Application of Kalman filtering", Marek Svojik
# Data analysis of CALL option prices using methods derived in Chapter 2


#############################################################################################


# input function variables

  # OptionPricesC         - standardized vector of CALL option prices
  # StrikePricesK         - vector of various strike prices K
  # BetaInitial           - the initial value of beta_{0|0}
  # NumberOfObservations  - for how many observations are estimates beta_t computed
  # Runs                  - how many times the usual Kalman filter runs


#############################################################################################
#############################################################################################


UsualKalmanFilter <- function(OptionPricesC,StrikePricesK,BetaInitial,NumberOfObservations,Runs)
 {

  SortedStrikePricesK <- sort(unique(StrikePricesK)) # vector of sorted "unique" strike prices K

  p <- length(SortedStrikePricesK)                   # p = # of various strike prices K
  T <- length(OptionPricesC)

  X <- matrix(0, nrow = p, ncol = p, byrow = TRUE)   # construction of the design matrix X
  X[,1] <- 1

  for (i in 1:(p-1))
   {
    for (j in 2:p)
    {
     if (p - i + 1 >= j) X[i,j] <- SortedStrikePricesK[p - j + 2] - SortedStrikePricesK[i]
    }
   }

  BETA <- matrix(0, nrow = T, ncol = p, byrow = TRUE)
  SIGMA <- matrix(0, nrow = p*T, ncol = p, byrow = TRUE)

  # BETA = matrix of all estimates beta_t in every time step t
  # SIGMA = matrix of all estimates sigma_t in every time step t



#############################################################################################

# 0. step: Setting of variances of eps_t and eta_t in the considered model:

        # C_t[k] = X*beta_t + eps_t, eps_t ~ N[0, var.eps]
        # beta_t = beta_{t-1} + eta_t, eta_t ~ N[0, var.eta]

  VarEps <- var(OptionPricesC)    # initialization of variance of eps_t
  VarEta <- diag(p)               # initialization of variance of eta_t
```

```
#############################################################################################

# 1. step: Initialization step [t = 0]; setting beta_{0|0}, sigma_{0|0}


  sigma.t <- diag(p)        # initialization of sigma_{0|0}
  beta.t <- rep(0,p)        # initialization of state vector beta_{0|0}

  beta.t[1] <- BetaInitial  # 1st component of beta_{0|0}
  beta.t[2:p] <- 1/(p-1)    # the second (p-1) components of beta_{0|0} are uniform distributed

  for (k in 1:Runs)         # Runs = how many times the Kalman filter runs
   {
    VarEpsSecond <- 0       # variance of eps_t in the 2nd run
    VarEtaSecond <- 0       # variance of eta_t in the 2nd run

    for (t in 1:NumberOfObservations)   # the main recursion, t -> t + 1
     {
      for (i in 1:p)        # corresponding row of X is selected
       {
        if (StrikePricesK[t] == SortedStrikePricesK[i]) X.SelectedRow <- X[i,]
       }

      BetaPrevious <- beta.t    # saving the previous value of beta_t

#############################################################################################

# 2. step: Prediction step, a priori estimates; beta_{t|t-1}, sigma_{t|t-1}

        # beta_{t|t-1} = beta_{t-1|t-1}
        # Sigma_{t|t-1} = Sigma_{t-1|t-1} + var(eta_t)

      P.t <- sigma.t + VarEta   # notation: P.t = Sigma_{t|t-1},  sigma.t = Sigma_{t-1|t-1}

#############################################################################################

# 3. step: Prediction error, its MSE; I_t, F_t


        # I_t = C_t[k] - X*beta_{t|t-1}
        # F_t = X*Sigma_{t|t-1}*t(X) + var(eps_t)

      Innovations <- OptionPricesC[t] - X.SelectedRow%*%beta.t
      F.t <- X.SelectedRow%*%P.t%*%X.SelectedRow + VarEps

#############################################################################################

# 4. step: Updating step; beta_{t|t}, Sigma_{t|t}, K_t


        # K_t = Sigma_{t|t-1}*t(X)*(F_t)^{-1}
        # beta_{t|t} = beta_{t|t-1} + K_t*I_t
        # Sigma_{t|t} = [I - K_t*X]*Sigma_{t|t-1}

      KalmanGain <- P.t%*%X.SelectedRow%*%(F.t)^{-1}
      beta.t <- beta.t + KalmanGain%*%Innovations
      sigma.t <- (diag(p) - KalmanGain%*%X.SelectedRow)%*%P.t

      VarEpsSecond <- VarEpsSecond + Innovations^2
      VarEtaSecond <- VarEtaSecond + (sum(beta.t) - sum(BetaPrevious))^2

      BETA[t,] <- beta.t
      SIGMA[(p*(t-1)+1):(p*t),] <- sigma.t


     }   # end of the main recursion for t = 1, ..., T
```

```
  VarEps <- (1/T)*VarEpsSecond                # new, more accurately variance of eps_t
  VarEta <- ((1/(T*p))*VarEtaSecond)*diag(p)  # new, more accurately variance of eta_t

  if (k == 1) BetaPreviousRun <- beta.t       # saving beta_t from the previous run


  }     # end of the cycle for k = 1, 2, ..., Runs


  return(beta.t)             # estimate beta.t we are looking for
  return(sum(beta.t[2:p]))   # checking the constraint


}       # end of the function
```

Extended Kalman filter algorithm

```
# Source code from the statistical program R
# Master thesis "Application of Kalman filtering", Marek Svojik
# Data analysis of CALL option prices using methods derived in Chapter 4


#############################################################################################


# input function variables

  # OptionPricesC         - standardized vector of CALL option prices
  # StrikePricesK         - vector of various strike prices K
  # BetaInitial           - the initial value of beta_{0|0}
  # NumberOfObservations  - for how many observations are estimates beta_t computed
  # Runs                  - how many times the extended Kalman filter runs


#############################################################################################
#############################################################################################


ExtendedKalmanFilter <- function(OptionPricesC,StrikePricesK,BetaInitial,NumberOfObservations,Runs)
 {

  SortedStrikePricesK <- sort(unique(StrikePricesK)) # vector of sorted "unique" strike prices K

  p <- length(SortedStrikePricesK)                   # p = # of various strike prices K
  T <- length(OptionPricesC)

  X <- matrix(0, nrow = p, ncol = p, byrow = TRUE)   # construction of the design matrix X
  X[,1] <- 1

  for (i in 1:(p-1))
   {
    for (j in 2:p)
    {
     if (p - i + 1 >= j) X[i,j] <- SortedStrikePricesK[p - j + 2] - SortedStrikePricesK[i]
    }
   }

  G <- matrix(0, nrow = T, ncol = p, byrow = TRUE)
  SIGMA <- matrix(0, nrow = p*T, ncol = p, byrow = TRUE)

  # G = matrix of all estimates g_t in every time step t
  # SIGMA = matrix of all estimates sigma_t in every time step t



#############################################################################################

# 0. step: Setting of variances of eps_t and eta_t in the considered model:

        # C_t[k] = X*g_t(beta_t) + eps_t, eps_t ~ N[0, var.eps]
        # beta_t = beta_{t-1} + eta_t, eta_t ~ N[0, var.eta]

  VarEps <- var(OptionPricesC)    # initialization of variance of eps_t
  VarEta <- diag(p)               # initialization of variance of eta_t

#############################################################################################
```

```
# 1. step: Initialization step [t = 0]; setting beta_{0|0}, sigma_{0|0}

  sigma.t <- diag(p)        # initialization of sigma_{0|0}
  beta.t <- rep(0,p)        # initialization of state vector beta_{0|0}

  beta.t[1] <- BetaInitial  # 1st component of beta_{0|0}
  beta.t[2:p] <- 1/(p-1)    # the second (p-1) components of beta_{0|0} are uniform distributed

  B.t <- matrix(0, nrow = p, ncol = p, byrow = TRUE)      # Jacobian matix B_{t|t-1}
  g.t <- rep(0,p)                                         # transformation of beta_t

  for (k in 1:Runs)         # Runs = how many times the Kalman filter runs
   {
    VarEpsSecond <- 0       # variance of eps_t in the 2nd run
    VarEtaSecond <- 0       # variance of eta_t in the 2nd run

    # state vector beta_t is transformed
    # S = sum_{j = 1}^{p-1} [exp(beta_t[j])], beta_t = (beta_t[0], ..., beta_t[p-1])

    S <- sum(exp(beta.t[2:p]))
    g.t[1] <- exp(beta.t[1])
    g.t[2:p] <- exp(beta.t[2:p])/S


    for (t in 1:NumberOfObservations)   # the main recursion, t -> t + 1
     {
      for (i in 1:p)                    # corresponding row of X is selected
       {
        if (StrikePricesK[t] == SortedStrikePricesK[i]) X.SelectedRow <- X[i,]
       }

      gPrevious <- g.t                  # saving the previous value of g_t

##########################################################################################################

# 2. step: Prediction step, a priori estimates; beta_{t|t-1}, sigma_{t|t-1}

      # beta_{t|t-1} = beta_{t-1|t-1}
      # Sigma_{t|t-1} = Sigma_{t-1|t-1} + var(eta_t)

      P.t <- sigma.t + VarEta   # notation: P.t = Sigma_{t|t-1},  sigma.t = Sigma_{t-1|t-1}

      # Computing of the Jacobian matrix B_{t|t-1}

      B.t[1,1] <- (S^2)*exp(beta.t[1])

      for (i in 2:p)
       {
        for (j in 2:p)
         {
          B.t[i,j] <- -exp(beta.t[i] + beta.t[j])
          if (i == j) B.t[i,i] <- exp(beta.t[i])*(S - exp(beta.t[i]))
         }
       }

      B.t <- (1/S^2)*B.t

##########################################################################################################

# 3. step: Prediction error, its MSE; I_t, F_t


        # I_t = C_t[k] - X*g_t(beta_{t|t-1})
        # F_t = X*B_{t|t-1}*Sigma_{t|t-1}*t(B_{t|t-1})*t(X) + var(eps_t)

      Innovations <- OptionPricesC[t] - X.SelectedRow%*%g.t
      F.t <- X.SelectedRow%*%B.t%*%P.t%*%t(B.t)%*%X.SelectedRow + VarEps
```

```
#############################################################################################################

# 4. step: Updating step; beta_{t|t}, Sigma_{t|t}, K_t


        # K_t = Sigma_{t|t-1}*t(B_{t|t-1})*t(X)*(F_t)^{-1}
        # beta_{t|t} = beta_{t|t-1} + K_t*I_t
        # Sigma_{t|t} = [I - K_t*X*]*B_{t|t-1}*Sigma_{t|t-1}

      KalmanGain <- P.t%*%t(B.t)%*%X.SelectedRow%*%(F.t)^{-1}
      beta.t <- beta.t + KalmanGain%*%Innovations
      sigma.t <- (diag(p) - KalmanGain%*%X.SelectedRow%*%B.t)%*%P.t


        # actual state vector beta_t is transformed to the actual estimate g_t

      S <- sum(exp(beta.t[2:p]))
      g.t[1] <- exp(beta.t[1])
      g.t[2:p] <- exp(beta.t[2:p])/S

      G[t,] <- g.t              # all estimates g_t
      SIGMA[(p*(t-1)+1):(p*t),] <- sigma.t

      VarEpsSecond <- VarEpsSecond + Innovations^2
      VarEtaSecond <- VarEtaSecond + (sum(g.t) - sum(gPrevious))^2

     }   # end of the main recursion for t = 1, ..., T


   VarEps <- (1/T)*VarEpsSecond                # new, more accurately variance of eps_t
   VarEta <- ((1/(T*p))*VarEtaSecond)*diag(p)  # new, more accurately variance of eta_t

   if (k == 1) gPreviousRun <- g.t             # saving g_t from the previous (1st) run

   }     # end of the cycle for k = 1, 2, ..., Runs

    return(g.t)               # estimate g.t we are looking for
    return(sum(g.t[2:p]))     # checking the constraint

 }       # end of the function
```

Bibliography

[1] Black F., Scholes M. (1973): The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**, 637-659.

[2] Breeden D., Litzenberger R. (1978): Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business* **51**, 621-651.

[3] Brown R. G., Hwang P. Y. C. (1996): Introduction to Random Signals and Applied Kalman Filtering. John Wiley & Sons, New York.

[4] Dupač V., Hušková M. (1999): Probability and Mathematical Statistics (in Czech). Karolinum, Prague.

[5] Franěk P. (2000): Adaptive Approach in the Kalman Filtering (in Czech). *Robust* **2000**, 59–64.

[6] Franke J., Härdle W., Hafner Ch. (2004): Statistics of Financial Markets II (in German). Springer, Berlin.

[7] Härdle W., Hlávka Z. (2005): Discuss Paper: Dynamics of State Price Densities. *Humboldt-Universität zu Berlin – SFB 649* **2005-021**, 1-29.

[8] Härdle W., Klinke S., Müller M. (2000): XploRe Learning Guide. Springer, Berlin.

[9] Härdle W., Simar L. (2003): Applied Multivariate Statistical Analysis. Springer, Berlin.

[10] Härdle W., Yatchew A. (2001): Dynamic Nonparametric State Price Density Estimation. *Humboldt-Universität zu Berlin – SFB 373*, 8-9.

[11] Hamilton J. D. (1994): Time Series Analysis. Princeton University Press, Princeton.

[12] Harrison J., Pliska S. (1981): Martingale and Stochastic Integral in the Theory of Continuous Trading. *Stochastic Processes and Their Applications* **11**, 215-260.

[13] Harvey A. C. (1989): Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

[14] Kalman R. E. (1960): A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering* **82**, Series D.

[15] Kellerhals B. P. (2001): Financial Pricing Models in Continuous Time and Kalman Filtering. Springer, Heidelberg.

[16] Künsch H. R. (2001): State Space and Hidden Markov Models. *Statistics and Applied Probability* **87**, Chapter 3.

[17] Lucas R. (1978): Asset Prices in an Exchange Economy. *Econometrica* **46**, 1429-1446.

[18] Merton R. (1973): Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* **4**, 141-183.

[19] Robertson T., Wright F. T., Dykstra R. L. (1988): Order Restricted Statistical Inference. Wiley, Chichester.

[20] Rubinstein M. (1976): The Valuation of Uncertain Income Streams and the Pricing of Options. *Bell Journal of Economics* **7**, 407-425.

[21] Welch G., Bishop G. (2004): An Introduction to the Kalman Filter. University of North Carolina Press, Chapel Hill.

ERRATA

Svojík M. (2007): Application of Kalman filtering (Master thesis). Charles University in Prague - Faculty of Mathematics and Physics, Prague

page	**change**	**to**
$9_{12,11}$	We say that the process is **adapted** if $\mathbf{X}_t \in \mathcal{F}_t$ for each time t, ... Further, if ...	We say that the process is **adapted** to the filtration $\mathbf{F}$ if $\mathbf{X}_t$ is $\mathcal{F}_t$-measurable for each time t. Further, if ...
10_{12}	()	(1.1)
17^3	< 0	$< \delta$
$31_{11,5}$	$\frac{\partial \mathbf{\Sigma}_{t\|t}}{\mathbf{K}_t}$	$\frac{\partial \mathbf{\Sigma}_{t\|t}}{\partial \mathbf{K}_t}$
38^7	For the sake of simplicity ... it will be assumed that the discount factor $e^{-r(T-t)} = 1$ ... Our data set ...	It will be assumed in the rest of this thesis that the discount factor $e^{-r(T-t)}$ in (2.2) is known (variables r, t and T are known and contained in our data set), so that the option prices $\mathbf{C}_t(K,T)$ on the left side of the equation in (2.2) can be divided by this factor. For the sake of simplicity of the following presentation, the notation of the option prices $\mathbf{C}_t(K,T)$ (now already divided by the discount factor) will be kept. Our data set ...

page	change	to
$42_{4,2}$	We further assume the normal distribution for the ... error term $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ ... The ... matrix $\text{Var}[\boldsymbol{\epsilon}_t] = \mathbf{H}_t(\boldsymbol{\psi})$ will now be known, assumed to be diagonal and heteroscedastic ...	We further assume the normal distribution for the ... error term $\boldsymbol{\epsilon}_t(\boldsymbol{\psi})$ (we can use for example the Shapiro-Wilk normality test) ... The ... matrix $\text{Var}[\boldsymbol{\epsilon}_t] = \mathbf{H}_t(\boldsymbol{\psi})$ will now be known, assumed to be diagonal and heteroscedastic (the Levene's test can be used) ...
$43_{4,2}$	We further assume the normal distribution for the ... noise term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$, ... variance matrix $\text{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t(\boldsymbol{\psi})$ ... is assumed to be diagonal and ... homoscedastic ...	We further assume the normal distribution for the ... noise term $\boldsymbol{\eta}_t(\boldsymbol{\psi})$ (we can again use the Shapiro-Wilk normality test), ... variance matrix $\text{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t(\boldsymbol{\psi})$ ... is assumed to be diagonal and ... homoscedastic (the Levene's test can be used) ...
54^{9}	... with the terms $\mathbf{g}\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ and $\mathbf{h}\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ denoting ...	... with the terms $\mathbf{g}_t\big(\boldsymbol{\xi}_t, \boldsymbol{\epsilon}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ and $\mathbf{h}_t\big(\boldsymbol{\xi}_{t-1}, \boldsymbol{\eta}_t(\boldsymbol{\psi}), \boldsymbol{\psi}\big)$ denoting ...
$54_{12,11}$	... the **non-linear** function $\mathbf{g}$ ... The **non-linear** function $\mathbf{h}$ in ...	... the **non-linear** function $\mathbf{g}_t$ ... The **non-linear** function $\mathbf{h}_t$ in ...
54_{8}	The most realistic models ...	The simplest models (we do not consider any dependence on $\boldsymbol{\psi}$) ...
68_{5}	... discount factor $e^{-r(T-t)}$ above (it was assumed to be 1, please ...) ...	... discount factor $e^{-r(T-t)}$ above (it was assumed to be known, please ...) ...
$70_{3,2,1}$	$\log L = -\frac{1}{2}\big\lvert\boldsymbol{\Sigma}_{0\lvert 0}\big\rvert - \ldots$ $-\frac{T}{2}\big\lvert\sigma^2_{\boldsymbol{\eta}}\boldsymbol{\mathcal{I}}\big\rvert - \ldots$ $-\frac{T}{2}\big\lvert\sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}\big\rvert - \ldots$	$\log L = -\frac{1}{2}\log\big\lvert\boldsymbol{\Sigma}_{0\lvert 0}\big\rvert - \ldots$ $-\frac{T}{2}\log\big\lvert\sigma^2_{\boldsymbol{\eta}}\boldsymbol{\mathcal{I}}\big\rvert - \ldots$ $-\frac{T}{2}\log\big\lvert\sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{\mathcal{I}}\big\rvert - \ldots$