6-6-2023

Survival Times and Investment Analysis with Dynamic Learning

Zhenzhen Li
Portland State University

Survival Times and Investment Analysis with Dynamic Learning

by

Zhenzhen Li

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Mathematical Sciences

Dissertation Committee:
Jong Sung Kim, Chair
Robert Fountain
Bin Jiang
Wayne Wakeland

Portland State University
2023

Abstract

The central statistical problem of survival analysis is to determine and characterize the conditional distribution of a survival time given a history of some observed health markers.

This dissertation contributes to the modeling of such conditional distributions in a setup where the health markers evolve randomly over time in a manner that can be represented by an Ito stochastic process, that is, a stochastic process that can be written as a sum of a time integral of some stochastic process and an Ito integral of some stochastic process, with both integrands subject to certain restrictions.

The random survival time is modeled as a deterministic function of a generating random variable that is related to the random evolution of the health markers, where the deterministic function is chosen so that the survival time has the desired distribution function.

The dissertation presents two families of such models. In the first family of models, the generating variable of the survival time is an Ito integral over the positive half-line, with the observable health marker at any given time represented by the same integral up to that time.

The second family of models involves a linear filtering framework, in which the generating variable affects linearly a number of observable health markers that evolve as Ito processes.

The dissertation offers formulas for conditional distribution, survival, and hazard functions of the survival time, and the relevance of each model is demonstrated with a simulation.

This application of a filtering model is not limited to the analysis of survival times. The dissertation shows that instead of a positive survival time we can use a return to a financial asset which can be positive or negative.

To apply the model in that context, we need to determine the distribution of log returns. The dissertation includes a goodness-of-fit investigation of some possible statistical distributions of a long history of log returns to the S & P 500 stock market index, concluding that we can use the generalized hyperbolic distribution to describe such returns.

With a number of investors, who think in terms of a normal distribution of log returns, providing the observable forecast markers, there is the problem of forcing the convergence of forecast markers to actual log returns at the end of the forecasting period. That problem is solved by using a multi-dimensional Brownian bridge process to model forecasting error.

Acknowledgments

I am grateful to all the people who helped make this dissertation possible.

First and foremost, I thank my dissertation advisor Dr. Jong Sung Kim, for his expertise, guidance, caring, patience, and encouragement. Dr. Kim has always made himself available to offer help, advice, and comments that significantly improved my research and this dissertation.

I thank members of my committee, Dr. Robert Fountain, Dr. Bin Jiang, and Dr. Wayne Wakeland, whose expertise, insights, advice, and encouragement have broadened and sharpened my knowledge and improved this dissertation.

I am grateful to the participants of the 2022 Statistics and Statistical Learning Seminer at Portland State University, where I presented part of my dissertation research, and in particular to Dr. Stephen Portnoy and Dr. Mara Tableman, for their thoughtful and helpful comments.

I extend my appreciation to Dr. Jeffrey Ovall, Director of the Ph.D. Program, and to Dr. Gerardo Lafferriere, Chair of the Fariborz Maseeh Department of Mathematics + Statistics, for their caring and support, and for helping me get the most out of the Ph.D. program.

I thank other departmental faculty whose courses I took, or with whom I had the opportunity to discuss mathematics and statistics, for guiding me and helping me learn, and to my fellow students and departmental staff who helped with many things.

I would also like to acknowledge financial support from the Fariborz Maseeh Department of Mathematics + Statistics during my initial years in the program.

Table of Contents

List of Tables

Table **Page**

List of Figures

Figure **Page**

1 Introduction

The central statistical problem of survival analysis is to determine and characterize the conditional distribution of a survival time given a history of some observed health markers.

A survival time, also called lifetime, time to outcome, time to event, time to failure, and random time, is a positive, unbounded, random variable.

In medical research survival time may be time from beginning of treatment to cure, time from exposure to risk factor to onset of medical condition, or time from diagnosis of medical condition to death.

In engineering and reliability theory survival time may be time to failure of an industrial system or its components, or the time to failure of a manufactured product, see for example Rausand et al. [29].

Economists analyze survival times of economic expansions and contractions, see for example Gamerman and West [20], Besedes and Prusa [3] and Tadeu et al. [37].

Survival analysis has natural applications in actuarial science, see for example Pitacco [28].

In sociology, survival times have been used to study the speed of diffusion of new ideas through social networks, see for example Wu et al. [42].

Frequently used existing models of survival analysis are the Kaplan-Meier empirical estimator, the Cox proportional hazards model, and the accelerated failure time model.

The Kaplan-Meier estimator is a complementary empirical distribution of survival time, adjusted for possible right censoring.

The proportional hazards model is a regression of log hazard function on a vector of covariates and an unknown baseline hazard function, and it can be estimated without the knowledge or estimation of the baseline hazard. This model provides ratios of hazard functions of pairs of patients rather than individual hazards.

The accelerated failure time model is an alternative to the proportional hazards model, and constitutes a regression of log survival time on a vector of covariates and log baseline survival time.

Those models are described in greater detail in Kalbfleisch and Prentice [23], Tableman and Kim [36], Aalen et al. [1], Collett [11], Emmert-Streib and Dehmer [17], and Ibrahim et al. [22].

The existing models do not account for the nature of the stochastic evolution of the health markers or covariates. In this dissertation, I introduce stochastic processes associated with stochastic differential equations, and methods of stochastic filtering, to provide new models grounded in the theory of probability and statistics, for capturing present knowledge about the future evolution of the health markers, in the calculation of conditional survival probabilities given the present value of the health markers.

The proposed new approach is an improvement over the widely used Cox proportional hazards model and the related accelerated failure time model. In addition to capturing the implications of the stochastic evolution of the health markers, the new models provide individual conditional survival functions and hazard functions, and work even when estimation of the proportional hazards model by partial likelihood fails. Finally, the new models are generalizable to richer setups that capture more complex stochastic behaviors of health markers and therapies.

Specifically, the research problem addressed in this dissertation is to develop, analyze, and describe applications of a model of random survival times and relevant, randomly arriving information, in which we can determine explicitly, at any desired time, the conditional distribution function, conditional survival function, and conditional hazard function of the survival time given the cumulative, randomly arriving information.

This dissertation describes two families of new models.

In the first family of models randomly arriving information is an Ito process that is a solution of a suitable stochastic differential equation involving the Wiener process, and where the survival time is generated by the limit of the solution Ito process at infinity.

A stochastic differential equation

$$dX_t = a(X_t, t)\,dt + b(X_t, t)\,dW_t$$

is short-hand notation for the integral equation in unknown stochastic process X_t

$$X_t = X_0 + \int_0^t a(X_s, s)\,ds + \int_0^t b(X_s, s)\,dW_s$$

where $a(x,t)$ and $b(x,t)$ are deterministic functions for which the integrals exist. Gihman and Skorohod [21] and Friedman [19] offer a rigorous discussion of stochastic differential equations.

For example, the stochastic process

$$X_t = X_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right]$$

is the solution of the stochastic differential equation

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

where μ is a constant, σ is a positive constant, and the initial value X_0 is a given log-normally distributed random variable. The solution X_t has a log-normal distribution.

Because the Wiener process W_t does not converge to a finite random variable when time t goes to infinity, the process X_t does not converge to a finite random variable.

The stochastic differential equation

$$dX_t = k(\mu - X_t)dt + \sigma dW_t$$

with constant μ and $k, \sigma > 0$, and a given initial, normally distributed, random variable X_0, has a solution, called the Ornstein-Uhlenbeck process, which, as t goes to infinity, converges to a normally distributed random variable with mean μ and variance $\frac{\sigma^2}{2k}$, which does not depend on the initial value X_0. Therefore, observing X_0, or more generally, observing the history of the process $X_u, 0 \leq u \leq s$ for some $s > 0$, does not help us predict the realization of the limit of X_t when t goes to infinity.

Next, consider a modified Ornstein-Uhlenbeck stochastic differential equation

$$dX_t = k\exp(-\beta t)(\mu - X_t)\,dt + \sigma\exp(-\beta t)\,dW_t$$

where β is a positive constant. This equation has a solution that converges to a finite random variable when time t goes to infinity, and the distribution of the limiting random variable depends on the initial value X_0. Therefore, observing the solution at some finite time helps us predict its value at infinity.

The construction of the first family of models of survival time utilizes stochastic differential equations whose solutions have the two properties: i) The solution converges to a finite random variable, and ii) The distribution of this limiting random variable depends on the initial value of the solution.

In such a situation, the survival time is defined as a certain deterministic function of the limiting random variable, giving the survival time the desired distribution, and the solution in finite time represents the randomly evolving health marker. The solution of the model is the conditional distribution of the survival time, given the observed history of the health marker.

The model is tractable because of the distributional characteristics of the Ito processes used, and produces explicit formulas for the conditional distribution function, survival function, and hazard function of the survival time, and delivers their random evolution over time.

The second family of models uses a normally distributed random variable θ that, again, generates a survival time with any desired distribution function. The random variable θ is unobservable, but the Ito stochastic process, representing a health marker or a vector of health markers, and defined by

$$d\xi_t = (A_0 + A_1\theta)\,dt + B_1\,dW_t$$

is assumed to be observable. The function, or vector and matrices A_0, A_1 and B_1 can be constants, functions of time, or, with some restrictions, functions of the health markers ξ_t and time. This sets up a filtering framework, due initially to Kalman [24] and Bucy and Joseph [9], and extended by many others, see for example Liptser and Shiryaev [27].

With suitable restrictions on the coefficients A_0, A_1, and B_1, the conditional distribution of the random variable, or random vector, θ, given the observed history $\mathcal{F}_t^{\xi}$ of the health marker(s) ξ_t, is normal, with conditional mean $m_t = \mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ and conditional variance $\gamma_t = \mathrm{var}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$, which are given as solutions of stochastic differential equations. Depending on the coefficients A_0, A_1 and B_1, the conditional variance γ_t may be a solution of a first order ordinary differential equation of Riccati type.

To demonstrate the relevance of this model the dissertation describes the results of a simulation study with a four-dimensional vector of health markers: Blood pressure, cholesterol, blood sugar, and hemoglobin.

This application of a filtering model is not limited to the analysis of survival times. The dissertation shows that instead of a positive survival time we can use a return to a financial asset which can be positive or negative.

To apply the model in that context, we need to determine the desired distribution of log returns. The dissertation includes a goodness-of-fit investigation of possible statistical distributions of a long history of log returns to the S & P 500 stock market index, concluding that we can use the generalized hyperbolic distribution to describe such returns.

With a number of investors, who think in terms of a normal distribution of log returns, providing the observable forecast markers, there is the problem of forcing the convergence of forecast markers to actual log returns at the end of the forecasting period. That problem is solved by using a multi-dimensional Brownian bridge process to model forecasting error. Critical to this modification is the fact that the unobservable random variable or vector θ in the observation equation can be replaced by a vector Ito process θ_t, so that we have a state equation of the form

$$d\theta_t \quad = \quad (a_0 + a_1\theta_t)\,dt + b_1\,dW_t$$

and an observation equation of the form

$$d\xi_t \quad = \quad (A_0 + A_1\theta_t)\,dt + B_1\,dW_t$$

Chapter 2 reviews the basic properties of random, or survival, times, the Wiener stochastic process, the Ito integral, and the Ito stochastic process. Subsequently, Ito processes are used to model dynamic health markers for a survival time, or dynamic forecast markers for financial investment, to compute the conditional distribution of a survival time, or stock index return, respectively.

Chapter 3 describes the first model, and starts by reviewing the concept of a stochastic differential equation and its solution, and characterizes solutions of stochastic differential equations that can be used to model a survival time and a dynamic health marker that can be used to compute the conditional distribution of the survival time given the observed history of the health marker.

Chapter 4 describes the results of a simulation study of the model constructed in Chapter 3. There are 30 patients for whom the survival time is the arrival time of some symptoms, and that survival time is generated by the limit of a solution of a certain stochastic differential equation. In addition, each patient possesses a health marker represented by the solution in finite time. The results of the model are described and compared with the results of the Cox proprtional hazards model.

Chapter 5 describes the stochastic filtering methodology that I propose to use to present a second family of models of random times and their conditional distributions given randomly arriving information. The chapter describes both a heuristic derivation of filtering equations and their relation to a familiar problem of conjugate distributions in Bayesian analysis, and some of the formal, general filtering context for the derivation of non-linear filtering equation and its common special case. The discussion includes both univariate and multivariate cases.

Chapter 6 describes the results of a simulation study of a model with one unobservable random variable and several observable health markers. There are 20 patients and four health markers: Diastolic blood pressure, high-density lipoprotein (HDL) cholesterol, fasting blood sugar level, and blood hemoglobin. The simulation includes the estimation of the parameters of the filtering model, the dynamics of the estimated conditional moments, and the resulting survival probabilities compared to actual survival times.

Chapter 7 describes the construction of a filtering model for the forecasting of stock market returns. Following an investigation of the distribution of a long history of log returns to the S & P 500 stock market index, the chapter describes a model with several forecast markers with dynamic noise represented by a multi-dimensional Brownian bridge process, to compute the conditional distribution of future log returns to the index.

2 General Concepts

2.1 Basic Properties of Random Times

2.1.1 Hazard Functions

A random time τ is a positive, unbounded random variable. In other words, τ satisfies $\mathbb{P}(\tau > 0) = 1$ and for every $t > 0$ we have $\mathbb{P}(\tau > t) > 0$.

In survival analysis, the complementary distribution function is called a survival or survivor function. If F is the distribution function of a random time τ, then the survival function is $S(t) = \mathbb{P}(\tau > t) = 1 - F(t)$.

If the distribution function of a random time τ has density $f(t)$, then we define a hazard function of τ as

$$\begin{aligned} h(t) &= \lim_{u \to 0} \frac{\mathbb{P}(\tau \leq t + u \big| \tau > t)}{u} \\ &= \lim_{u \to 0} \frac{F(t+u) - F(t)}{uS(t)} \\ &= \frac{f(t)}{S(t)} \end{aligned} \tag{2.1}$$

It is easy to see that

$$h(t) = -\frac{d \log S(t)}{dt} \tag{2.2}$$

Therefore, we can recover the distribution function from the hazard function

$$F(t) = 1 - \exp\left[-\int_0^t h(u)du\right] \tag{2.3}$$

The function $H(t) = \int_0^t h(u)du$ is called cumulative hazard function.

2.1.2 Mean Residual Life and Mean Tail Life

Mean Residual Life

For a random time τ and a positive, non-random time t denote

$$m(t) = \mathbb{E}\left(\tau \middle| \tau > t\right) - t \tag{2.4}$$

The function $m(t)$ is called mean residual life and describes conditional expected time to event in excess of time t given that time to event is greater than time t.

We can compute mean residual life for a random time τ with distribution function $F(t)$, density function $f(t)$, and survival function $S(t)$ as follows.

For $t < u$, the conditional probability distribution is

$$\mathbb{P}\left(\tau \leq u \middle| \tau > t\right) = \frac{F(u) - F(t)}{1 - F(t)} \tag{2.5}$$

Differentiating with respect to u, we get the conditional density $\frac{f(u)}{S(t)}$. Therefore

$$\mathbb{E}\left(\tau \middle| \tau > t\right) = \frac{\int_t^\infty u f(u)du}{S(t)} \tag{2.6}$$

Assuming that $\lim_{u\to\infty} uS(u) = 0$, we can rewrite the numerator on the right side of Equation (2.6)

$$\begin{aligned} \int_t^\infty uf(u)du &= \int_t^\infty udF(u) \\ &= -\int_t^\infty udS(u) \\ &= -uS(u)\Big|_t^\infty + \int_t^\infty S(u)du \\ &= tS(t) + \int_t^\infty S(u)du \end{aligned} \tag{2.7}$$

Using the definition in Equation (2.4) and Equation (2.6), we get

$$m(t) = \frac{\int_t^\infty S(u)du}{S(t)} \tag{2.8}$$

For example, when the distribution of the random time τ is Weibull with shape parameter k and rate parameter λ, the survival function is $S(t) = \exp\Big[-(\lambda t)^k\Big]$. Therefore

$$m(t) = \frac{\int_t^\infty \exp\Big[-(\lambda u)^k\Big]du}{\exp\Big[-(\lambda t)^k\Big]} \tag{2.9}$$

To compute the integral in the preceding equation, substitute $v = (\lambda u)^k$, then $dv = k\lambda(\lambda u)^{k-1}du$. Next, $\lambda u = v^{\frac{1}{k}}$, therefore, $(\lambda u)^{k-1} = v^{\frac{k-1}{k}}$. It follows that $du = \dfrac{dv}{k\lambda v^{\frac{k-1}{k}}}$. The integral becomes now

$$\frac{1}{k\lambda}\int_{(\lambda t)^k}^{\infty} e^{-v} v^{\frac{1}{k}-1} dv = \frac{1}{k\lambda}\Gamma\left[\frac{1}{k},(\lambda t)^k\right] \tag{2.10}$$

and, therefore

$$m(t) = \frac{\exp\left[(\lambda t)^k\right]}{k\lambda}\Gamma\left[\frac{1}{k},(\lambda t)^k\right] \tag{2.11}$$

where Γ is the upper incomplete gamma function. In the special case when $k = 1$, we get that mean residual life of an exponentially distributed random time is $m(t) = \frac{1}{\lambda}$, which reflects the memoryless nature of the exponential distribution.

Mean Tail Life

It is useful, especially for applications to investment losses, not to subtract t, and introduce a related concept in which the argument of mean residual life is changed from time t to the quantile $\alpha = S(t)$, and which I will call mean tail life. For $0 < \alpha \leq 1$, define

$$n(\alpha) = \mathbb{E}\left[\tau \middle| \tau > S^{-1}(\alpha)\right] \tag{2.12}$$

The definition implies that for $0 < \alpha \leq 1$ we have $n(\alpha) = m\left[S^{-1}(\alpha)\right] + S^{-1}(\alpha)$. In particular, for a strictly decreasing survival function S, we have $n(1) = m(0)$.

Using Equation (2.8) we can compute mean tail life from

$$n(\alpha) = \frac{\int_{S^{-1}(\alpha)}^{\infty} S(u)du}{\alpha} + S^{-1}(\alpha) \tag{2.13}$$

For example, the survival function of a Weibull distribution with shape parameter k and scale parameter λ is $S(t) = \exp\left[-(\lambda t)^k\right]$. Therefore

$$n(\alpha) = \frac{\int_{S^{-1}(\alpha)}^{\infty} \exp\left[-(\lambda u)^k\right] du}{\alpha} + S^{-1}(\alpha) \tag{2.14}$$

To compute the integral in the preceding equation, substitute $v = (\lambda u)^k$, then $dv = k\lambda(\lambda u)^{k-1} du$. Next, $\lambda u = v^{\frac{1}{k}}$, therefore, $(\lambda u)^{k-1} = v^{\frac{k-1}{k}}$. It follows that $du = \frac{dv}{k\lambda v^{\frac{k-1}{k}}}$. The integral becomes now

$$\frac{1}{k\lambda}\int_{-\log\alpha}^{\infty} e^{-v} v^{\frac{1}{k}-1} dv = \frac{1}{k\lambda}\Gamma\left[\frac{1}{k}, -\log\alpha\right] \tag{2.15}$$

where Γ above is the upper incomplete gamma function. The lower limit of integration is $-\log\alpha$ because when $u = S^{-1}(\alpha)$ then $(\lambda u)^k = -\log\alpha$.

To compute the additive term $S^{-1}(\alpha)$, start with $S(t) = \alpha = \exp\left[-(\lambda t)^k\right]$. Solving for t we get $t = S^{-1}(\alpha) = \frac{1}{\lambda}(-\log\alpha)^{\frac{1}{k}}$. Therefore

$$n(\alpha) = \frac{1}{k\lambda\alpha}\Gamma\left[\frac{1}{k}, -\log\alpha\right] + \frac{1}{\lambda}(-\log\alpha)^{\frac{1}{k}} \tag{2.16}$$

We can simplify Equation (2.16) by using the following property of the upper incomplete gamma function

$$\Gamma(x+1, y) = x\Gamma(x, y) + y^x \exp(-y) \tag{2.17}$$

Therefore

$$\Gamma\left(1+\frac{1}{k},-\log\alpha\right) = \frac{1}{k}\Gamma\left(\frac{1}{k},-\log\alpha\right)+\alpha(-\log\alpha)^{\frac{1}{k}} \tag{2.18}$$

and

$$\frac{1}{\lambda\alpha}\Gamma\left(1+\frac{1}{k},-\log\alpha\right) = \frac{1}{k\lambda\alpha}\Gamma\left(\frac{1}{k},-\log\alpha\right)+\frac{1}{\lambda}(-\log\alpha)^{\frac{1}{k}} \tag{2.19}$$

We have the simplified formula

$$n(\alpha) = \frac{1}{\lambda\alpha}\Gamma\left(1+\frac{1}{k},-\log\alpha\right) \tag{2.20}$$

Of course, if we parametrize the Weibull distribution with scale λ in Equation (2.20), we need to replace λ by $\frac{1}{\lambda}$. In the special case when $k = 1$, we get for the exponential distribution $n(\alpha) = \frac{1-\log\alpha}{\lambda}$. This simple case helps illustrate the difference between mean residual life and mean tail life.

2.2 Stochastic Processes

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, where Ω is a sample space, $\mathcal{A}$ is a σ-field of events, and $\mathbb{P}$ is a probability measure.

A stochastic process X_t is a family of random variables indexed by time $0 \le t < \infty$. The stochastic processes in this dissertation are the Wiener process, also called Brownian motion, and its extension to the so-called Ito process. Stochastic processes carry information, that I will use to condition the distribution of random times. I will list the definitions of those concepts below.

2.2.1 The Wiener Process and Its Natural Filtration

A Wiener process on a given probability space is a stochastic process W_t that can be written as a measurable function $W\colon \Omega \times \mathbb{R}_+ \to \mathbb{R}$, such that

1. $W(\omega, 0) = 0$ for all $\omega \in \Omega$.
2. For any $0 \leqslant s \leqslant t \leqslant u$ the random variables $W(\omega, u) - W(\omega, t)$ and $W(\omega, t) - W(\omega, s)$ are independent (the Wiener process has independent increments).
3. For any $t > 0$ the random variable $W(\omega, t)$ has normal distribution with zero mean and variance t.

It can be shown that a Wiener process exists, that its sample paths are continuous with probability one and not differentiable with probability one Borodin and Salminen [5].

It is useful to formalize the concept of information carried by the Wiener process. The definition is general and applies to any stochastic process.

A filtration is an increasing family $\{\mathcal{F}_t; t \in \mathbb{R}_+\}$ of σ-fields $\mathcal{F}_t \subset \mathcal{A}$, where $\mathcal{F}_t$ contains events that are decidable at time t. Increasing here means that information becomes increasingly refined over time, that is, if $s \leqslant t$ then $\mathcal{F}_s \subset \mathcal{F}_t$.

A filtration satisfies the usual conditions if and only if:

1. It is right continuous, meaning that for every $t \geq 0$ we have $\bigcap_{u>t} \mathcal{F}_u = \mathcal{F}_t$. The interpretation of this condition is that information at time t is exactly equal to information just after t, there is no jump in information right after time t. The intersection $\bigcap_{u>t} \mathcal{F}_u$ represents information right after t, and is sometimes denoted by $\mathcal{F}_{t+}$.

2. It satisfies $\sigma\left(\bigcup_{t\geq 0}\mathcal{F}_t\right)\subset\mathcal{A}$. The σ-field $\sigma\left(\bigcup_{t\geq 0}\mathcal{F}_t\right)$ is frequently denoted $\mathcal{F}_\infty$. This requirement means that all decidable events are in $\mathcal{A}$.

3. It is complete, meaning that $\mathcal{F}_0$ contains all the negligible sets of $\mathcal{A}$. Notice that because $\mathcal{F}_0\subset\mathcal{F}_t$ for every $t\geq 0$, we have that every $\mathcal{F}_t$ contains all the negligible sets of $\mathcal{A}$.

A probability space $(\Omega,\mathcal{A},\mathbb{P})$ equipped with a filtration $\mathcal{F}_t$ is called a filtered probability space and denoted $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F}_t)$

For example, the natural filtration of a Wiener process W_t is an increasing family of σ-fields $\mathcal{F}_t$ such that $\mathcal{F}_t$ is the smallest σ-field with respect to which all the random variables $\left\{W_s\middle|0\leqslant s\leqslant t\right\}$ are measurable.

The natural filtration of a Wiener process satisfies the usual conditions.

For a further discussion of the Wiener process see, for example Revuz and Yor [32] and Borodin and Salminen [5].

2.2.2 Martingales on a Filtered Probability Space

A random variable X on the probability space $(\Omega,\mathcal{A},\mathbb{P})$ is called integrable if and only if $\mathbb{E}\left(\left|X\right|\right)<\infty$. The normed space of integrable random variables on $(\Omega,\mathcal{A},\mathbb{P})$ is denoted $\mathcal{L}^1(\Omega,\mathcal{A},\mathbb{P})$.

A stochastic process X_t on a filtered probability space $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F}_t)$ is called adapted to the filtration $\mathcal{F}_t$ if and only if for every $t\geqslant 0$ the random variable X_t is $\mathcal{F}_t$-measurable.

Let X be an integrable random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{G}$ a sub-σ-field of $\mathcal{A}$. We say that the random variable Y is a conditional mean of X given $\mathcal{G}$ if and only if:

1. Y is $\mathcal{G}$-measurable.

2. Y is integrable.

3. For every $A \in \mathcal{G}$ we have $\mathbb{E}(\mathbb{1}_A Y) = \mathbb{E}(\mathbb{1}_A X)$.

The existence of conditional mean is guaranteed by the Radon-Nikodym theorem which says that if $\mathbb{Q}$ is a second probability measure on $(\Omega, \mathcal{A})$ such that, if $\mathbb{P}(A) = 0$ then $\mathbb{Q}(A) = 0$, then there is an almost surely unique random variable ξ, called the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$, such that $\mathbb{Q}(A) = \mathbb{E}(\mathbb{1}_A \xi)$.

Let X be a positive random variable, otherwise we separate X into positive and negative parts. On $\mathcal{G}$ define the probability measure $\mathbb{Q}(A) = \mathbb{E}(\mathbb{1}_A X)$. Then the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is the conditional expectation $\mathbb{E}\left(X \middle| \mathcal{G}\right)$.

Conditional mean has the following tower property: If the sub-σ-fields $\mathcal{G}$ and $\mathcal{H}$ of $\mathcal{A}$ are such that $\mathcal{G} \subset \mathcal{H}$ then $\mathbb{E}\left[\mathbb{E}\left(X \middle| \mathcal{H}\right) \middle| \mathcal{G}\right] = \mathbb{E}\left(X \middle| \mathcal{G}\right)$. The tower property says that a coarse average of a fine average is the coarse average.

On the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ let $A \in \mathcal{A}$ and $\mathcal{G}$ be a sub-σ-field of $\mathcal{A}$. The conditional probability $\mathbb{P}\left(A \middle| \mathcal{G}\right)$ is the conditional mean $\mathbb{E}\left(\mathbb{1}_A \middle| \mathcal{G}\right)$.

A stochastic process X_t on a filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$ is called a $(\mathbb{P}, \mathcal{F})$-martingale if and only if:

1. The process X_t is adapted to the filtration $\mathcal{F}_t$.

2. For each $t \geqslant 0$ the random variable X_t is integrable.

3. For each $0 \leqslant s \leqslant t$ we have, $\mathbb{P}$-almost-surely the equality

$$\mathbb{E}\left(X_t \middle| \mathcal{F}_s\right) = X_s \tag{2.21}$$

If conditions 1 and 2 are satisfied, and instead of condition 3 we have the inequality $\mathbb{E}\left(X_t \middle| \mathcal{F}_s\right) \geq X_s$, then the process X_t is called a sub-martingale.

For example, the Wiener process W_t is a martingale with respect to its natural filtration. Indeed, $W_t - W_s$ is independent of $\mathcal{F}_s$, and therefore

$$\mathbb{E}\left(W_t - W_s \middle| \mathcal{F}_s\right) = \mathbb{E}\left(W_t - W_s\right) = 0 \tag{2.22}$$

Other examples of martingales on the natural filtration the Wiener process are $W_t^2 - t$ and $\exp\left(-\frac{1}{2}t + W_t\right)$.

A martingale has uncorrelated increments. Indeed, for $0 \leqslant u \leqslant v \leqslant s \leqslant t$ we have by the tower property of conditional expectation

$$\mathbb{E}\left[(X_v - X_u)(X_t - X_s)\right] = \mathbb{E}\left\{\mathbb{E}\left[(X_v - X_u)(X_t - X_s) \middle| \mathcal{F}_v\right]\right\} \tag{2.23}$$

By pullout

$$\mathbb{E}\left\{\mathbb{E}\left[(X_v - X_u)(X_t - X_s) \middle| \mathcal{F}_v\right]\right\} = \mathbb{E}\left\{(X_v - X_u)\mathbb{E}\left[(X_t - X_s) \middle| \mathcal{F}_v\right]\right\} \tag{2.24}$$

By another application of the tower property

$$\mathbb{E}\left[(X_t - X_s) \middle| \mathcal{F}_v\right] = \mathbb{E}\left[\mathbb{E}\left(X_t - X_s \middle| \mathcal{F}_s\right) \middle| \mathcal{F}_v\right] = 0 \tag{2.25}$$

Finally, $\mathbb{E}(X_v - X_u) = \mathbb{E}\left[\mathbb{E}\left(X_v - X_u \middle| \mathcal{F}_u\right)\right] = 0$ and similarly $\mathbb{E}(X_t - X_s) = 0$, so that the increments $X_v - X_u$ and $X_t - X_s$ are uncorrelated. It follows that a martingale represents cumulative serially uncorrelated, but possibly serially dependent, noise.

2.2.3 Ito Integrals

Informally, an Ito integral is a special integral whose integrator is a Wiener process. The integral is special because the Wiener process has infinite variation on finite time intervals with probability one, and the usual construction of the Stieltjes integral fails for such integrators.

In addition, the quadratic variation of the Wiener process is positive, which introduces another difference between the Ito integral and the Stieltjes integral. This difference appears even with simple integrands. For example, if g is a continuous, bounded function of time (on finite intervals) such that $g(0) = 0$, then

$$\int_0^t g(s)\, dg(s) = \frac{g^2(t)}{2} \tag{2.26}$$

By contrast, if W_t is a Wiener process, which is also a continuous, bounded function of time (a.s. on finite intervals) and satisfies $W_0 = 0$, then we have

$$\int_0^t W_s\, dW_s = \frac{W_t^2 - t}{2} \tag{2.27}$$

Consider a partition $0 = t_0 < t_1 < \ldots < t_n = t$ of the interval $\left[0, t\right]$ and the sum

$$\sum_{i=1}^{n} W_{t_{i-1}}\left(W_{t_i} - W_{t_{i-1}}\right) = W_{t_0}\left(W_{t_1} - W_{t_0}\right) + \cdots + W_{t_{n-1}}\left(W_{t_n} - W_{t_{n-1}}\right) \tag{2.28}$$

From the identity

$$\begin{aligned} 2W_{t_{i-1}}\left(W_{t_i} - W_{t_{i-1}}\right) &= 2W_{t_{i-1}}\Delta W_{t_i} \\ &= \left(W_{t_{i-1}} + \Delta W_{t_i}\right)^2 - W_{t_{i-1}}^2 - \left(\Delta W_{t_i}\right)^2 \end{aligned} \tag{2.29}$$

We get the approximating sum of the integral $2\int_0^t W_s\,dW_s$

$$2\sum_{i=1}^{n} W_{t_{i-1}}\Delta W_{t_i} = \sum_{i=1}^{n}\left[\left(W_{t_{i-1}} + \Delta W_{t_i}\right)^2 - W_{t_{i-1}}^2\right] - \sum_{i=1}^{n}\left(\Delta W_{t_i}\right)^2 \tag{2.30}$$

The first sum on the right side of Equation (2.30) telescopes into $W_t^2 - W_0^2 = W_t^2$, and the second sum is interesting. If we consider for a moment $W_t = g(t)$ to be a Stieltjes-integrable function, then the second sum goes to zero. If, however, W_t is a Wiener process, then the second sum goes to t, the quadratic variation of the Wiener process on the interval $\left[0, t\right]$. That is how we get the difference between Equation (2.26) and Equation (2.27).

It is important to recognize that one of the aspects of the Ito integral is that in the approximating sum the integrand is evaluated at the left point of the partition interval, therefore $W_{t_{i-1}}\Delta W_{t_i}$ but not $W_{\tau_i}\Delta W_{t_i}$ where $t_{i-1} \leqslant \tau_i \leqslant t_i$, as in the definition of the Stieltjes integral. In general, because of the infinite variation of the Wiener process, the sum $W_{\tau_i}\Delta W_{t_i}$ converges only for $\tau_i = t_{i-1}$. The silver lining of that restriction is that, given suitable technical restrictions on the integrand function, the Ito integral is a martingale.

Formally, let $\mathcal{F}_t$ be a family of σ-fields generated by the Wiener process W_t, such family is called the natural filtration of W_t. Let $g(t)$ be a separable, progressively measurable stochastic process that is adapted to the filtration $\mathcal{F}_t$ and such that $\mathbb{P}\left(\int_0^t g^2(s)\,ds < \infty\right) = 1$. Then the Ito integral $\int_0^t g(s)\,dW_s$ can be defined and satisfies for admissible integrands g and h and constants α and β

$$\int_0^t \left[\alpha\, g(s) + \beta\, h(s)\right] dW_s \quad = \quad \alpha \int_0^t g(s)\,dW_s + \beta \int_0^t h(s)\,dW_s \tag{2.31}$$

If, in addition, the stochastic process $g(t)$ satisfies $\mathbb{E}\left(\int_0^t g^2(s)\,ds\right) < \infty$ then the Ito integral has the following three properties:

1. $\mathbb{E}\left(\int_0^t g(s)\,dW_s\right) = 0$.
2. $\mathbb{E}\left[\int_0^t g(s)\,dW_s\right]^2 = \mathbb{E}\left[\int_0^t g^2(s)\,ds\right]$.
3. The stochastic process $X_t = \int_0^t g(s)\,dW_s$ is a continuous $\mathcal{F}_t$-martingale.

Property 2 above is called Ito isometry.

For further details of the definitions and properties of Ito integrals see, for example, Friedman [19] and Revuz and Yor [32].

2.2.4 Ito Processes

Let $a(t)$ be a separable, progressively measurable, and adapted process such that $\mathbb{P}\left(\int_0^t |a(s)|\,ds < \infty\right) = 1$, and $b(t)$ be a separable, progressively measurable, and adapted process such that $\mathbb{P}\left(\int_0^t b^2(s)\,ds < \infty\right) = 1$. Then the process

$$X_t = X_0 + \int_0^t a(s)\,ds + \int_0^t b(s)\,dW_s \tag{2.32}$$

is called an Ito process. The Ito process in Equation (2.32) is frequently written in the shorthand notation

$$dX_t = a(t)\,dt + b(t)\,dW_t \tag{2.33}$$

For example, Equation (2.27) can be written

$$W_t^2 = \int_0^t ds + \int_0^t 2W_s\,dW_s \tag{2.34}$$

or in shorthand notation, $dW_t^2 = dt + 2W_t\,dW_t$. Therefore, W_t^2 is an Ito process.

The Ito integral with respect to a Wiener process can be extended to an Ito integral with respect to an Ito process. We define an integral with respect to an Ito process

$$\int_0^t g(s)\,dX_s = \int_0^t g(s)\,a(s)\,ds + \int_0^t g(s)\,b(s)\,dW_s \tag{2.35}$$

where g is a separable, progressively measurable, adapted process such that $\mathbb{P}\left(\int_0^t g^2(s)\,ds < \infty\right) = 1$. In shorthand notation the definition is

$$g(t)\,dX_t = g(t)\,a(t)\,dt + g(t)\,b(t)\,dW_t \tag{2.36}$$

There is an important formula for the change of variables in an Ito integral. Its simplified version is as follows. Let X_{1t} and X_{2t} are the Ito processes, written in shorthand notation

$$dX_{it} = a_i(t)\,dt + b_i(t)\,dW_t \quad \text{for} \quad i = 1,2 \tag{2.37}$$

then it can be shown that the process $X_{1t}X_{2t}$ is also an Ito process, written in shorthand notation

$$d(X_{1t}X_{2t}) = X_{1t}dX_{2t} + X_{2t}dX_{1t} + b_1(t)\,b_2(t)\,dt \tag{2.38}$$

Equation (2.38) is a special case of the following theorem, which is called Ito's formula.

Let

$$dX_t = a(t)\,dt + b(t)\,dW_t \tag{2.39}$$

$a(t)$ and $b(t)$ are stochastic processes such that the integrals $\int_0^t a(u)du$ and $\int_0^t b(u)dW_u$ exist, $\alpha < \beta$ real numbers, and let $f(x,t) : \left[\alpha,\beta\right] \times \left[0,T\right] \to \mathbb{R}$ be continuous together with the partial derivatives f_x, f_{xx}, and f_t. Then $f(X_t,t)$ is the Ito process

$$df(X_t,t) = \left[f_t(X_t,t) + f_x(X_t,t)\,a(t) + \frac{1}{2}f_{xx}(X_t,t)\,b^2(t)\right]dt + f_x(X_t,t)\,b(t)\,dW_t \tag{2.40}$$

Equation (2.40) is similar to the standard calculus formula for total differential, but has the additional term $\frac{1}{2}f_{xx}(X_t,t)\,b^2(t)\,dt$. Informally, the additional term is from second-order Taylor expansion of $f(X_t, t)$ and heuristically that $(dW_t)^2 = dt$

$$\begin{aligned} df(X_t,t) &= f_t(X_t,t)\,dt + f_x(X_t,t)\,dX_t + \frac{1}{2}f_{tt}(X_t,t)(dt)^2 \\ &+ \frac{1}{2}f_{xx}(X_t,t)(dX_t)^2 + f_{tx}(X_t,t)\,dt\,dX_t \\ &= f_t(X_t,t)\,dt + a(t)f_x(X_t,t)\,dt + b(t)f_x(X_t,t)\,dW_t + \frac{1}{2}b^2(t)f_{xx}(X_t,t)\,dt \\ &= \left[f_t(X_t,t) + a(t)f_x(X_t,t) + \frac{1}{2}b^2(t)f_{xx}(X_t,t)\right]dt + b(t)f_x(X_t,t)\,dW_t \end{aligned} \tag{2.41}$$

The first two lines of Equation (2.41) contain one term $(dX_t)^2$ that appears to be of second order but actually is of first order because the quadratic variation of the Wiener process is positive $(dW_t)^2 = dt$. The fourth line of Equation (2.41) omits the other terms of order greater than one.

For example, consider the Ito process

$$dX_t = \mu X_t dt + \sigma X_t dW_t \tag{2.42}$$

For intuitive understanding, we can rewrite it as $\frac{dX_t}{X_t} = \mu dt + \sigma dW_t$, which describes a combination of exponential growth at rate μ and scaled white noise with variance σ^2. Choose $f(x,t) = \log x$ and apply Ito's formula to get

$$\begin{aligned} d\log X_t &= \left(\mu\frac{1}{X_t}X_t - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 X_t^2\right)dt + \sigma\frac{1}{X_t}X_t dW_t \\ &= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t \end{aligned} \tag{2.43}$$

Integrating, we get the representation

$$\log X_t - \log X_0 = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t \tag{2.44}$$

and finally

$$X_t = X_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right] \tag{2.45}$$

For further details of the definitions and properties of Ito processes see, for example, Friedman [19] and Gihman and Skorohod [21].

2.2.5 Predictable Quadratic Covariation Process

The concept of predictable quadratic covariation process is readily defined in terms of the following theorem called the Doob-Meyer decomposition theorem:

Let $\mathbb{P}$ be a given probability measure, $\mathcal{F}_t$ be a given filtration, and let X_t be a non-negative $(\mathbb{P}, \mathcal{F}_t)$-sub-martingale.

Then there is a predictable, non-decreasing, right-continuous process A_t such that the process $X_t - A_t$ is a $(\mathbb{P}, \mathcal{F}_t)$-martingale. The process A_t is called a compensator of the process X_t. If we also require that $A_0 = 0$ then the process A_t is unique.

I will offer only a heuristic explanation of a predictable process. Heuristically, a predictable process A_t is adapted to the filtration $\mathcal{F}_{t-}$. For example, a deterministic process is predictable, and a left-continuous process is predictable. There are many predictable processes that are not deterministic or left-continuous.

The word "decomposition" in the name of the theorem comes from the fact that if we denote the martingale $X_t - A_t$ by M_t then we can write $X_t = A_t + M_t$, which is a decomposition of the sub-martingale X_t into a predictable process and a martingale.

For example, the Wiener process $X_t = W_t$ has a compensator $A_t = 0$ for all t because $W_t = 0 + W_t$ and W_t is a martingale.

As another example, the process $X_t = W_t^2$ has a compensator $A_t = t$. First, $X_t = W_t^2$ is a sub-martingale because of Jensen's inequality: Convex function of an average is less or equal than the average of a convex function, $\phi\left[\mathbb{E}(X)\right] \leq \mathbb{E}\left[\phi(X)\right]$, where X is an integrable random variable and ϕ is a convex function from $\mathbb{R}$ into $\mathbb{R}$. Square is a convex function, therefore, $\left[\mathbb{E}(X)\right]^2 \leq \mathbb{E}\left(X^2\right)$. Applying this inequality to $X = W_t$ and using conditional means given $\mathcal{F}_s$ we have

$$\mathbb{E}\left(W_t^2 \middle| \mathcal{F}_s\right) \quad \geq \quad \left[\mathbb{E}\left(W_t \middle| \mathcal{F}_s\right)\right]^2 \quad = \quad W_s^2 \tag{2.46}$$

We got that W_t^2 is a sub-martingale, and of course, it is adapted and non-negative. Now we can apply the Doob-Meyer decomposition theorem, which tells us that there is a unique predictable, non-decreasing, right-continuous process A_t such that $A_0 = 0$ and $W_t^2 - A_t$ is a martingale. We want to show that $A_t = t$.

It is easy to see that $W_t^2 - t$ is a martingale, and then the claim will follow from the uniqueness of A_t such that $A_0 = 0$. First, for $0 \leq s \leq t$, we have $W_t^2 = \Big[(W_t - W_s) + W_s\Big]^2 = (W_t - W_s)^2 + 2(W_t - W_s)W_s + W_s^2$.

Therefore

$$\begin{aligned}
\mathbb{E}\left(W_t^2 \middle| \mathcal{F}_s\right) &= \mathbb{E}\left[(W_t - W_s)^2 \middle| \mathcal{F}_s\right] + 2\mathbb{E}\left[(W_t - W_s)\, W_s \middle| \mathcal{F}_s\right] + \mathbb{E}\left(W_s^2 \middle| \mathcal{F}_s\right) \\
&= \mathbb{E}\left[(W_t - W_s)^2\right] + 2\mathbb{E}\,(W_t - W_s)\, W_s + W_s^2 \qquad (2.47) \\
&= t - s + 0 + W_s^2
\end{aligned}$$

In deriving Equation (2.47) we used the fact that the increment $W_t - W_s$ is independent of the history of the Wiener process up to time s, and that the variance of the increment $W_t - W_s$ is $t - s$.

We got that $\mathbb{E}\left(W_t^2 \middle| \mathcal{F}_s\right) = W_s^2 + t - s$, which is the same as $\mathbb{E}\left(W_t^2 - t \middle| \mathcal{F}_s\right) = W_s^2 - s$.

The preceding example involved the calculation of a compensator of a sub-martingale which is a square of the martingale W_t. We can generalize this to the square of any martingale M_t because $X_t = M_t^2$ is a non-negative sub-martingale. The proof is the same as the proof for $X_t = W_t$ using Jensen's inequality

$$\mathbb{E}\left(M_t^2 \middle| \mathcal{F}_s\right) \;\geq\; \left[\mathbb{E}\left(M_t \middle| \mathcal{F}_s\right)\right]^2 \;=\; M_s^2 \qquad (2.48)$$

The Doob-Meyer decomposition theorem tells us that M_t^2 has a unique compensator A_t such that $A_0 = 0$. This unique compensator is called the predictable quadratic variation process of the martingale M_t and is denoted by $\langle M, M\rangle_t$. From the Doob-Meyer theorem we know that $\langle M, M\rangle_t$ is a predictable, non-decreasing, right-continuous process and that $M_t^2 - \langle M, M\rangle_t$ is a martingale.

The fact that $W_t^2 - t$ is a martingale tells us that $\langle W, W\rangle_t = t$.

Next, consider the Ito integral $M_t = \int_0^t g(u)dW_u$. We know that M_t is a martingale, and therefore, $\langle M, M\rangle_t$ exists. Using a method similar to the calculation of $\langle W, W\rangle_t$, we can show that

$$\langle M, M\rangle_t = \int_0^t g^2(u)du \tag{2.49}$$

Indeed

$$\begin{aligned}
\mathbb{E}\left(M_t^2 \middle| \mathcal{F}_s\right) &= \mathbb{E}\left[(M_t - M_s)^2 \middle| \mathcal{F}_s\right] + 2\mathbb{E}\left[(M_t - M_s)M_s \middle| \mathcal{F}_s\right] + \mathbb{E}\left(M_s^2 \middle| \mathcal{F}_s\right) \\
&= \mathbb{E}\left[(M_t - M_s)^2\right] + 2M_s\mathbb{E}\left(M_t - M_s \middle| \mathcal{F}_s\right) + M_s^2 \\
&= \int_s^t g^2(u)du + 0 + M_s^2
\end{aligned} \tag{2.50}$$

We get that $M_t^2 - \int_0^t g^2(u)du$ is a martingale, and therefore, $\langle M, M\rangle_t = \int_0^t g^2(u)du$. We can now take one more step and define the predictable quadratic covariation process of two martingales M_t and N_t as

$$\langle M, N\rangle_t = \frac{1}{2}\Big(\langle M+N, M+N\rangle_t - \langle M, M\rangle_t - \langle N, N\rangle_t\Big) \tag{2.51}$$

With this definition, we can do a little algebra to show that the process $M_tN_t - \langle M, N\rangle_t$ is a martingale.

In the special case when $M_t = \int_0^t g(u)dW_u$ and $N_t = \int_0^t h(u)dW_u$ we have

$$\langle M, N\rangle_t = \int_0^t g(u)h(u)du \tag{2.52}$$

so that $M_t N_t - \int_0^t g(u)h(u)du$ is a martingale.

3 A Random Time with Learning

3.1 Three Revealing Examples

3.1.1 First Example

Consider the following three examples of a stochastic differential equation and its solution. The first example is

$$dX_t = aX_t dt + bX_t dW_t \tag{3.1}$$

where a is a constant and b, X_0 are positive constants. To solve this stochastic differential equation, apply Ito's formula to the function $f(x) = \log x$ to get

$$d\log X_t = \left(a - \frac{1}{2}b^2\right) dt + b dW_t \tag{3.2}$$

It follows that $X_t = X_0 \exp\left[\left(a - \frac{1}{2}b^2\right) t + bW_t\right]$. The process X_t is called exponential or geometric Wiener process. Because the Wiener process W_t does not converge to a finite random variable when t goes to infinity, the exponential Wiener process X_t does not converge to a finite random variable either.

3.1.2 Second Example

The second example is the stochastic differential equation

$$dX_t = \kappa(\mu - X_t)\,dt + \sigma dW_t \tag{3.3}$$

where μ, X_0 are constants, and κ, σ are positive constants. The solution to this stochastic differential equation is the mean-reverting Ornstein-Uhlenbeck process. To obtain a solution, apply Ito's formula to the function $f(x,t) = A(t) + B(t)x$, where $A(t)$ and $B(t)$ are functions of time to be chosen later. We get the partial derivatives $f_t = A'(t) + B'(t)x$, $f_x = B(t)$, and $f_{xx} = 0$. It follows that

$$df(X_t, t) = \Big[A'(t) + B'(t)X_t + \kappa(\mu - X_t)B(t)\Big]dt + \sigma B(t)dW_t \tag{3.4}$$

Now choose the functions $A(t), B(t)$ such that $A'(t) + \kappa\mu B(t) = 0$ and $B'(t) - \kappa B(t) = 0$. Then the stochastic differential equation becomes $df(X_t, t) = \sigma B(t) dW_t$. Solving those two ordinary differential equations yields

$$\begin{aligned} B(t) &= \exp(\kappa t) \\ A(t) &= -\kappa\mu \int_0^t \exp(\kappa u)du \\ &= -\mu\Big[\exp(\kappa t) - 1\Big] \end{aligned} \tag{3.5}$$

Putting those calculations together, we get $f(X_t, t) - f(X_0, 0) = \sigma \int_0^t B(u) dW_u$.

$$-\mu\Big[\exp(\kappa t)-1\Big]+\exp(\kappa t)X_t - X_0 \quad = \quad \sigma\int_0^t \exp(\kappa u)dW_u \tag{3.6}$$

$$X_t \quad = \quad \exp(-\kappa t)X_0 + \mu\Big[1-\exp(-\kappa t)\Big] + \sigma\exp(-\kappa t)\int_0^t \exp(\kappa u)dW_u \tag{3.7}$$

From the properties of the Ito integral, X_t has normal distribution with mean and variance

$$\begin{aligned}
\mathbb{E}(X_t) &= \exp(-\kappa t)X_0 + \mu\Big[1-\exp(-\kappa t)\Big] \\
\text{var}(X_t) &= \sigma^2\exp(-2\kappa t)\int_0^t \exp(2\kappa u)du \\
&= \sigma^2\exp(-2\kappa t)\frac{\exp(2\kappa t)-1}{2\kappa} \\
&= \sigma^2\frac{1-\exp(-2\kappa t)}{2\kappa}
\end{aligned} \tag{3.8}$$

When t goes to infinity, the process X_t converges to a normally distributed random variable with mean μ and variance $\frac{\sigma^2}{2\kappa}$. It is important to recognize that the limiting random variable does not depend on the initial value X_0, and as I will show next, is not suitable for generating a random time with learning.

3.1.3 Third Example

The third example is the stochastic differential equation

$$dX_t \quad = \quad b(t)dW_t \tag{3.9}$$

where the positive function $b(t)$ is such that $\int_0^\infty b^2(t)dt < \infty$. Then the solution $X_t = X_0 + \int_0^t b(u)dW_u$ has normal distribution with mean X_0 and variance $\int_0^t b^2(u)du$. The process X_t converges to a limiting random variable

$$X = X_0 + \int_0^\infty b(u)dW_u \tag{3.10}$$

which has normal distribution with mean X_0 and variance $\int_0^\infty b^2(u)du$. Note that the distribution of X depends on the initial value X_0.

More generally, let $\mathcal{F}_t^W$ be the natural filtration of the Wiener process W_t. For any $t > 0$ we can write

$$X = X_t + \int_t^\infty b(u)dW_u \tag{3.11}$$

Because X_t is $\mathcal{F}_t^W$ - measurable and for $u > t$ the increment $W_u - W_t$ is independent of $\mathcal{F}_t^W$, the integral $\int_t^\infty b(u)dW_u$ is independent of $\mathcal{F}_t^W$, and the conditional mean of the integral given $\mathcal{F}_t^W$ is equal to its marginal mean $\mathbb{E}\left[\int_t^\infty b(u)dW_u\right] = 0$. Similarly, the conditional variance of X given $\mathcal{F}_t^W$ is the marginal variance of the integral $\int_t^\infty b(u)dW_u$. We get

$$\begin{aligned} \mathbb{E}\left(X\middle|\mathcal{F}_t^W\right) &= X_t \\ \operatorname{var}\left(X\middle|\mathcal{F}_t^W\right) &= \int_t^\infty b^2(u)du \end{aligned} \tag{3.12}$$

Equation (3.12) suggests the following model of a random time. Let F be a strictly increasing distribution function on $\mathbb{R}_+$, and denote by Φ the standard normal distribution function. Then, the function $\psi\colon \mathbb{R} \to \mathbb{R}_+$ defined by $\psi(z) = F^{-1}\Big[\Phi(z)\Big]$ is strictly increasing, and if Z has standard normal distribution then the random variable $\psi(Z)$ has distribution function F.

As a concrete illustration, consider two possible choices for the distribution function F of the random time τ: A Weibull distribution and a gamma distribution. Panel (A) of Figure 3.1 shows a Weibull density and a gamma density with parameters chosen to make the distributions similar in the sense of equal means and equal variances. The value of the shape parameter for the Weibull distribution is $k = 2.00$, and the value of the scale parameter is $\lambda = 10.00$ (rate parameter is 0.1). The value of the shape parameter for the gamma distribution is $k = 3.66$, and the value of the scale parameter is $\lambda = 2.42$ (rate parameter is about 0.4132). The common value of the mean of the two distributions is 8.86, and the common value of the variance is 24.60. Panel (B) of Figure 3.1 shows the resulting functions ψ for the two distributions. We see that the function ψ is increasing and convex for both distributions, and that the function ψ corresponding to the gamma distribution grows faster for large values of its argument than the function ψ corresponding to the Weibull distribution.

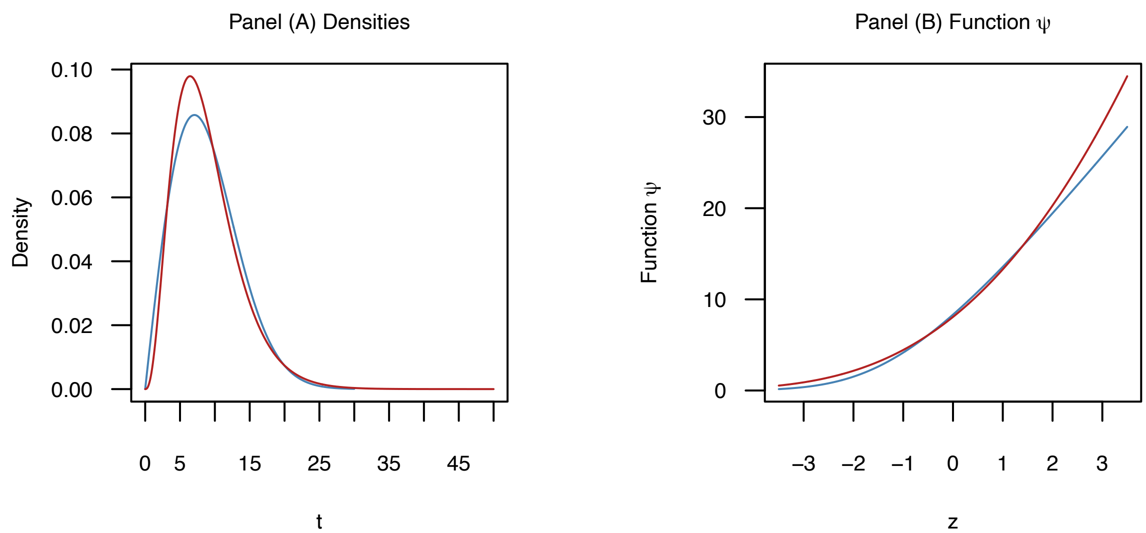

Figure 3.1. The function ψ for Weilbull and gamma distributions. The blue line corresponds to the Weibull distribution and the red line corresponds to the gamma distribution.

Assume, for simplicity, that $X_0 = 0$ and that $\int_0^\infty b^2(u)du = 1$ (by rescaling the function b). Define a random time $\tau = \psi(X)$. Noting that $\mathcal{F}_t^W = \mathcal{F}_t^X$, we have

$$\begin{aligned} \mathbb{P}\left(\tau \leq y \middle| \mathcal{F}_t^X\right) &= \mathbb{P}\left[\psi(X) \leq y \middle| \mathcal{F}_t^X\right] \\ &= \mathbb{P}\left[X \leq \psi^{-1}(y) \middle| \mathcal{F}_t^X\right] \\ &= \Phi\left[\frac{\psi^{-1}(y) - X_t}{\sigma_t}\right] \end{aligned} \tag{3.13}$$

where $\sigma_t^2 = \int_t^\infty b^2(u)du$. Equation (3.13) allows us to calculate the conditional distribution of the random time τ given the history of the process X_t. In addition, given the conditional distribution, we can calculate the conditional survival function, the conditional hazard function, the conditional mean residual life, and the conditional mean tail life of the random time τ.

It is important to recognize that the first example in Equation (3.1) does not allow such construction of a random time because the process X_t there does not have a limiting random variable X. Similarly, the second example in Equation (3.3) does not allow a construction of a random time because the distribution of the limiting random variable X does not depend on the initial value X_0. I will discuss below how we can modify the first and second example to make the construction of a random time feasible.

To continue the concrete illustration of this example, choose $b(t) = \sqrt{\delta}\exp\left(-\frac{\delta}{2}t\right)$. Then, $\int_0^\infty b^2(t)dt = 1$, and $\sigma_t^2 = \mathrm{var}\left(X \middle| \mathcal{F}_t^X\right) = \int_t^\infty b^2(u)du = \exp(-\delta t)$.

In the context of this illustration, the conditional survival function at time t, given $\mathcal{F}_t^X$ is

$$
\begin{aligned}
S\left(y \middle| \mathcal{F}_t{}^W\right) &= \mathbb{P}\left(\tau > y \middle| \mathcal{F}_t{}^X\right) \\
&= \Phi\left[\frac{X_t - \psi^{-1}(y)}{\sigma_t}\right]
\end{aligned}
\tag{3.14}
$$

where $\sigma_t = \exp\left(-\frac{\delta}{2}t\right)$.

Figure 3.2 shows the effect of randomly arriving information in the current illustrative example. Panel (A) displays conditional survival functions given unfavorable information $X_t = -0.8$ for some time $t > 0$ and a range of values of time $0 \leq y \leq 40$ years. The solid gray line corresponds to unconditional Weibull distribution of the random time, and the dashed gray line corresponds to unconditional gamma distribution of the random time. The solid blue line corresponds to conditional survival function when the unconditional distribution is Weibull, and the dashed blue line corresponds to conditional survival function when the unconditional distribution is gamma. Panel (B) displays conditional survival functions given favorable information $X_t = 0.8$ for some time $t > 0$ and a range of values of time $0 \le y \le 40$ years. As in Panel (A), the solid gray line corresponds to unconditional Weibull distribution of the random time, and the dashed gray line corresponds to unconditional gamma distribution of the random time. The solid red line corresponds to conditional survival function when the unconditional distribution is gamma, and the dashed red line corresponds to conditional survival function when the unconditional distribution is gamma.

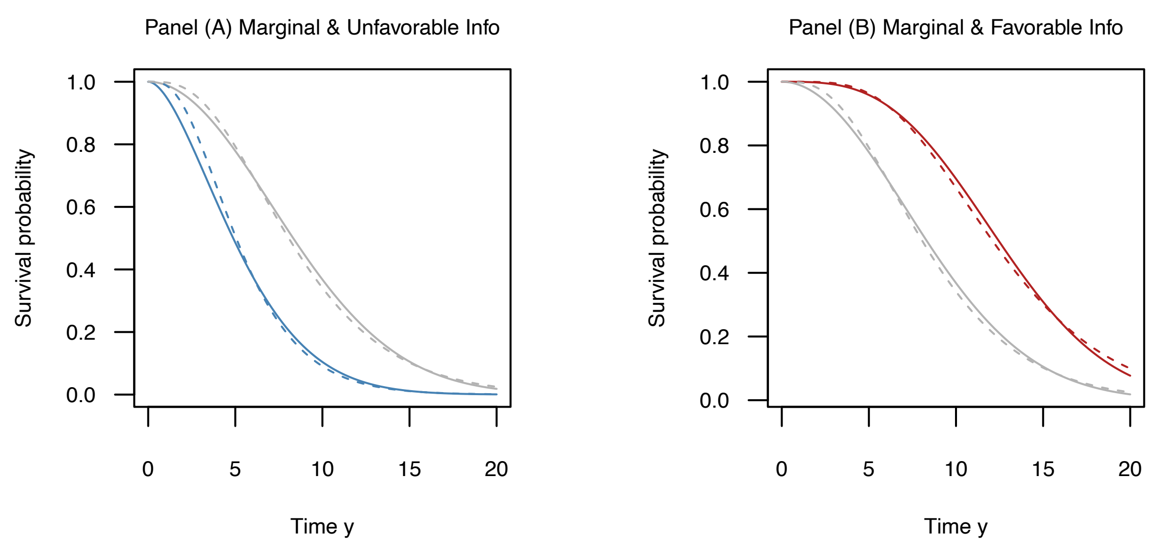

Figure 3.2. Marginal and conditional survival functions for Weibull and gamma. Gray lines refer to marginal survival, blue lines are conditional on unfavorable information $X_t = -0.8$, and red lines are conditional on favorable information $X_t = 0.8$. Solid lines refer to Weibull distribution, dashed lines refer to gamma distribution. The value of the parameter in the integrand function $b(t)$ is $\delta = 0.2$.

Figure 3.3 shows random arrival of information and survival probabilities for a fixed future time horizon $y = 10$ years. Panel (A) shows two sample paths of the health marker X_t. The blue line corresponds to a health marker trending mostly up, signifying largely favorable information, and the red line corresponds to a health marker trending down starting at time $t = 7$, signifying increasing unfavorable information. Panel (B) shows the corresponding survival probabilities past a fixed time horizon of 10 years. The blue line shows this probability trending up because of the favorable information in the health marker. The red line shows this probability beginning to fall at time $t = 7$ because the down-trending health marker. The lines correspond to conditional probabilities when the unconditional distribution is Weibull. The evolution of survival probabilities is similar when the unconditional distribution of the random time is a matched gamma.

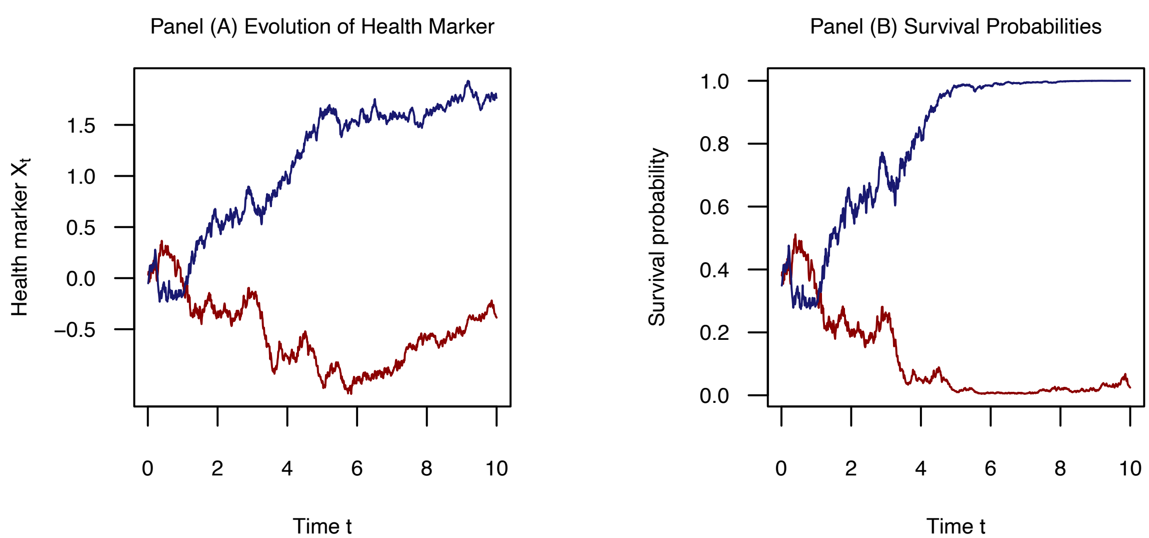

Figure 3.3. Random arrival of information and survival probabilities for a fixed future time horizon $y = 10$ years. Panel (A) shows two types of randomly arriving information, and Panel (B) shows the corrsponding survival probabilities for a fixed future time horizon $y = 10$. The lines correspond to conditional probabilities when the unconditional distribution is Weibull. The value of the parameter in the integrand function $b(t)$ is $\delta = 0.2$.

3.2 Linear Stochastic Differential Equations for Construction of Random Time

In this section, I will characterize linear stochastic differential equations that allow the construction of a random time. Based on the three examples above, we are looking for equations whose solutions converge to a limiting random variable that depends on the initial value of the solution.

3.2.1 Solution of the Linear Stochastic Differential Equation

Consider the one-dimensional linear stochastic differential equation

$$dX_t = \Big[a_0(t) + a_1(t)X_t\Big]dt + \Big[b_0(t) + b_1(t)X_t\Big]dW_t \tag{3.15}$$

The solution procedure is to solve first the associated homogeneous stochastic differential equation

$$dY_t = a_1(t)Y_t dt + b_1(t)Y_t dW_t \tag{3.16}$$

and then apply the two-dimensional Ito's formula to the process $\frac{X_t}{Y_t}$. We get the solution of the homogeneous equation by applying the one-dimensional Ito's formula to the function $f(x) = \log(x)$

$$d\log(Y_t) = \Big[a_1(t) - \tfrac{1}{2}b_1^2(t)\Big]dt + b_1(t)dW_t \tag{3.17}$$

Integrating both sides we get

$$Y_t = Y_0 \exp\left\{ \int_0^t \left[a_1(u) - \tfrac{1}{2} b_1^2(u) \right] du + \int_0^t b_1(u) dW_u \right\} \tag{3.18}$$

Next, consider the two-dimensional Ito's formula for the Ito processes

$$\begin{aligned} dX_t &= a(X_t, t)\, dt + b_1(X_t, t)\, dW_{1t} + b_2(X_t, t)\, dW_{2t} \\ dY_t &= A(Y_t, t)\, dt + B_1(Y_t, t)\, dW_{1t} + B_2(Y_t, t)\, dW_{2t} \end{aligned} \tag{3.19}$$

For the function $f(x, y, t)$ Ito's formula is

$$\begin{aligned} df(X_t, Y_t, t) &= \frac{\partial f}{\partial x}(X_t, Y_t, t)\, dX_t + \frac{\partial f}{\partial y}(X_t, Y_t, t)\, dY_t + \frac{\partial f}{\partial t}(X_t, Y_t, t)\, dt \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_t, Y_t, t)(dX_t)^2 + \frac{\partial^2 f}{\partial x \partial y}(X_t, Y_t, t)\, dX_t dY_t \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(X_t, Y_t, t)(dY_t)^2 \end{aligned} \tag{3.20}$$

where

$$\begin{aligned} (dX_t)^2 &= \left[b_1^2(X_t, t) + b_2^2(X_t, t) \right] dt \\ dX_t dY_t &= \left[b_1(X_t, t) B_1(Y_t, t) + b_2(X_t, t) B_2(Y_t, t) \right] dt \\ (dY_t)^2 &= \left[B_1^2(Y_t, t) + B_2^2(Y_t, t) \right] dt \end{aligned} \tag{3.21}$$

For the processes in Equation (3.15) and Equation (3.16) we get

$$
\begin{aligned}
d\left(\frac{X_t}{Y_t}\right) &= \frac{dX_t}{Y_t} - \frac{X_t dY_t}{Y_t^2} - \frac{dX_t dY_t}{Y_t^2} + \frac{1}{2}\frac{2X_t(dY_t)^2}{Y_t^3} \\
&= \frac{\left[a_0(t) + a_1(t)X_t\right]dt + \left[b_0(t) + b_1(t)X_t\right]dW_t}{Y_t} \\
&- \frac{a_1(t)X_t Y_t dt + b_1(t)X_t Y_t dW_t}{Y_t^2} \qquad (3.22) \\
&- \frac{\left[b_0(t) + b_1(t)X_t\right]b_1(t)Y_t dt}{Y_t^2} + \frac{X_t b_1^2(t) Y_t^2 dt}{Y_t^3} \\
&= \frac{\left[a_0(t) - b_0(t)b_1(t)\right]dt + b_0(t)dW_t}{Y_t}
\end{aligned}
$$

Integrating both sides

$$
\frac{X_t}{Y_t} - \frac{X_0}{Y_0} = \int_0^t \frac{a_0(u) - b_0(u)b_1(u)}{Y_u}du + \int_0^t \frac{b_0(u)}{Y_u}dW_u \qquad (3.23)
$$

We can see that Equation (3.23) does not depend on the value of Y_0. Therefore, we can set $Y_0 = 1$ and get

$$
X_t = Y_t\left[X_0 + \int_0^t \frac{a_0(u) - b_0(u)b_1(u)}{Y_u}du + \int_0^t \frac{b_0(u)}{Y_u}dW_u\right] \qquad (3.24)
$$

3.2.2 First Example Revisited

Modify the stochastic differential equation in Equation (3.1) into the form

$$
dX_t = a(t)X_t dt + b(t)X_t dW_t \qquad (3.25)
$$

where $a(t)$ is a real function of time such that $\int_0^\infty |a(t)| dt < \infty$, and $b(t)$ is a real positive function of time such that $\int_0^\infty b^2(t) dt < \infty$. From Equation (3.24) we get the solution

$$X_t = X_0 \exp\left\{ \int_0^t \left[a(u) - \tfrac{1}{2} b^2(u) \right] du + \int_0^t b(u) dW_u \right\} \tag{3.26}$$

The process X_t in Equation (3.26) converges to a limiting random variable X. The distribution of X is normal and depends on the initial value of X_0.

3.2.3 Second Example Revisited

Modify the stochastic differential equation in Equation (3.3) into the form

$$dX_t = \kappa \, e^{-\alpha t} (\mu - X_t) \, dt + \sigma \, e^{-\beta t} \, dW_t \tag{3.27}$$

where $\kappa, \alpha, \beta > 0$. To solve this equation let $f(x,t) = a(t) + b(t)x$, then $f_t = a'(t) + b'(t)x$, $f_x = b(t)$, and $f_{xx} = 0$. Using Ito's formula

$$df(X_t, t) = \left[a'(t) + b'(t) X_t + \kappa \, e^{-\alpha t} (\mu - X_t) \, b(t) \right] dt + \sigma \, e^{-\beta t} \, b(t) dW_t \tag{3.28}$$

To make the term $a'(t) + b'(t)X_t + \kappa \, e^{-\alpha t} (\mu - X_t) \, b(t)$ equal to zero set $b'(t) = \kappa \, e^{-\alpha t} \, b(t)$ and $a'(t) = -\kappa \mu \, e^{-\alpha t} \, b(t)$.

The first differential equation can be rewritten $\dfrac{d \log b(t)}{dt} = \kappa \, e^{-\alpha t}$ so that $\log b(t) = -\dfrac{\kappa}{\alpha} e^{-\alpha t} + A$ and $b(t) = B exp\left[-\dfrac{\kappa}{\alpha} e^{-\alpha t} \right]$. We want $b(0) = 1$ so that
$b(t) = \exp\left[\dfrac{\kappa}{\alpha} (1 - e^{-\alpha t}) \right]$.

The derivative of $\exp(-h\mathrm{e}^{-\alpha t})$ is $\alpha h \exp(-h\mathrm{e}^{-\alpha t})\mathrm{e}^{-\alpha t}$. Therefore

$$\begin{aligned}\int_0^t \alpha h \exp(-h\mathrm{e}^{-\alpha u})\mathrm{e}^{-\alpha u}\,du &= \int_0^t d\exp(-h\mathrm{e}^{-\alpha u}) \\ &= \exp(-h\mathrm{e}^{-\alpha t}) - \exp\left(-h\mathrm{e}^{0}\right) \\ &= \exp(-h\mathrm{e}^{-\alpha t}) - \exp(-h)\end{aligned} \tag{3.29}$$

We get that

$$\begin{aligned}a(t) &= -\kappa\mu\exp\left(\frac{\kappa}{\alpha}\right)\int_0^t \exp\left(-\frac{\kappa}{\alpha}\mathrm{e}^{-\alpha u}\right)\mathrm{e}^{-\alpha u}\,du \\ &= -\mu\exp\left(\frac{\kappa}{\alpha}\right)\left[\exp\left(-\frac{\kappa}{\alpha}\mathrm{e}^{-\alpha t}\right) - \exp\left(-\frac{\kappa}{\alpha}\right)\right] \\ &= \mu\left\{1 - \exp\left[\frac{k}{\alpha}(1-\mathrm{e}^{-\alpha t})\right]\right\} \\ &= \mu\left[1 - b(t)\right]\end{aligned} \tag{3.30}$$

To do a check for correctness, when α goes to zero we get

$$\begin{aligned}a(t) &= \mu\left[1 - \exp(\kappa t)\right] \\ b(t) &= \exp(\kappa t)\end{aligned} \tag{3.31}$$

which is the result when solving the Ornstein-Uhlenbeck stochastic differential equation.

The solution of our stochastic differential equation is

$$
\begin{aligned}
a(t)+b(t)X_t-X_0 &= \sigma\int_0^t e^{-\beta u}\, b(u)\,dW_u \\
X_t &= \frac{1}{b(t)}\left[X_0-a(t)+\sigma\int_0^t e^{-\beta u}\, b(u)\,dW_u\right] \qquad (3.32)\\
&= \frac{X_0}{b(t)}+\mu\left[1-\frac{1}{b(t)}\right]+\frac{\sigma}{b(t)}\int_0^t e^{-\beta u}\, b(u)\,dW_u
\end{aligned}
$$

and $\frac{1}{b(t)}=\exp\left[-\frac{\kappa}{\alpha}(1-e^{-\alpha t})\right]$, so that

$$
\begin{aligned}
X_t &= X_0\exp\left[-\frac{\kappa}{\alpha}(1-e^{-\alpha t})\right]+\mu\left\{1-\exp\left[-\frac{\kappa}{\alpha}(1-e^{-\alpha t})\right]\right\} \qquad (3.33)\\
&+ \sigma\exp\left[-\frac{\kappa}{\alpha}(1-e^{-\alpha t})\right]\int_0^t e^{-\beta u}\, b(u)\,dW_u
\end{aligned}
$$

When t goes to infinity in Equation (3.33), the stochastic process X_t converges to a limiting random variable whose distribution depends on the initial value X_0. Therefore, the solution of the modified Ornstein-Uhlenbeck process is suitable for generating a random time with learning.

I defer using the modified Ornstein-Uhlenbeck process for generating a random time to future research, and look now only at the behavior of the integral term

$$
J_t = \sigma\exp\left[-\frac{\kappa}{\alpha}(1-e^{-\alpha t})\right]\int_0^t e^{-\beta u}\exp\left[\frac{\kappa}{\alpha}(1-e^{-\alpha u})\right]dW_u \qquad (3.34)
$$

Perhaps not surprisingly, the behavior depends again on the magnitude of the parameter β.

3.3 A Random Time with Learning

3.3.1 Choice of a Weight Function

In the preceding modeling options, the positive integrand $b(t)$ in the Ito integral $\int_0^t b(u)dW_u$ can be called a weight function. Here I offer two additional options for a weight function and plot their shapes and the shapes of the conditional variance functions.

The first weight function is based on the exponential density.

$$\begin{aligned} f(t) &= \lambda \exp(-\lambda t) \\ b(t) &= \left[f(t)\right]^{\frac{1}{2}} \\ S(t) &= \exp(-\lambda t) \\ \text{var}\left(X\middle|\mathcal{F}_t^X\right) &= S(t) \end{aligned} \tag{3.35}$$

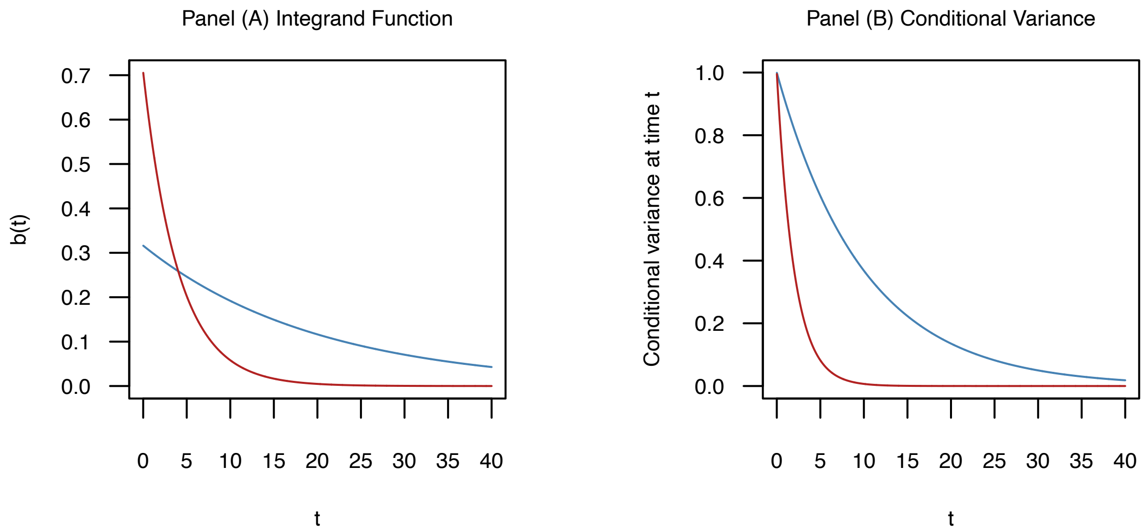

Figure 3.4. Integrand function and conditional variance based on exponential density. The blue lines are calculated with rate parameter $\lambda = 0.1$. The red lines are calculated with rate parameter $\lambda = 0.5$.

The second weight function is based on the lognormal density.

$$
\begin{aligned}
f(t) &= \frac{1}{\sqrt{2\pi\sigma^2}\,t} \exp\left[\frac{(\log t - \mu)^2}{2\sigma^2}\right] \\
b(t) &= \left[f(t)\right]^{\frac{1}{2}} \\
S(t) &= \Phi\left(\frac{\mu - \log t}{\sigma}\right) \\
\operatorname{var}\left(X \middle| \mathcal{F}_t^X\right) &= S(t)
\end{aligned}
\tag{3.36}
$$

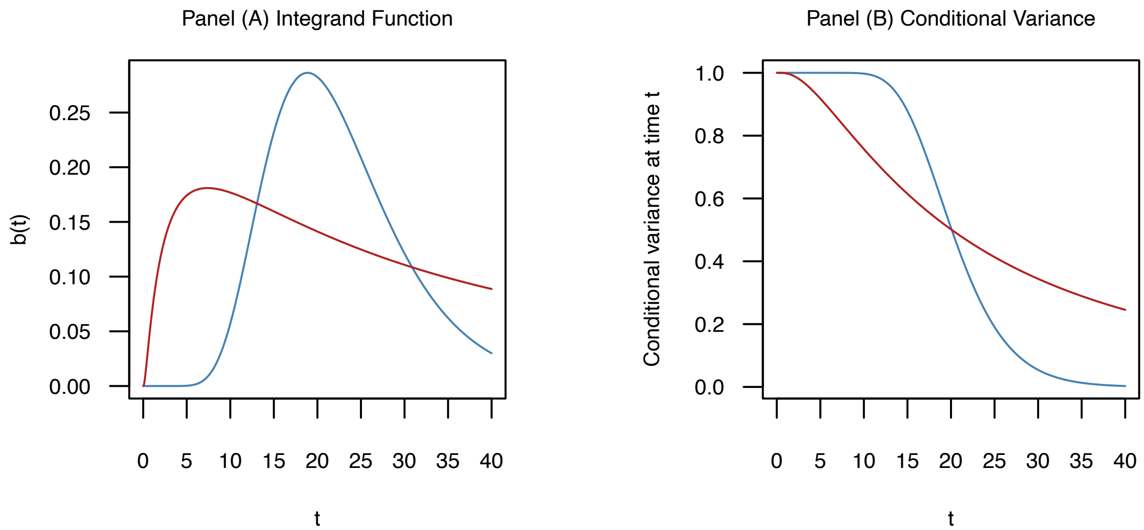

Figure 3.5. Integrand function and conditional variance based on log-normal density. The blue lines are calculated with location parameter $\mu = 3$ and scale parameter $\sigma = 0.25$. The red lines are calculated with location parameter $\mu = 3$ and scale parameter $\sigma = 1$.

3.3.2 Conditional Distribution Function

I am going to refer to two time variables, t and y, the time variable t denotes the time of observation of the health marker X_t, and the time variable y serves as an argument of the conditional distribution function, conditional density function, conditional survival function, *etc.* Both t and y are measured from a common origin that can be the time the patient was born, the time the patient was put under observation for the first time, or another time that is suitable for the modeling requirements. There is a flexibility in choosing the time origin, depending on the specific application of the model. Given the interpretation of the time variables t and y, we must have $y \geq t$.

For $y < t$, we have $\mathbb{P}\left(\tau \leq y \middle| \mathcal{F}_t^X, \tau > t\right) = 0$, and for $y \geq t$, we have

$$\begin{aligned} F\left(y \middle| \mathcal{F}_t^X, \tau > t\right) &= \mathbb{P}\left(\tau \leq y \middle| \mathcal{F}_t^X, \tau > t\right) \\ &= \frac{\mathbb{P}\left(t < \tau \leq y \middle| \mathcal{F}_t^X\right)}{\mathbb{P}\left(\tau > t \middle| \mathcal{F}_t^X\right)} \\ &= \frac{\Phi\left[\frac{\psi^{-1}(y) - X_t}{\sigma_t}\right] - \Phi\left[\frac{\psi^{-1}(t) - X_t}{\sigma_t}\right]}{1 - \Phi\left[\frac{\psi^{-1}(t) - X_t}{\sigma_t}\right]} \end{aligned} \tag{3.37}$$

where $\sigma_t = \exp\left(-\frac{\delta}{2}t\right)$.

To calculate the conditional density function observe that $\psi^{-1}(y) = \Phi^{-1}\left[F(y)\right]$. Therefore

$$\begin{aligned} \Phi\left[\psi^{-1}(y)\right] &= \Phi\left\{\Phi^{-1}\left[F(y)\right]\right\} \\ &= F(y) \end{aligned} \tag{3.38}$$

and $\phi\left[\psi^{-1}(y)\right]\frac{d\psi^{-1}(y)}{dy} = f(y)$ where f is the density of F.

It follows that

$$\frac{d\psi^{-1}(y)}{dy} = \frac{f(y)}{\phi\left[\psi^{-1}(y)\right]} \tag{3.39}$$

and for $y \geq t$

$$f\left(y \middle| \mathcal{F}_t^X, \tau > t\right) = \frac{\phi\left[\frac{\psi^{-1}(y) - X_t}{\sigma_t}\right]\frac{f(y)}{\sigma_t \phi\left[\psi^{-1}(y)\right]}}{1 - \Phi\left[\frac{\psi^{-1}(t) - X_t}{\sigma_t}\right]} \tag{3.40}$$

where $\sigma_t = \exp\left(-\frac{\delta}{2}t\right)$. Note that when $t = 0$ and $X_0 = 0$ then $f\left(y\middle|\mathcal{F}_0^X, \tau > 0\right) = f(y)$.

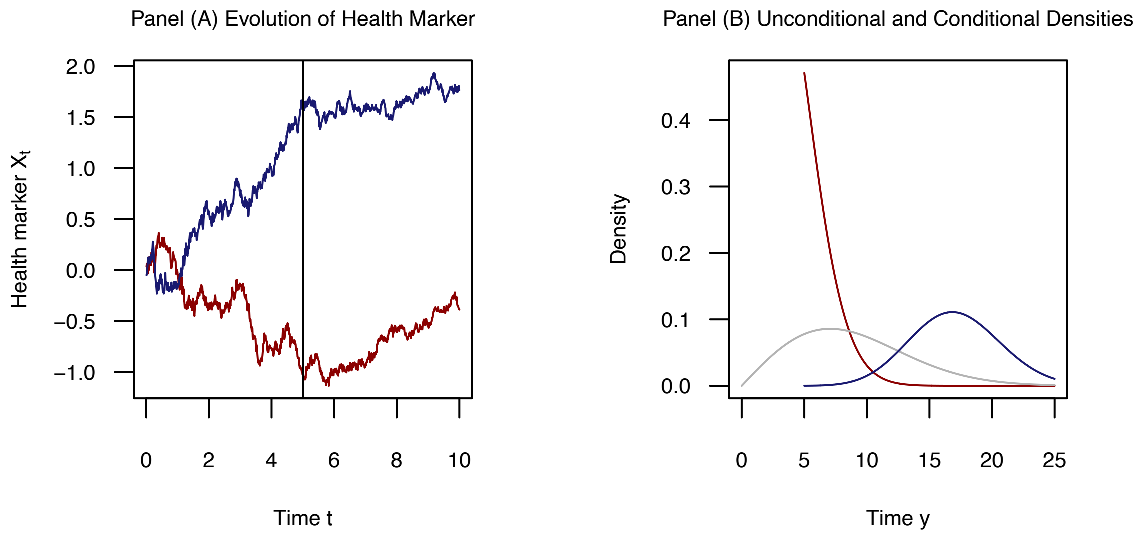

Figure 3.6. Random arrival of information, unconditional survival density, and conditional densities of survival time given information at time $t = 5$ and given that the patient is alive at time $t = 5$. Panel (A) shows two types of randomly arriving information. In Panel (B), the unconditional density is Weibull with shape parameters 2 and scale parameter 10, and the conditional densities are based on the unconditional Weibull density. The blue line corresponds to $X_5 = 0.889$, the red line corresponds to $X_5 = 0.240$. The value of the parameter in the integrand function $b(t)$ is $\delta = 0.2$.

3.3.3 Conditional Survival Function

For $y < t$, we have $\mathbb{P}\left(\tau > y\middle|\mathcal{F}_t^X, \tau > t\right) = 1$, and for $y \geq t$, we have

$$S\left(y\middle|\mathcal{F}_t^X, \tau > t\right) = \mathbb{P}\left(\tau > y\middle|\mathcal{F}_t^X, \tau > t\right) = \frac{1 - \Phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right]}{1 - \Phi\left[\dfrac{\psi^{-1}(t) - X_t}{\sigma_t}\right]} \tag{3.41}$$

where $\sigma_t = \exp\left(-\frac{\delta}{2} t\right)$.

3.3.4 Conditional Hazard Function

Using Equation (3.40) for the conditional density of the random time and Equation (3.41) for conditional survival function, we get for $y \geq t$

$$
\begin{aligned}
h\left(y \middle| \mathcal{F}_t^X, \tau > t\right) &= \frac{f\left(y \middle| \mathcal{F}_t^X, \tau > t\right)}{S\left(y \middle| \mathcal{F}_t^X, \tau > t\right)} \\
&= \frac{\dfrac{\phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right] \dfrac{f(y)}{\sigma_t \phi\left[\psi^{-1}(y)\right]}}{1 - \Phi\left[\dfrac{\psi^{-1}(t) - X_t}{\sigma_t}\right]}}{\dfrac{1 - \Phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right]}{1 - \Phi\left[\dfrac{\psi^{-1}(t) - X_t}{\sigma_t}\right]}} \\
&= \frac{\phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right] \dfrac{f(y)}{\sigma_t \phi\left[\psi^{-1}(y)\right]}}{1 - \Phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right]}
\end{aligned}
\tag{3.42}
$$

It is important to keep in mind that the time variables t and y are measured from the same origin.

Figures 3.7 through 3.9 illustrate the behavior and flexibility of the conditional hazard functions in this model.

Figure 3.7 illustrates the behavior of unconditional and conditional hazard functions when the unconditional distribution of the random time is Weibull with shape parameter $k = 2$ and scale parameter $\lambda = 10$. Panel (A) of Figure 3.7 shows the random evolution of the health marker X_t in two patients. Focusing on observation time $t = 2$, we get for patient 1 (blue line) the value $X_2 = 0.576$ and for patient 2 (red line) the value $X_2 = -0.342$. Panel (B) of Figure 3.7 shows the unconditional hazard function (gray line), the conditional hazard function for patient 1 (blue line) and the conditional hazard function for patient 2 (red line), all as functions of the time variable y. As is the case for shape parameter greater than 1, the unconditional hazard function is increasing. Because of the favorable value of the health marker for patient 1, the hazard line for patient 1 starts below the unconditional hazard line, but eventually rises above the unconditional hazard line reflecting the uncertainty about the future evolution of the health marker. The hazard line for patient 2 starts above the unconditional hazard line due to the unfavorable value of the health marker, and increases faster than the unconditional hazard line.

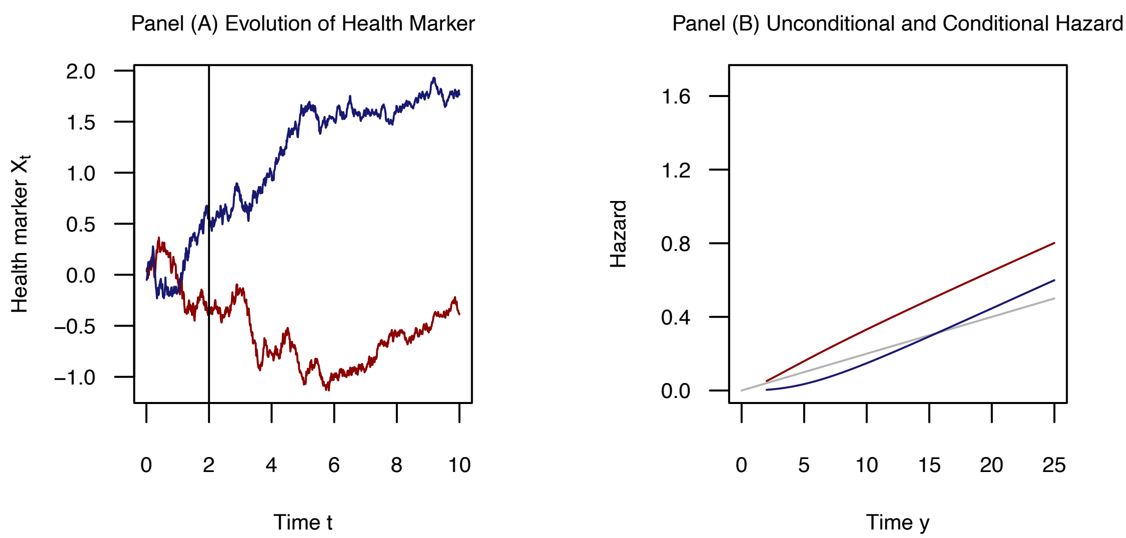

Figure 3.7. Random arrival of information, unconditional hazard function, and conditional hazard functions given information at time $t = 2$ and given that the patients are alive at time $t = 2$. The blue line in Panel (A) corresponds to $X_2 = 0.576$ and the red line corresponds to $X_2 = -0.342$. Panel (B) shows the unconditional and conditional hazards corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

Figure 3.8 resembles Figure 3.7 with a shift in observation time from $t = 2$ to $t = 5$. As Panel (A) shows, the shift corresponds to the value $X_5 = 1.592$ for patient 1 (blue line), and $X_5 = -1.028$ for patient 2 (red line). Panel (B) of Figure 3.8 is drawn with the same vertical scale as Figure 3.7 for convenient comparison. The unconditional hazard function (gray line) remains unchanged, the conditional hazard function for patient 1 (blue line) is more convex than in Panel (B) of Figure 3.7, reflecting both the increase in the value of the observed health marker and the shift in observation time from $t = 2$ to $t = 5$. The hazard function of patient 2 (red line) rises for all values of the time variable y, reflecting the more unfavorable value of the observed health marker.

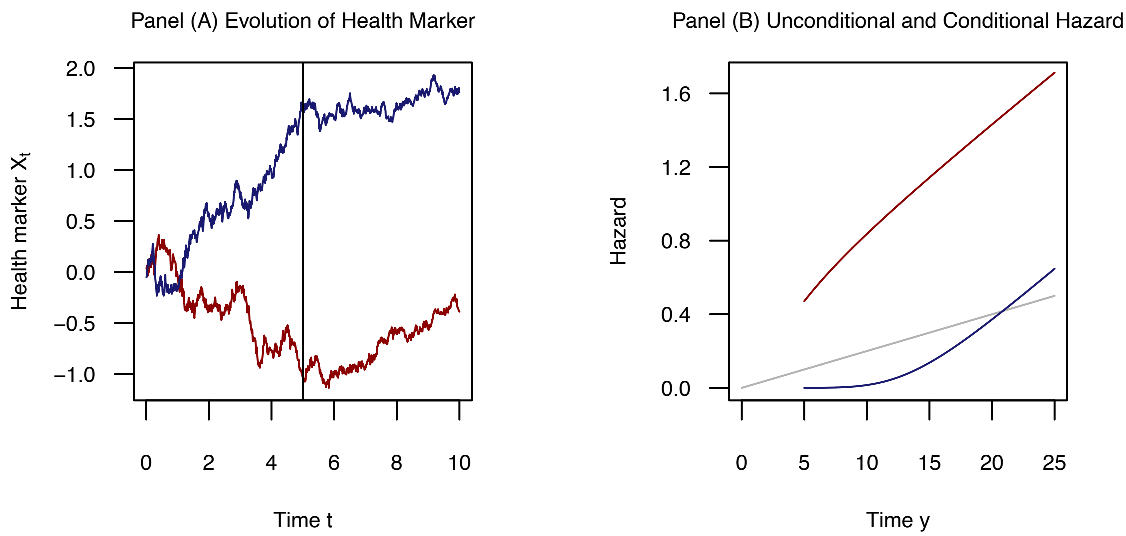

Figure 3.8. Random arrival of information, unconditional hazard function, and conditional hazard functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional hazards corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

Figure 3.9 shows again observation time $t = 5$, and differs from Figure 3.8 in the value of the shape parameter $k = 0.5$ of the unconditional Weibull distribution. The unconditional hazard function (gray line) in Panel (B) reflects the fact that the hazard function of the Weibull distribution is decreasing for $k < 1$. The conditional hazard function of patient 1 (blue line) is now almost flat and the conditional hazard function of patient 2 (red line) is decreasing, almost parallel to the unconditional hazard function, but with higher values because of the unfavorable value of the health marker for patient 2.

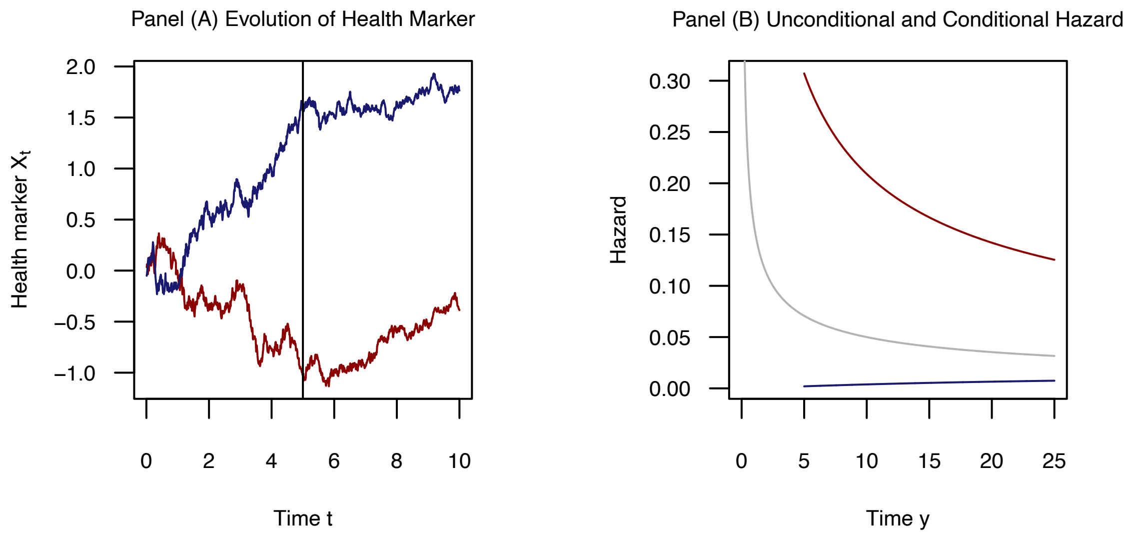

Figure 3.9. Random arrival of information, unconditional hazard function, and conditional hazard functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional hazards corresponding to an unconditional Weibull distribution with shape parameter $k = 0.5$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

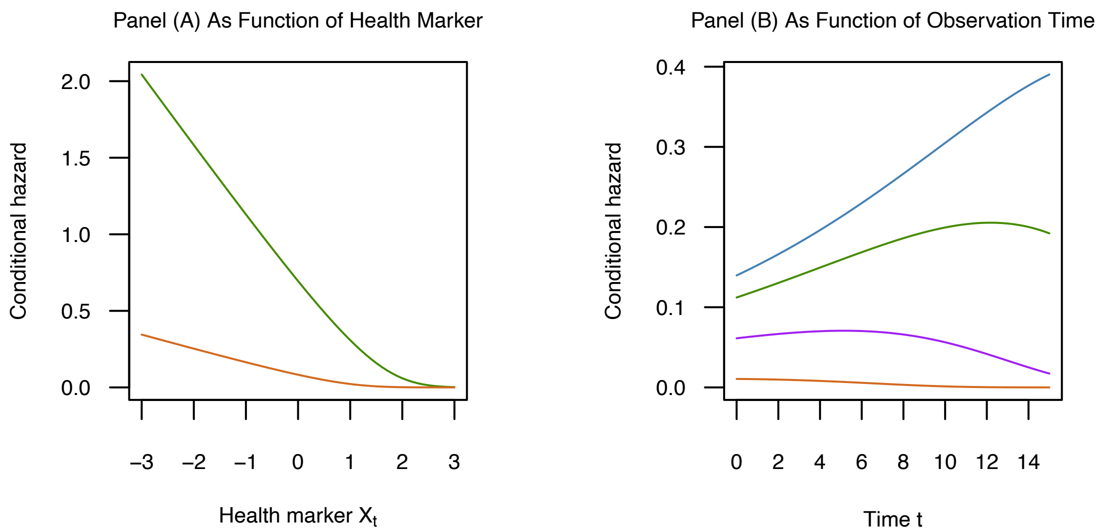

Figure 3.10. Conditional hazard as function of single argument with other arguments held constant. Panel (A) shows conditional hazard as a function of the health marker X_t for $k = 2$ (green line) and $k = 0.5$ (brown line) at $t = 5$. Panel (B) shows conditional hazard as a function of observation time t for $k = 2.2$ (blue line), $k = 2$ (green line), $k = 1.5$ (purple line) and $k = 0.5$ (brown line) at $X_t = 1.5$. The value of time y is constant at $y = 15$, the scale parameter of the unconditional Weibull distribution is constant at $\lambda = 10$, and the parameter of the weight function is constant at $\delta = 0.2$.

3.3.5 Conditional Mean Residual Life

Using Equation (2.8) for the integral representation of the mean residual life and Equation (3.41) for the conditional survival function, we get for $y \geq t$

$$m\left(y \middle| \mathcal{F}_t^X, \tau > t\right) = \frac{\int_y^\infty S\left(u \middle| \mathcal{F}_t^X, \tau > t\right) du}{S\left(y \middle| \mathcal{F}_t^X, \tau > t\right)} \tag{3.43}$$

where $S\left(y \middle| \mathcal{F}_t^X, \tau > t\right) = \dfrac{1 - \Phi\left[\dfrac{\psi^{-1}(y) - X_t}{\sigma_t}\right]}{1 - \Phi\left[\dfrac{\psi^{-1}(t) - X_t}{\sigma_t}\right]}$ and $\sigma_t = \exp\left(-\dfrac{\delta}{2} t\right)$. The integral in Equation (3.43) has to be evaluated numerically.

Figure 3.11 and Figure 3.12 illustrate the behavior of unconditional and conditional mean residual life for observation times $t = 2$ and $t = 5$, respectively, when the unconditional distribution is Weibull with shape parameter $k = 2$ and scale parameter $\lambda = 10$. In accordance with intuition, the conditional mean residual life for patient 1 (blue line) dominates the unconditional mean residual life (gray line), which in turn, dominates the mean residual life for patient 2 (red line). This is explained by the more favorable value of the health marker for patient 1 than for patient 2. When we shift the observation time from $t = 2$ to $t = 5$, all three mean residual life functions shift down, corresponding to the shift in observation time. Moreover, the mean residual life of patient 1 (blue line) declines less than the mean residual life of patient 2 (red line) because of the improvement in the value of the health marker of patient 1 and worsening in the value of the health marker of patient 2.

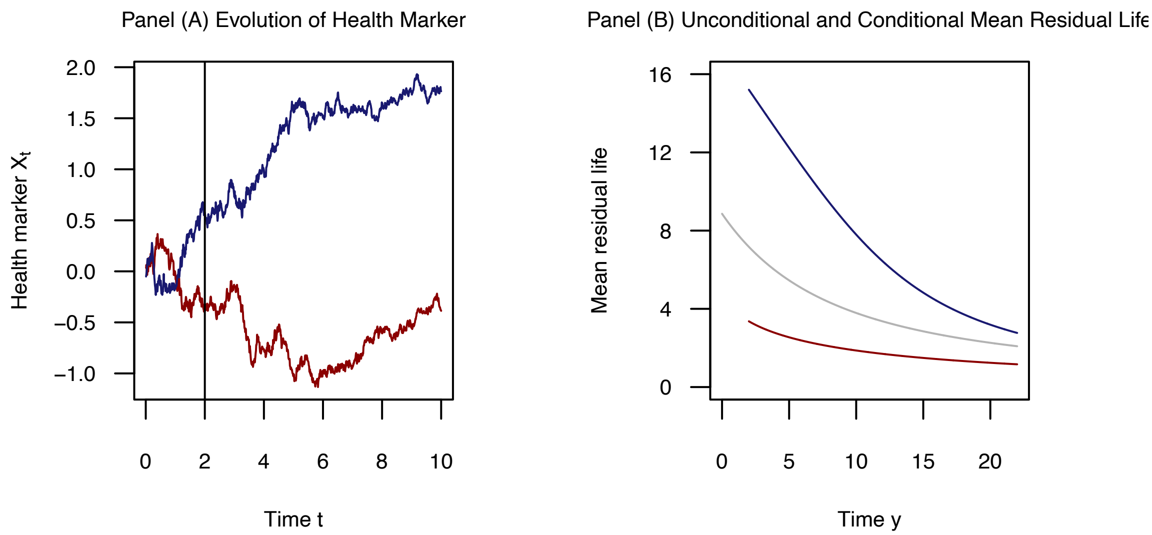

Figure 3.11. Random arrival of information, unconditional mean residual life function, and conditional mean residual life functions given information at time $t = 2$ and given that the patients are alive at time $t = 2$. The blue line in Panel (A) corresponds to $X_2 = 0.576$ and the red line corresponds to $X_2 = -0.342$. Panel (B) shows the unconditional and conditional mean residual life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

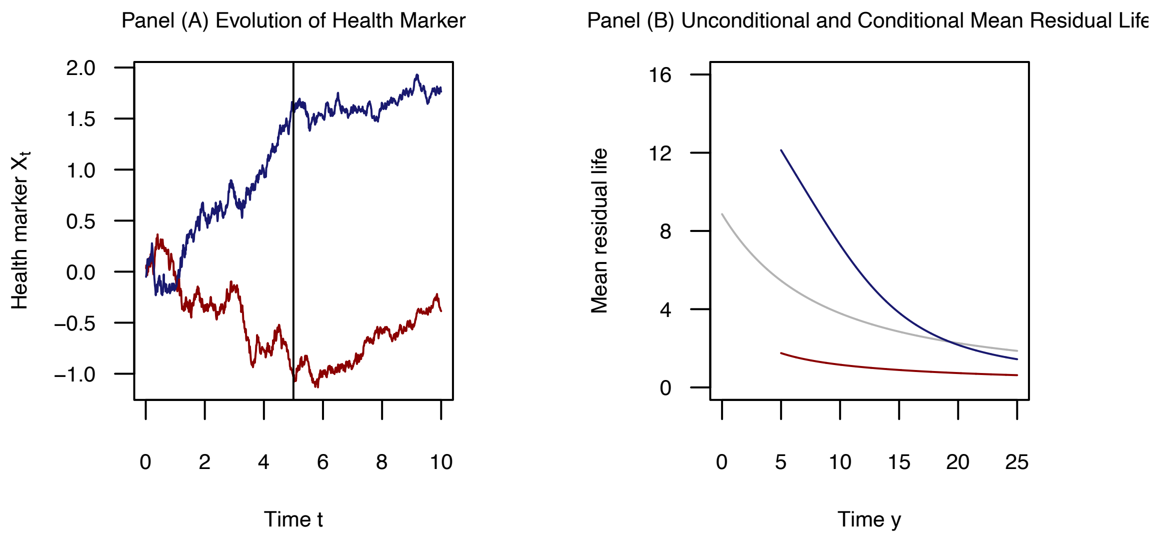

Figure 3.12. Random arrival of information, unconditional mean residual life function, and conditional mean residual life functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional mean residual life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

Figure 3.13 illustrates the behavior of mean residual life when the unconditional distribution of the random time is Weibull with shape parameter $k = 0.5$ and scale parameter $\lambda = 10$. Because the unconditional hazard function of the Weibull distribution with shape parameter $k < 1$ is decreasing, the unconditional mean residual life function (gray line) shifts up and becomes increasing. The conditional mean residual life for patient 1 (blue line) also shifts up, but remains decreasing, because of the uncertainty associated with the random evolution of the health marker. Finally, the mean residual life of patient 2 (red line) shifts up but remains almost flat.

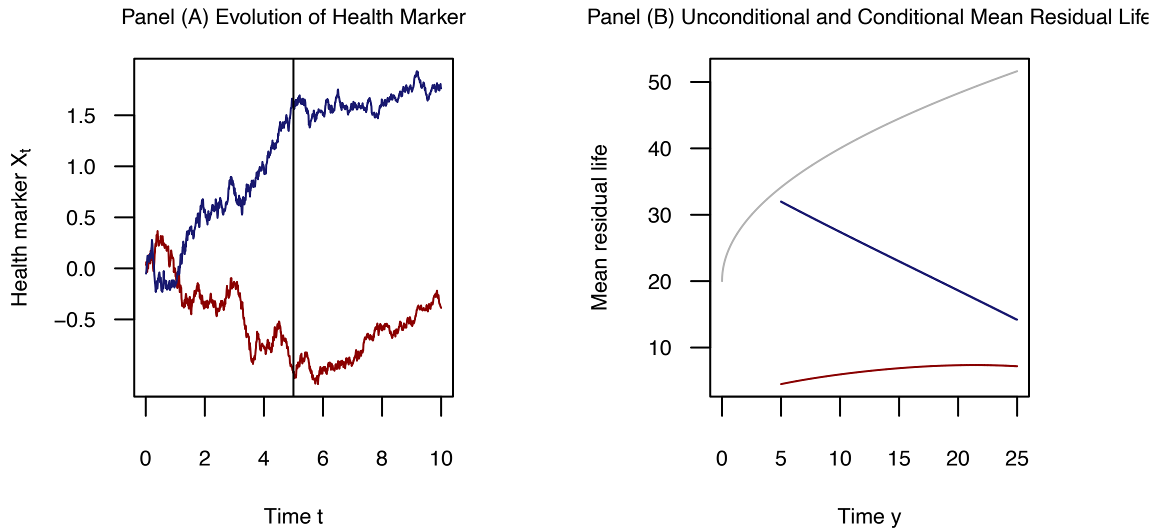

Figure 3.13. Random arrival of information, unconditional mean residual life function, and conditional mean residual life functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional mean residual life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 0.5$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

3.3.6 Conditional Mean Tail Life

For $0 < \alpha \leq 1$, we have

$$n\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) = \mathbb{E}\left[\tau\middle|\mathcal{F}_t^X, \tau > t, \tau > S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right)\right] \tag{3.44}$$

Time $y = S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right)$ is such that $S\left(y\middle|\mathcal{F}_t^X, \tau > t\right) = \alpha$. Because the conditional survival function $S\left(y\middle|\mathcal{F}_t^X, \tau > t\right)$ is defined only for $y \geq t$, we must have $S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) \geq t$.

Comparing the definitions of conditional mean residual life and conditional mean tail life implies that for $0 < \alpha \leq 1$

$$n\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) = m\left[S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right)\middle|\mathcal{F}_t^X, \tau > t\right] + S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) \tag{3.45}$$

In particular

$$n\left(1\middle|\mathcal{F}_t^X, \tau > t\right) = m\left(t\middle|\mathcal{F}_t^X, \tau > t\right) + t \tag{3.46}$$

Alternatively, by direct calculation, denote $y = S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right)$, then $S\left(y\middle|\mathcal{F}_t^X, \tau > t\right) = \alpha$, and substituting $\alpha = 1$, we get $y = t$ (We assumed that the distribution function F is strictly increasing).

Therefore, $n\left(1\middle|\mathcal{F}_t^X, \tau > t\right) = E\left(\tau\middle|\mathcal{F}_t^X, \tau > t\right)$, and recalling that $m\left(t\middle|\mathcal{F}_t^X, \tau > t\right) = E\left(\tau\middle|\mathcal{F}_t^X, \tau > t\right) - t$ and we get the desired result. This is reflected in Figure 3.11 and Figure 3.14.

Using Equation (2.13) we can compute mean tail life from

$$n\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) = \frac{\int_{S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right)}^{\infty} S\left(u\middle|\mathcal{F}_t^X, \tau > t\right) du}{\alpha} + S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) \tag{3.47}$$

Inverting Equation (3.41), we get

$$S^{-1}\left(\alpha\middle|\mathcal{F}_t^X, \tau > t\right) = \psi\left\{X_t + \sigma_t \Phi^{-1}\left[g(\alpha, X_t, \sigma_t)\right]\right\} \tag{3.48}$$

where $\sigma_t = \exp\left(-\frac{\delta}{2}t\right)$, and $g(\alpha, X_t, \sigma_t) = 1 - \alpha\left\{1 - \Phi\left[\frac{\psi^{-1}(t) - X_t}{\sigma_t}\right]\right\}$. The integral in Equation (3.44) has to be evaluated numerically.

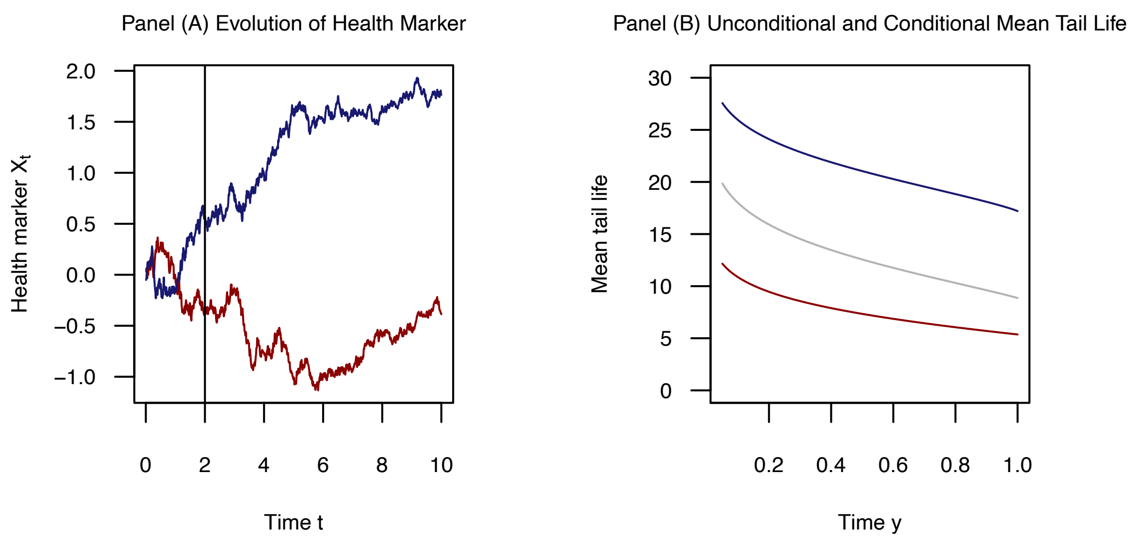

Figure 3.14. Random arrival of information, unconditional mean tail life function, and conditional mean tail life functions given information at time $t = 2$ and given that the patients are alive at time $t = 2$. The blue line in Panel (A) corresponds to $X_2 = 0.576$ and the red line corresponds to $X_2 = -0.342$. Panel (B) shows the unconditional and conditional mean tail life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

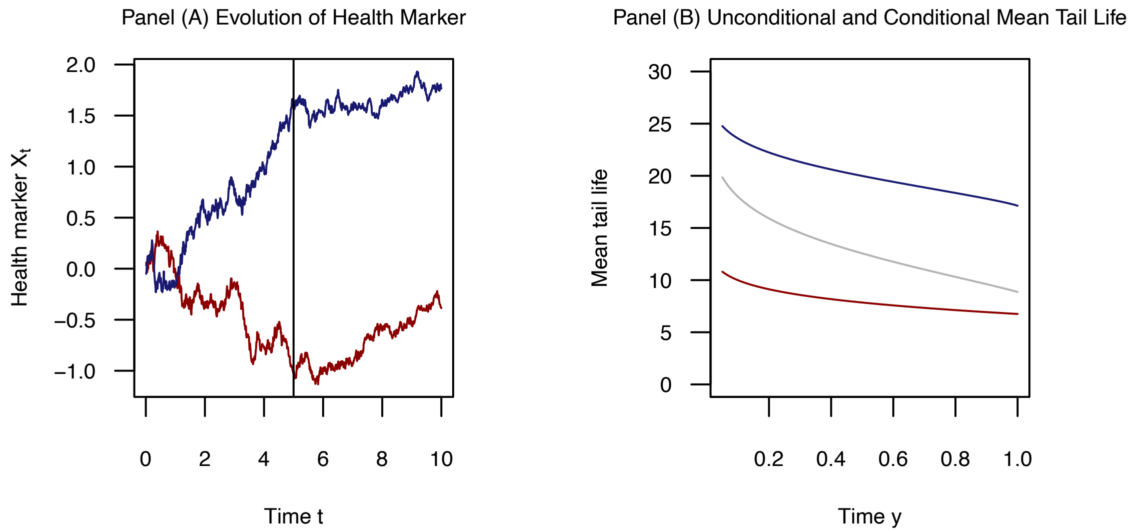

Figure 3.15. Random arrival of information, unconditional mean tail life function, and conditional mean tial life functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional mean tail life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 2$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

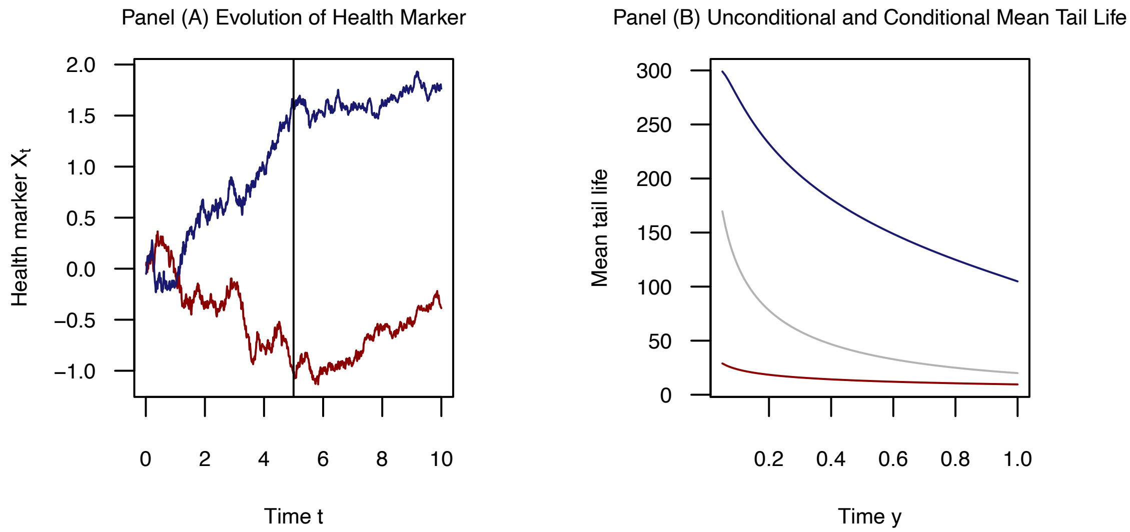

Figure 3.16. Random arrival of information, unconditional mean tail life function, and conditional mean tail life functions given information at time $t = 5$ and given that the patients are alive at time $t = 5$. The blue line in Panel (A) corresponds to $X_5 = 1.592$ and the red line corresponds to $X_5 = -1.028$. Panel (B) shows the unconditional and conditional mean tail life functions corresponding to an unconditional Weibull distribution with shape parameter $k = 0.5$ and scale parameter $\lambda = 10$. The value of the parameter in the weight function $b(t)$ is $\delta = 0.2$.

4 Simulation and Comparison with Cox Model

4.1 Simulation

The essential feature of the model introduced in Chapter 3 is the stochastic evolution of the health marker and the ability to compute conditional survival and hazard functions. To illustrate the differences between the new model and the Cox proportional hazards model it is helpful to have a health marker that has a non-zero drift and which converges to a finite random variable Y when time goes to infinity. This limiting random variable then defines the survival time of the model $\tau = \psi(Y)$.

Consider the following setup. A number n of subjects have been diagnosed with a medical condition and were instructed to isolate at home until onset of symptoms, at which time they will be hospitalized. To help estimate the arrival time of symptoms, the subjects are given a test with a single health marker, first at time 0, and then every 10 days until hospitalized. This health marker corresponds to a single covariate in the Cox proportional hazards model analysis of the data and represents conditioning information in the model introduced in Chapter 3.

The health marker $Y_t = \log X_t$ is generated by the Ito process

$$dX_t \quad = \quad \mathrm{e}^{-\delta t}(\mu X_t dt + \sigma X_t dW_t) \tag{4.1}$$

The solution of the stochastic differential equation in Equation (4.1) is obtained as follows.

Consider a function $f(x,t) = a(t) + \log x$, then

$$\begin{aligned} f_x &= \frac{1}{x} \\ f_{xx} &= -\frac{1}{x^2} \\ f_t &= a'(t) \end{aligned} \tag{4.2}$$

and, using Ito's formula

$$df(X_t,t) = \left(a' + \frac{1}{X_t}\mu e^{-\delta t} X_t - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 e^{-2\delta t} X_t^2\right)dt + \frac{1}{X_t}\sigma e^{-\delta t} X_t dW_t \tag{4.3}$$

Simplifying

$$df(X_t,t) = \left(a' + \mu e^{-\delta t} - \frac{1}{2}\sigma^2 e^{-2\delta t}\right)dt + \sigma e^{-\delta t} dW_t \tag{4.4}$$

We will solve for $a(t)$ such that

$$a' = \frac{1}{2}\sigma^2 e^{-2\delta t} - \mu e^{-\delta t} \tag{4.5}$$

then

$$a(t) = \frac{\mu}{\delta} e^{-\delta t} - \frac{\sigma^2}{4\delta} e^{-2\delta t} + c \tag{4.6}$$

If we set c so that $a(0) = 0$ then $f(X_0, 0) = \log X_0$ and

$$a(t) = \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta t}\right) - \frac{\mu}{\delta}\left(1 - e^{-\delta t}\right) \tag{4.7}$$

Next we have

$$f(X_t, t) = f(X_0, 0) + \sigma \int_0^t e^{-\delta u} \, dW_u \tag{4.8}$$

and therefore

$$\log X_t = \log X_0 + \frac{\mu}{\delta}\left(1 - e^{-\delta t}\right) - \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta t}\right) + \sigma \int_0^t e^{-\delta u} \, dW_u \tag{4.9}$$

Consider the drift of the process $Y_t = \log X_t$ in Equation (4.9). This drift is 0 at time $t = 0$. Because

$$\frac{d}{dt}\left[\frac{\mu}{\delta}\left(1 - e^{-\delta t}\right) - \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta t}\right)\right] = \mu e^{-\delta t} - \frac{\sigma^2}{2} e^{-2\delta t} \tag{4.10}$$

the drift is positive and increasing if $\mu \geq \frac{\sigma^2}{2}$. As time goes to infinity, the drift tends to a positive limit $\frac{\mu}{\delta} - \frac{\sigma^2}{4\delta}$.

Figure 4.1 illustrates a sample path and the mean function of this process for parameter values $\mu = 0.08$, $\sigma = 0.4$, and $\delta = 0.05$. Each individual sample path can be distant from the mean function, and the probability weighted average of all the sample paths converges to the mean function.

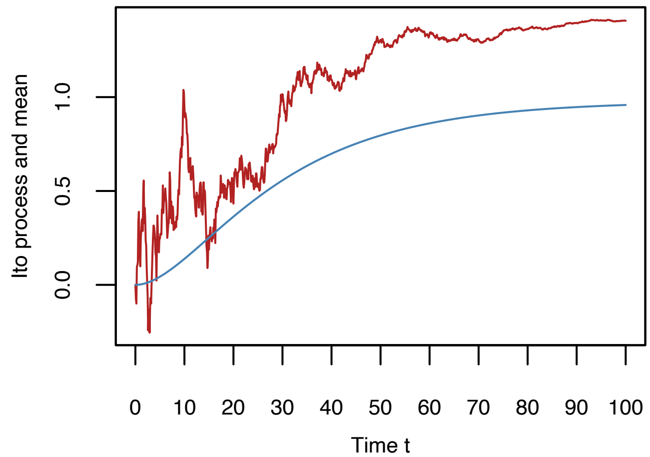

Figure 4.1. A sample path of the Ito process in Equation (4.9) (red line) and its mean function (blue line). Parameter values are $\mu = 0.08$, $\sigma = 0.4$, $\delta = 0.05$.

The process $Y_t = \log X_t$ has normal distribution with mean and variance

$$\begin{aligned} \mathbb{E}(Y_t) &= Y_0 + \frac{\mu}{\delta}\left(1 - e^{-\delta t}\right) - \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta t}\right) \\ \operatorname{var}(Y_t) &= \frac{\sigma^2\left(1 - e^{-2\delta t}\right)}{2\delta} \end{aligned} \tag{4.11}$$

The random variable $Y = \lim_{t \to \infty} Y_t$ has normal distribution with mean and variance

$$\begin{aligned} \mathbb{E}(Y) &= Y_0 + \frac{\mu}{\delta} - \frac{\sigma^2}{4\delta} \\ \operatorname{var}(Y) &= \frac{\sigma^2}{2\delta} \end{aligned} \tag{4.12}$$

The time of arrival of symptoms is $\tau = F^{-1}\left[\Phi^\star(Y)\right]$, where $\Phi^\star$ is the normal distribution function with mean and variance shown in Equation (4.12) and F is Weibull with shape parameter $k = 1.8$ and scale parameter $\lambda = 22$. The random time τ has the specified Weibull distribution.

Table 4.1 shows the generated values of the health marker process Y_t and, in the last column, the generated time of onset of symptoms in days.

Table 4.1. Simulated values of health marker and time of onset of symptoms

	Scores							
	Time 0	Time 10	Time 20	Time 30	Time 40	Time 50	Time 60	Time of Onset
1	-0.31	-0.35	-0.44					28
2	-0.54	-0.37						13
3	-0.42	-0.66						19
4	1.23	1.43	2.12	1.87	1.69	1.55		53
5	-0.26	-0.23	0.11	0.07				34
6	0.39	0.47	0.42	0.51				39
7	-0.59	-0.56						14
8	-1.26	-1.09	-1.07					22
9	-0.98	-0.73	-0.85					26
10	-0.61	-0.40						12
11	-0.74	-0.58						17
12	0.37	0.72	0.63	0.77				34
13	-0.55	-0.62						11
14	1.00	1.03	1.92	1.85	1.99	1.76		58
15	1.84	1.47	1.52	2.11	2.06			47
16	-0.31	-0.45	-0.38					21
17	-0.40	-0.36						18
18	-3.22							7
19	-0.26	-0.15						16
20	-1.26	-1.09	-1.07					21
21	-0.76	-0.64	-0.53					24
22	-3.01	-2.53						14
23	-0.56	-0.39	-0.26					29
24	-0.48	-0.55	-0.53					21
25	-0.83	-0.66						18
26	-1.01	-1.14	-0.85					23
27	-2.66							8
28	-0.42	-0.29						11
29	0.03	-0.37	0.24	0.17				31
30	-0.63	-0.57						18

4.2 Parameter Estimation of the Ito Process Y_t

Let $Y_t = h(t) + \sigma \int_0^t e^{-\delta u}\, dW_u$. The conditional distribution of Y_t given $\mathcal{F}_s^Y$ is normal and its moments follow from writing

$$Y_t = Y_s + Y_t - Y_s \tag{4.13}$$

Because $Y_t - Y_s$ is independent of Y_s we can write

$$\begin{aligned} \mathbb{E}\left(Y_t \middle| \mathcal{F}_s^Y\right) &= Y_s + \mathbb{E}(Y_t - Y_s) \\ &= Y_s + h(t) - h(s) + \sigma \mathbb{E}\left(\int_s^t e^{-\delta u}\, dW_u\right) \\ &= Y_s + h(t) - h(s) \end{aligned} \tag{4.14}$$

Also

$$\begin{aligned} \operatorname{var}\left(Y_t \middle| \mathcal{F}_s^Y\right) &= \operatorname{var}(Y_t - Y_s) \\ &= \sigma^2 \operatorname{var}\left(\int_s^t e^{-\delta u}\, dW_u\right) \\ &= \sigma^2 \int_s^t e^{-2\delta u}\, du \\ &= \sigma^2 \frac{e^{-2\delta s} - e^{-2\delta t}}{2\delta} \end{aligned} \tag{4.15}$$

The function $h(t) = Y_0 + \frac{\mu}{\delta}\left(1 - e^{-\delta t}\right) - \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta t}\right)$ depends on the parameters δ, μ, and σ. Letting $t \to \infty$ in Equation (4.14) we get

$$\mathbb{E}\left(Y \middle| \mathcal{F}_s^Y\right) = Y_s + \frac{\mu}{\delta} e^{-\delta s} - \frac{\sigma^2}{4\delta} e^{-2\delta s} \tag{4.16}$$

Similarly, from Equation (4.15) we have

$$\operatorname{var}\left(Y \middle| \mathcal{F}_s^Y\right) = \sigma^2 \frac{e^{-2\delta s}}{2\delta} \tag{4.17}$$

We observe a sample $y_1, \ldots, y_n$ of the process Y_t at times $s_1, \ldots, s_n$. The joint density of the sample is the product

$$f(y_1, \ldots, y_n) = f(y_1) f\left(y_2 \middle| y_1\right) f\left(y_3 \middle| y_1, y_2\right) \cdots f\left(y_n \middle| y_1, \ldots, y_{n-1}\right) \tag{4.18}$$

Therefore, we can write the log-likelihood function for the parameters δ, μ, σ, indexed by the given sample

$$\begin{aligned} \ell(\delta, \mu, \sigma; y_1, \ldots, y_n) &= -\frac{1}{2} \sum_{i=1}^{n} \log\left[\eta(s_i) - \eta(s_{i-1})\right] \\ &- \frac{1}{2} \sum_{i=1}^{n} \frac{\left[y_i - y_{i-1} - h(s_i) + h(s_{i-1})\right]^2}{\eta(s_i) - \eta(s_{i-1})} \end{aligned} \tag{4.19}$$

where $h(s) = \frac{\mu}{\delta}\left(1 - e^{-\delta s}\right) - \frac{\sigma^2}{4\delta}\left(1 - e^{-2\delta s}\right)$, $h(t) - h(s)$ is the conditional mean of $Y_t - Y_s$ given $\mathcal{F}_s^Y$, $\eta(s) = -\sigma^2 \frac{e^{-2\delta s}}{2\delta}$, $\eta(t) - \eta(s)$ is the conditional variance of Y_t given $\mathcal{F}_s^Y$, and $y_0 = s_0 = 0$.

For estimation, I used the R function `optim`. Table 4.2 shows the true and the estimated parameter values.

Table 4.2. True and estimated parameter values.

Parameter	True	Estimated
δ	0.05	0.045
μ	0.08	0.053
σ	0.40	0.494

4.3 Cox Proportional Hazards Model and Partial Likelihood

Cox Cox [12] and Cox [13] proposed a relative risk model with explanatory variables, or covariates, on which the failure times of individuals may depend. Consider a vector of covariates $Z' = \left(Z_1, \ldots, Z_p\right)$ and a hazard function indexed by those covariates in the following way

$$h(t;Z) = h_0(t)\exp(Z'\beta) \tag{4.20}$$

where $h_0(t)$ is an unknown baseline hazard function and β is the vector of parameters $\beta' = \left(\beta_1, \ldots, \beta_p\right)$. In the basic model, the covariates Z and the parameters β are constant, and known at the time origin $t = 0$. The baseline hazard $h_0(t)$ can be interpreted as the hazard $h(t;0)$ at $Z = 0$. The Cox model can be also written in a logarithmic form

$$\log h(t;Z) = \log h_0(t) + \exp(Z'\beta) \tag{4.21}$$

For many random times τ the Cox proportional hazards model can be derived as follows. Denote by $H(t)$ the cumulative hazard function of the survival time τ, and let g be a strictly increasing function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $g(0) = 0$. Then

$$\begin{aligned} \mathbb{P}\Big[g(\tau) > t\Big] &= \mathbb{P}\Big[\tau > g^{-1}(t)\Big] \\ &= \exp\Big\{-H\Big[g^{-1}(t)\Big]\Big\} \end{aligned} \tag{4.22}$$

Therefore, $H\Big[g^{-1}(t)\Big]$ is the cumulative hazard function of the random time $g(\tau)$.

Now let τ have an exponential distribution with rate parameter λ, and suppose that the function g is differentiable. Then the random time $g(\tau)$ has the cumulative hazard function $\lambda g^{-1}(t)$ and the hazard function $\dfrac{\lambda}{g'\left[g^{-1}(t)\right]}$. If we set $\lambda = \exp(Z'\beta)$ then we get the Cox proportional hazards model

$$h(t;Z) = \frac{1}{g'\left[g^{-1}(t)\right]}\exp(Z'\beta) \tag{4.23}$$

where $h(t;Z)$ is the hazard function of the random time $g(\tau)$. Moreover, if we choose $g(t) = F^{-1}\Big[G(t)\Big]$ where F is a strictly increasing distribution function on $(0,\infty)$, and G is the exponential distribution function with rate parameter $\lambda = 1$, then the random time $g(\tau)$ has distribution function F.

Cox Cox [12] and Cox [13] proposed a partial likelihood method for the estimation of the parameters β which treats the baseline hazard $h_0(t)$ as a nuisance function. Assume for now that there is no censoring and that there are no ties of failure times, so that exactly one individual fails at each failure time, I will return to these issues later. Consider n individuals with failure times τ_i where $1 \leq i \leq n$. Denote by A_i the event that individual i does not fail before time t and fails in the time interval $[t, t+\Delta t)$. Then the probability of the event A_i is

$$\mathbb{P}(A_i) = \frac{\mathbb{P}(t \le \tau_i < t+\Delta t)}{\mathbb{P}(t \le \tau_i)} \tag{4.24}$$

Select a specific individual i and compute the conditional probability that i will fail in the time interval $[t, t+\Delta t)$, given that exactly one of the n individuals fails in the time interval $[t, t+\Delta t)$. In other words, assuming that there are no tied failure times, and given that one of them fails at time t, what is the conditional probability it will be individual i.

We are interested in the conditional probability of specific individual i surviving up to but not including time t and failing in the time interval $[t, t+\Delta t)$, given that exactly one of the n individuals alive just prior to time t fails in the time interval $[t, t+\Delta t)$. Then the conditioning event is $A_1 \cup \cdots \cup A_n$ and $\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$. Therefore

$$\begin{aligned} \mathbb{P}(A_i | A_1 \cup \cdots \cup A_n) &= \frac{\mathbb{P}(A_i)}{\mathbb{P}(A_1 \cup \cdots \cup A_n)} \\ &= \frac{\mathbb{P}(A_i)}{\sum_{i=1}^{n} \mathbb{P}(A_i)} \\ &= \frac{\dfrac{\mathbb{P}(t \le \tau_i < t+\Delta t)}{\mathbb{P}(t \le \tau_i)}}{\sum_{j=1}^{n} \dfrac{\mathbb{P}\left(t \le \tau_j < t+\Delta t\right)}{\mathbb{P}\left(t \le \tau_j\right)}} \end{aligned} \tag{4.25}$$

Because the distribution function of each failure time τ_j is differentiable, it is also continuous, and when Δt goes to zero we get that the desired conditional probability is

$$\mathbb{P}(A_i | A_1 \cup \cdots \cup A_n) = \frac{h_i(t)}{\sum_{j=1}^{n} h_j(t)} \tag{4.26}$$

where $h_j(t)$ is the hazard function of the failure time τ_j computed at time t.

The way this is converted into a partial likelihood function is illustrated by the following example. We have a sample of $n = 5$ individuals $N = (1, 2, 3, 4, 5)$ and their failure times, without censoring, $T = (7, 3, 11, 5, 8)$. We also have one covariate Z with corresponding values $Z = (0, 1, 2, -1, 3)$.

Arrange the failure times in increasing order and arrange the individuals and covariate values in the corresponding order, that is, $T = (3, 5, 7, 8, 11), N = (2, 4, 1, 5, 3)$, and $Z = (1, -1, 0, 3, 2)$. The conditional probability that the failure time of individual $i = 2$ is $t_i = 3$, given that the failure time of one individual is $t = 3$, is

$$\frac{\exp(\beta)}{\exp(\beta) + \exp(-\beta) + 1 + \exp(3\beta) + \exp(2\beta)} \tag{4.27}$$

Equation (4.27) comes from the Cox model in Equation (4.20), and the fact that we can cancel the baseline hazard $H_0(t)$ in the numerator and the denominator of Equation (4.27).

Next, the conditional probability that the failure time of individual $i = 4$ is $t_i = 5$, given that the failure time of one individual is $t = 5$, is

$$\frac{\exp(-\beta)}{\exp(-\beta) + 1 + \exp(3\beta) + \exp(2\beta)} \tag{4.28}$$

Continuing in this manner, we get the partial likelihood function

$$
\begin{aligned}
\mathcal{L}_p &= \frac{\exp(\beta)}{\exp(\beta)+\exp(-\beta)+1+\exp(3\beta)+\exp(2\beta)} \\
&\times \frac{\exp(-\beta)}{\exp(-\beta)+1+\exp(3\beta)+\exp(2\beta)} \\
&\times \frac{1}{1+\exp(3\beta)+\exp(2\beta)} \\
&\times \frac{\exp(3\beta)}{\exp(3\beta)+\exp(2\beta)}
\end{aligned}
\tag{4.29}
$$

The partial likelihood function in Equation (4.29), and the corresponding partial log-likelihood, have a maximum at $\beta = -0.6$, as shown in Panel (A) of Figure 4.2. It is important to note that the partial likelihood function does not always have a maximum at a finite value of β. For example, if we change the vector of the covariate Z to $Z = (3, 5, 1, 4, 2)$, then the partial log-likelihood function is increasing asymptotically toward zero, as illustrated Panel (B) of Figure 4.2.

The reason for the non-existence of a maximum is that, when sorted in the order of failure times, the covariate vector is $Z = (5, 4, 3, 2, 1)$ and the partial likelihood function is

$$
\begin{aligned}
\mathcal{L}_p &= \frac{\exp(5\beta)}{\exp(5\beta)+\exp(4\beta)+\exp(3\beta)+\exp(2\beta)+\exp(\beta)} \\
&\times \frac{\exp(4\beta)}{\exp(4\beta)+\exp(3\beta)+\exp(2\beta)+\exp(\beta)} \\
&\times \frac{\exp(3\beta)}{\exp(3\beta)+\exp(2\beta)+\exp(\beta)} \\
&\times \frac{\exp(2\beta)}{\exp(2\beta)+\exp(\beta)} \\
&= \frac{1}{1+\exp(-\beta)+\exp(-2\beta)+\exp(-3\beta)+\exp(-4\beta)} \\
&\times \frac{1}{1+\exp(-\beta)+\exp(-2\beta)+\exp(-3\beta)} \\
&\times \frac{1}{1+\exp(-\beta)+\exp(-2\beta)} \\
&\times \frac{1}{1+\exp(-\beta)}
\end{aligned}
\tag{4.30}
$$

It is clear from the last four lines of Equation (4.30) that each one of the multiplicative terms there is an increasing function of β, therefore the product does not have a maximum for a finite value of β.

This problem with maximization of the partial likelihood is known, but does not appear to be widely discussed in the literature. One mention of this problem appears in Kalbfleisch and Prentice [23], where it is relegated to item 4.1 in the section *Exercises and Complements* in Chapter 4.

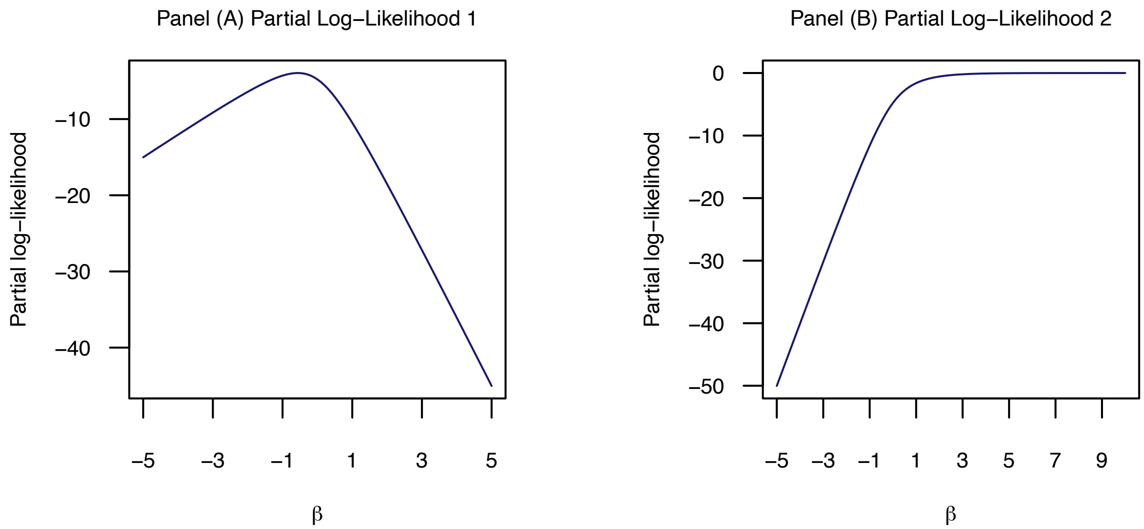

Figure 4.2. Panel (A) shows the partial log-likelihood function in this example when $Z = (0,1,2,-1,3)$, and Panel (B) shows the partial log-likelihood function when $Z = (3,5,1,4,2)$.

The partial likelihood estimator of the parameter β is the value of β that maximizes the partial likelihood function or partial log-likelihood function, if a maximum exists.

4.3.1 General Formulation of Partial Likelihood

We can write an expression for the survival function in the Cox model, but it depends on the unknown baseline survival function. First, the cumulative hazard function is $H(t;Z) = H(t_0)\exp(Z'\beta)$, therefore

$$\begin{aligned} \log S(t;Z) &= -H(t;Z) \\ &= -H_0(t)\exp(Z'\beta) \\ &= \exp(Z'\beta)\log S_0(t) \end{aligned} \tag{4.31}$$

Therefore, $S(t;Z) = \left[S_0(t)\right]^{\exp(Z'\beta)}$. Next, we can write the likelihood function for n individuals with independent failure times $t_1 < t_2 < \cdots < t_n$ (no ties of failure times). Each density at failure time is $f(t_i;Z_i) = h(t_i;Z_i)S(t_i;Z_i)$. The likelihood function is

$$\begin{aligned} \mathcal{L}(\beta;t_1,\ldots,t_n,Z_1,\ldots,Z_n) &= \prod_{i=1}^{n} f(t_i;Z_i) \\ &= \prod_{i=1}^{n} h(t_i;Z_i)S(t_i;Z_i) \end{aligned} \tag{4.32}$$

For each $1 \le i \le n$ the risk set R_i is the set of all the indexes $i \le j \le n$, that is, the index set of all the individuals who are at risk at time t_i. Multiplying and dividing by $\sum_{j \in R_i} h_0(t_i)\exp\left(Z_j'\beta\right)$, we can write the likelihood function

$$\begin{aligned} \mathcal{L}(\beta;t_1,\ldots,t_n,Z_1,\ldots,Z_n) &= \prod_{i=1}^{n} \frac{h_0(t_i)\exp(Z_i'\beta)}{\sum_{j \in R_i} h_0(t_i)\exp\left(Z_j'\beta\right)} \left[\sum_{j \in R_i} h_0(t_i)\exp\left(Z_j'\beta\right)\right] S(t_i;Z_i) \\ &= \prod_{i=1}^{n} \frac{\exp(Z_i'\beta)}{\sum_{j \in R_i} \exp\left(Z_j'\beta\right)} \left[\sum_{j \in R_i} h_0(t_i)\exp\left(Z_j'\beta\right)\right] S(t_i;Z_i) \end{aligned} \tag{4.33}$$

The right side of the second line of Equation (4.33) is a product of three factors, the first of which is $\frac{\exp(Z_i'\beta)}{\sum_{j \in R_i} \exp\left(Z_j'\beta\right)}$ corresponding to the calculation of conditional probability in Equation (4.26). This factorization of the likelihood function, treating the baseline hazard $h_0(t)$ as a nuisance function, is the basis for the partial likelihood

$$\mathcal{L}_p(\beta;Z_1,\ldots,Z_n) = \prod_{i=1}^{n} \frac{\exp(Z_i'\beta)}{\sum_{j \in R_i} \exp\left(Z_j'\beta\right)} \tag{4.34}$$

When we use only one factor in Equation (4.33) and ignore the two other factors that also involve β, we are not using all the information in the observed sample. The benefits are, as argued by Cox [12] and Cox [13], simplification and an increase in robustness.

When right censoring is present, the partial likelihood function becomes

$$\mathcal{L}_p(\beta; Z_1, \ldots, Z_n, \delta_1, \ldots, \delta_n) = \prod_{i=1}^{n} \left[\frac{\exp(Z_i'\beta)}{\sum_{j \in R_i} \exp\left(Z_j'\beta\right)} \right]^{\delta_i} \tag{4.35}$$

where δ_i is the censoring indicator at time t_i, that is, $\delta_i = 1$ if t_i is a failure time, and $\delta_i = 0$ if t_i is a censoring time.

Cox Cox [12] points out that the covariates Z can be deterministic functions of time. The form of the partial likelihood function will then be a straightforward extension of the partial likelihood with constant covariates. For example, consider $n = 5$ individuals with ordered failure times $T = (31, 32, 45, 52, 53)$, corresponding values of the first covariate $Z_1 = (1, 0, 1, 1, 0)$, and a second covariate defined as the function of time $Z_2(t) = Z_1(\log t - c)$, where c is the average value of $\log t$ in this sample. Then, $Z_2 = (-0.292, 0, 0.081, 0.226, 0)$, and we get that partial likelihood is maximized for $\beta_1 = 2.004$ and $\beta_2 = -8.884$. If, however, $Z_2(t) = \log t - c$ then the partial likelihood function does not have a maximum.

The partial likelihood function can also be modified to allow tied failure times. Cox Cox [12] proposed the following method. Denote by d_i the number of individuals failing at time i. Then, the calculation of conditional probability in Equation (4.25) is replaced by the conditional probability of the specific d_i individuals surviving up to but not including time t and failing in the time interval $[t, t+\Delta t)$, given that exactly d_i of the n individuals alive just prior to time t fail in the time interval $[t, t+\Delta t)$. With that modification, the partial likelihood changes to

$$\mathcal{L}_p(\beta; Z_1, \ldots, Z_n, d_1, \ldots, d_n) = \prod_{i=1}^{n} \frac{\exp(U_i'\beta)}{\sum_{j \in S_i} \exp\left(V_{jK}'\beta\right)} \tag{4.36}$$

where S_j is the set of all d_j-tuples from the risk set at time t_j, U_i is the sum of the covariates Z for the d_i individuals that fail at time t_i, and V_{jk} is the sum of the covariates Z for the d_j individuals in d_j-tuple number k. As this calculation may be time consuming in the presence of many ties, Efron [16] has proposed a fast and accurate approximation.

An asymptotic theory of maximum likelihood estimation based on a partial likelihood appears in Wong [41].

For limitations of the Cox model with time-dependent covariates see Fisher and Lin [18].

The following is an important and well-known fact about partial likelihood estimation. If instead of the sample $T = (7, 3, 11, 5, 8)$, we have a sample such that $S = \left[g(7), g(3), g(11), g(5), g(8)\right]$ where g is a positive increasing function, then the partial likelihood function, constructed as above, would be exactly the same. This has an important implication for the comparison of the Cox model with the model introduced in Chapter 3.

In the model introduced in Chapter 3, we have a random variable X with distribution function Φ, that can be, but doesn't have to be, normal. We then consider another distribution F and define a survival time $\tau = F^{-1}\left[\Phi(X)\right]$ which has the distribution function F.

Consider two random times generated by the random variable $X = \int_0^\infty b(u)\,dW_u$, defined by $\tau_1 = F_1^{-1}\left[\Phi(X)\right]$ and $\tau_2 = F_2^{-1}\left[\Phi(X)\right]$.

Then $F_1(\tau_1) = F_2(\tau_2)$, and τ_2 is a positive increasing function of τ_1, *i.e.*, $\tau_2 = F_2^{-1}\left[F_1(\tau_1)\right]$, and the estimated parameters of the Cox model are the same for both survival times τ_1 and τ_2.

In the model introduced in Chapter 3, the conditional hazard functions of the two survival times, given $X_t = \int_0^t b(u)\,dW_u$, are different.

For clinical applications of the Cox model, see for example Spruance et al. [33].

4.4 Estimating a Cox Model with Time Dependent Covariates

Next, I use the partial likelihood method of Cox to estimate the proportional hazards model. The idea is to recognize that we have time dependent covariate and to assume that it is a step functions of time. Thus, for subject 1 in Table 4.1 the covariate $X_1(t)$ has a constant value of -0.31 between time 0 and time 9, a constant value of -0.35 between time 10 and time 19, and a constant value of -0.44 between time 20 and time 28.

For estimation, I used the function `coxph` of the R package `survival`. The function `coxph` requires as input an object of class `survival`, which I built with the `Surv` function by entering each subject several times using censoring intervals without an event, and one last time with an event. For example, for subject 1 in Table 4.1 the entries are shown in Table 4.3.

Table 4.3. Data entry for subject 1 in Table 4.1.

Start	Stop	Event
0	10	0
10	20	0
20	28	1

Patient 18 in Table 4.1 gets only one input line as shown in Table 4.4.

Table 4.4. Data entry for subject 18 in Table 4.1.

Start	Stop	Event
0	7	1

Table 4.5 shows the estimation results of the parameter of the Cox proportional hazards model for all the 30 subjects. The 30 subjects generate 86 lines of data with 30 events. Likelihood ratio test statistic is 27.76 with p-value of 1×10^{-7}, and the Wald test statistic is 24.03 with p-value of 9×10^{-7}.

Table 4.5. Estimation results for the Cox proportional hazards model with time dependent covariate.

Parameter	z-Score	p-Value
-1.3655	-4.9020	9.47×10^{-7}

exp(Parameter)	Lower 0.95	Upper 0.95
0.2553	0.1479	0.4406

4.5 Model Comparison

We can use the 30 values of time to onset of symptoms, as shown in the last column of Table 4.1, to compute the Kaplan-Meier estimator of the time to onset of symptoms. Using the `survfit` function of the R package `survival` we get the survival function presented in Figure 4.3.

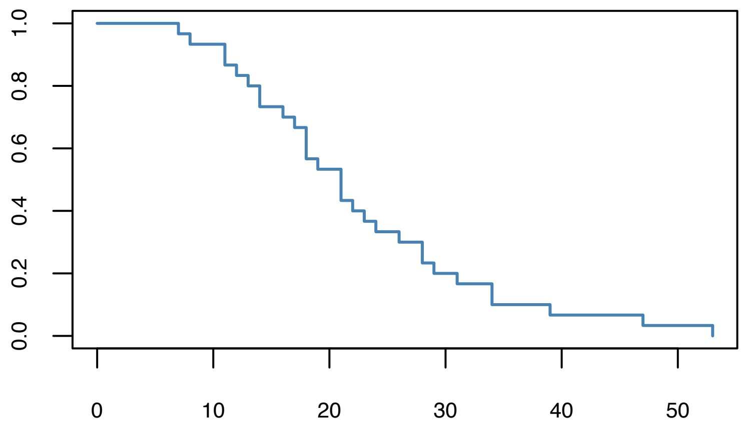

Figure 4.3. Kaplan-Meier marginal survival function for all subjects.

Using the general formula for conditional survival functions in Equation (3.41) and the estimated values of model parameters in Table 4.2, we can plot the individual conditional survival functions for every subject. Figure 4.4 shows those conditional survival functions, given the measured value of the health marker at time 0, for subjects 1 (blue), 8 (red), 15 (olive), 22 (magenta), and 29 (gray).

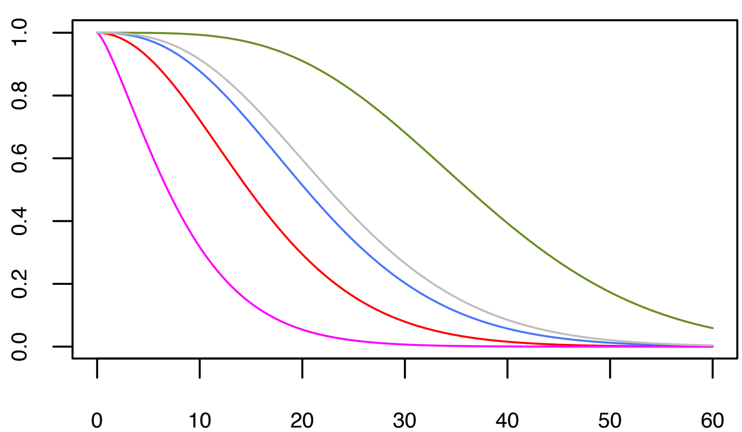

Figure 4.4. New model individual conditional survival function for subjects 1 (blue), 8 (red), 15 (olive), 22 (magenta), and 29 (gray).

We can see in Table 4.1 that larger values of the health marker are associated with longer time to onset of symptoms. Table 4.1 and Figure 4.4 show a good correspondence between between observed health markers at time 0, individual survival times, and individual conditional survival functions.

The Cox model allows the calculation of hazard ratios, and the new model introduced in Chapter 3 allows the calculation of the individual conditional hazard and survival functions. To do a comparison of the two models, I calculated and plotted the hazard ratios for both models for some of the 30 subjects listed in Table 4.1.

An examination of Table 4.1 shows that subject 14 has the longest time to the onset of symptoms, and therefore, it makes sense to use the results for subject 14 as the denominator in hazard ratios for all the other subjects.

Consider for example subject 12. From Table 4.1, the measured health marker of subject 12 at time 0 is 0.37, and the measured health marker for subject 14 at time 0 is 1.00. The estimated parameter of the Cox model is a single number $\beta = -1.3655$, therefore, the estimated hazard ratio of subject 12 to subject 14 is constant in the time interval from day 0 to day 9 with the value

$$\begin{aligned} \frac{\exp\left[\beta z_{12}(0)\right]}{\exp\left[\beta z_{14}(0)\right]} &= \frac{\exp(-1.3655 \times 0.37)}{\exp(-1.3655 \times 1.00)} \\ &= 2.364 \end{aligned} \tag{4.37}$$

Continuing this comparison to time 10, we get the hazard ratio for times 10 through 19

$$\frac{\exp\left[\beta z_{12}(10)\right]}{\exp\left[\beta z_{14}(10)\right]} = \frac{\exp(-1.3655 \times 0.72)}{\exp(-1.3655 \times 1.03)} = 1.527 \tag{4.38}$$

For times 20 to 29 the hazard ratio is

$$\frac{\exp\left[\beta z_{12}(20)\right]}{\exp\left[\beta z_{14}(20)\right]} = \frac{\exp(-1.3655 \times 0.63)}{\exp(-1.3655 \times 1.92)} = 5.821 \tag{4.39}$$

For times 30 to 34 we get the hazard ratio

$$\frac{\exp\left[\beta z_{12}(30)\right]}{\exp\left[\beta z_{14}(30)\right]} = \frac{\exp(-1.3655 \times 0.77)}{\exp(-1.3655 \times 1.85)} = 4.370 \tag{4.40}$$

Figure 4.5 shows the log of estimated hazard ratio in the Cox model for subjects 12 and 14.

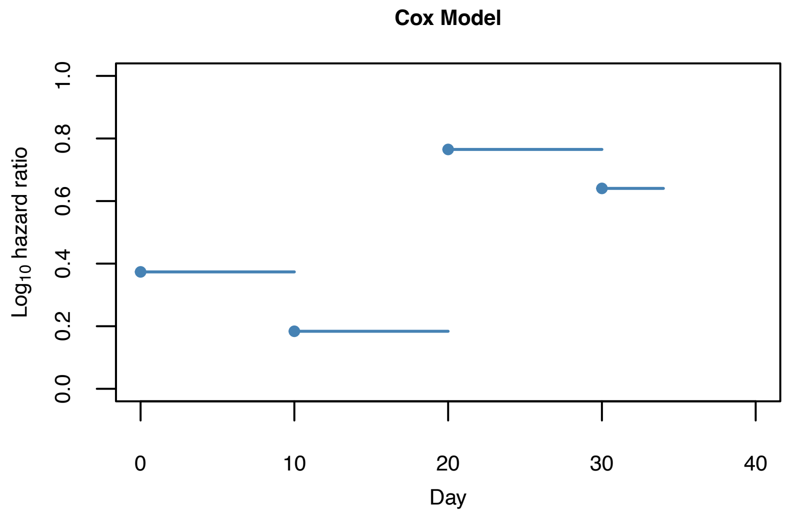

Figure 4.5. Log hazard ratio for subjects 12 and 14 in a Cox model with time-varying covariates.

Figure 4.6 shows the log of estimated hazard ratio in the new model for subjects 12 and 14. The hazard ratio is not a step function here, and there is a particularly large increase on day 20, when the measured health marker of subject 14 increases significantly. An examination of the two figures reveals that all the piecewise constant values of the hazard ratio intersect the corresponding lines in the new model.

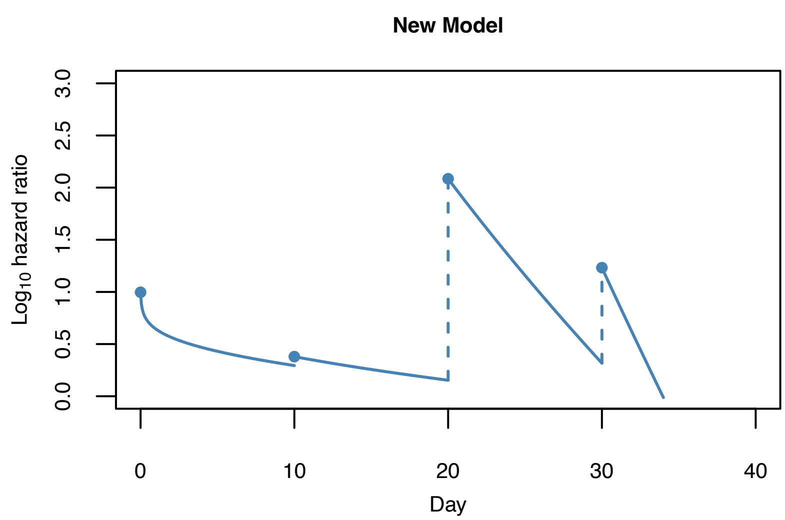

Figure 4.6. Log hazard ratio for subjects 12 and 14 in the new model.

4.6 A Case When the Cox Model Cannot Be Used

The new model delivers results even when the Cox model cannot be used because the partial likelihood function does not have a maximum.

Consider a health marker that evolves as

$$dX_t = b e^{-\delta t} X_t dW_t \quad b \text{ and } \delta \text{ positive constants} \tag{4.41}$$

Apply Ito's formula with $f(x) = \log x$ to get

$$d \log X_t = -\frac{1}{2} b^2 e^{-2\delta t} dt + b e^{-\delta t} dW_t$$

and assume parameter values: $X_0 = 100$, $b = 0.05$, and $\delta = 0.03$.

Integrating both sides

$$\begin{aligned} \log \frac{X_t}{X_0} &= -\frac{1}{2} b^2 \int_0^t e^{-2\delta u} du + b \int_0^t e^{-\delta u} dW_u \\ &= -\frac{b^2}{2} \frac{1 - e^{-2\delta t}}{2\delta} + b \int_0^t e^{-\delta u} dW_u \end{aligned} \tag{4.42}$$

Therefore

$$X_t = X_0 \exp\left(-\frac{b^2}{2} \frac{1 - e^{-2\delta t}}{2\delta} + b \int_0^t e^{-\delta u} dW_u \right) \tag{4.43}$$

The positive, continuous martingale

$$X_t = X_0 \exp\left(-\frac{b^2}{2}\,\frac{1-e^{-2\delta t}}{2\delta} + b\int_0^t e^{-\delta u}\, dW_u\right) \tag{4.44}$$

converges with probability 1 to the random variable

$$X = X_0 \exp\left(-\frac{b^2}{4\delta} + b\int_0^\infty e^{-\delta u}\, dW_u\right) \tag{4.45}$$

The variance of the Ito integral is

$$\operatorname{var}\left(\int_0^t e^{-\delta u}\, dW_u\right) = \int_0^t e^{-2\delta u}\, du = \frac{1-e^{-2\delta t}}{2\delta} \tag{4.46}$$

Therefore:

1. Distribution of X_t is log-normal
   $Y_t = \log X_t \sim \mathbb{N}\left(\log X_0 - \frac{b^2}{2}\frac{1-e^{-2\delta t}}{2\delta}, b^2\frac{1-e^{-2\delta t}}{2\delta}\right)$.
2. Distribution of X is log-normal $Y = \log X \sim \mathbb{N}\left(\log X_0 - \frac{b^2}{4\delta}, \frac{b^2}{2\delta}\right)$.

Because the increment $Y - Y_t$ is independent of $\mathcal{F}_t^Y$, the conditional distribution of Y given the history of observations $\mathcal{F}_t^Y$ is normal with conditional mean and variance

$$\begin{aligned}
\mathbb{E}\left(Y \middle| \mathcal{F}_t^Y\right) &= \mathbb{E}\left(Y_t + Y - Y_t \middle| \mathcal{F}_t^Y\right) \\
&= Y_t + \mathbb{E}\left(Y - Y_t \middle| \mathcal{F}_t^Y\right) \\
&= Y_t + \mathbb{E}\left(Y - Y_t\right) \\
&= Y_t - \frac{b^2}{4\delta} e^{-2\delta t}
\end{aligned} \tag{4.47}$$

$$
\begin{aligned}
\mathrm{var}\left(Y \middle| \mathcal{F}_t^{\,Y}\right) &= \mathrm{var}\left(Y_t + Y - Y_t \middle| \mathcal{F}_t^{\,Y}\right) \\
&= \mathrm{var}\left(Y - Y_t \middle| \mathcal{F}_t^{\,Y}\right) \\
&= \mathrm{var}\left(Y - Y_t\right) \\
&= b^2 \int_t^{\infty} \mathrm{e}^{-2\delta u}\, du \\
&= \frac{b^2}{2\delta}\,\mathrm{e}^{-2\delta t}
\end{aligned}
\tag{4.48}
$$

As before, define the random survival time as

$$
\tau \;=\; \psi(Y) \;=\; F^{-1}\left\{\Phi\left[\frac{Y - \mathbb{E}(Y)}{\sqrt{\mathrm{var}(Y)}}\right]\right\}
\tag{4.49}
$$

The conditional distribution of the survival time τ given cumulative information at observation time t is for future time $r \geq t$

$$
\begin{aligned}
\mathbb{P}\left(\tau \leq r \middle| \mathcal{F}_t^{X}\right) &= \mathbb{P}\left(Y \leq \psi^{-1}(r) \middle| \mathcal{F}_t^{\,Y}\right) \\
&= \Phi\left[\frac{\psi^{-1}(r) - \mathbb{E}\left(Y \middle| \mathcal{F}_t^{\,Y}\right)}{\sqrt{\mathrm{var}\left(Y \middle| \mathcal{F}_t^{\,Y}\right)}}\right] \\
&= \Phi\left[\frac{\psi^{-1}(r) - \log X_t}{\sigma_t} + \frac{\sigma_t}{2}\right]
\end{aligned}
\tag{4.50}
$$

where ψ is strictly increasing, and therefore, invertible, and $\sigma_t^2 = \dfrac{b^2}{2\delta}\,\mathrm{e}^{-2\delta t}$.

The conditional survival function of the survival time τ, given cumulative information at observation time t and the condition $\tau > t$, written as a function of future time $r \geq t$

$$\mathbb{P}\left(\tau > r \middle| \mathcal{F}_t^X, \tau > t\right) = \frac{1 - \Phi\left[\frac{\psi^{-1}(r) - \log X_t}{\sigma_t} + \frac{\sigma_t}{2}\right]}{1 - \Phi\left[\frac{\psi^{-1}(t) - \log X_t}{\sigma_t} + \frac{\sigma_t}{2}\right]} \tag{4.51}$$

The conditional hazard function of the survival time τ, given cumulative information at observation time t and the condition $\tau > t$, written as a function of future time $r \geq t$, is

$$\begin{aligned} h\left(r \middle| \mathcal{F}_t^X, \tau > t\right) &= \frac{f\left(r \middle| \mathcal{F}_t^X, \tau > t\right)}{\mathbb{P}\left(\tau > r \middle| \mathcal{F}_t^X, \tau > t\right)} \\ &= \frac{\phi\left[\frac{\psi^{-1}(r) - \log X_t}{\sigma_t} + \frac{\sigma_t}{2}\right]}{1 - \Phi\left[\frac{\psi^{-1}(r) - \log X_t}{\sigma_t} + \frac{\sigma_t}{2}\right]} \frac{\sigma f(r)}{\sigma_t \phi\left[\frac{\psi^{-1}(r) - \mu}{\sigma}\right]} \end{aligned} \tag{4.52}$$

Suppose Y_0 is known and we observe the process Y_t at times $t_1 < t_2 < \cdots < t_p$. Denote the observed values $y_1, y_2, \ldots, y_p$. The observations do not come from i.i.d. random variables, but accounting for the fact that the process Y_t has the Markov property, we can write the joint density

$$f\left(y_1, y_2, \ldots, y_p\right) = f(y_1) f\left(y_2 \middle| y_1\right) f\left(y_3 \middle| y_2\right) \cdots f\left(y_p \middle| y_{p-1}\right) \tag{4.53}$$

From properties of normal distribution we have for $t_{i-1} < t_i$

$$
\begin{aligned}
m_i &= \mathbb{E}\left(Y_{t_i} \middle| Y_{t_{i-1}}\right) = \mathbb{E}\left(Y_{t_i}\right) + \frac{\operatorname{cov}\left(Y_{t_i}, Y_{t_{i-1}}\right)}{\operatorname{var}\left(Y_{t_{i-1}}\right)}\left[Y_{t_{i-1}} - \mathbb{E}\left(Y_{t_{i-1}}\right)\right] \\
&= Y_{t_{i-1}} - \frac{b^2}{2}\frac{\mathrm{e}^{-2\delta t_{t_{i-1}}} - \mathrm{e}^{-2\delta t_i}}{2\delta} \\
&= Y_{t_{i-1}} - \frac{1}{2}\gamma_i
\end{aligned}
\tag{4.54}
$$

$$
\begin{aligned}
\gamma_i &= \operatorname{var}\left(Y_{t_i} \middle| Y_{t_{i-1}}\right) = \operatorname{var}\left(Y_{t_i}\right) - \frac{\operatorname{cov}^2\left(Y_{t_i}, Y_{t_{i-1}}\right)}{\operatorname{var}\left(Y_{t_{i-1}}\right)} \\
&= b^2\frac{\mathrm{e}^{-2\delta t_{i-1}} - \mathrm{e}^{-2\delta t_i}}{2\delta}
\end{aligned}
\tag{4.55}
$$

The formulas are also valid for $\mathbb{E}\left(Y_{t_1}\right)$ and $\operatorname{var}\left(Y_{t_1}\right)$ if we set $t_0 = 0$. Log-likelihood is

$$
\ell = -\frac{1}{2}\sum_{i=1}^{p}\log\gamma_i - \frac{1}{2}\sum_{i=1}^{p}\frac{(y_i - m_i)^2}{\gamma_i} + \cdots
\tag{4.56}
$$

Now generate 10 simulated subjects with health marker covariate process parameters $X_0 = 100$, $b = 0.05$, and $\delta = 0.03$, and use for the distribution function F a Weibull distribution with shape parameter 2.3 and scale parameter 20. Table 4.6 shows the values of the health marker observed at time $t = 5$ together with the values of the random survival time computed from Equation (4.49).

Table 4.6. Simulation Results of 10 Subjects.

Patient ID	Covariate X_t at $t = 5$	Standardized Covariate X_t at $t = 5$	Log Covariate Y_t at $t = 5$	Limit Log Covariate Y_t as $t \to \infty$	Survival Time τ in Years
1	100.58	0.534	4.611	4.880	30.32
2	89.16	-0.458	4.490	4.576	16.69
3	110.14	1.365	4.702	4.998	35.94
4	82.73	-1.017	4.416	4.356	8.54
5	90.37	-0.353	4.504	4.687	21.49
6	87.14	-0.634	4.468	4.510	14.04
7	112.35	1.557	4.722	5.020	36.97
8	102.60	0.710	4.631	4.890	30.81
9	91.44	-0.260	4.516	4.722	23.03
10	77.82	-1.444	4.354	4.290	6.58

Maximizing the likelihood function in Equation (4.56) I got estimates of the model parameters $\widehat{b} = 0.0493$ and $\widehat{\delta} = 0.0283$.

Figure 4.7 shows the 10 individual conditional survival functions computed from Equation (4.51) using estimated model parameters and the value of the health marker covariate observed at tine $t = 5$.

Figure 4.8 shows the 10 individual conditional hazard functions computed from Equation (4.52) using estimated model parameters and the value of the health marker covariate observed at tine $t = 5$.

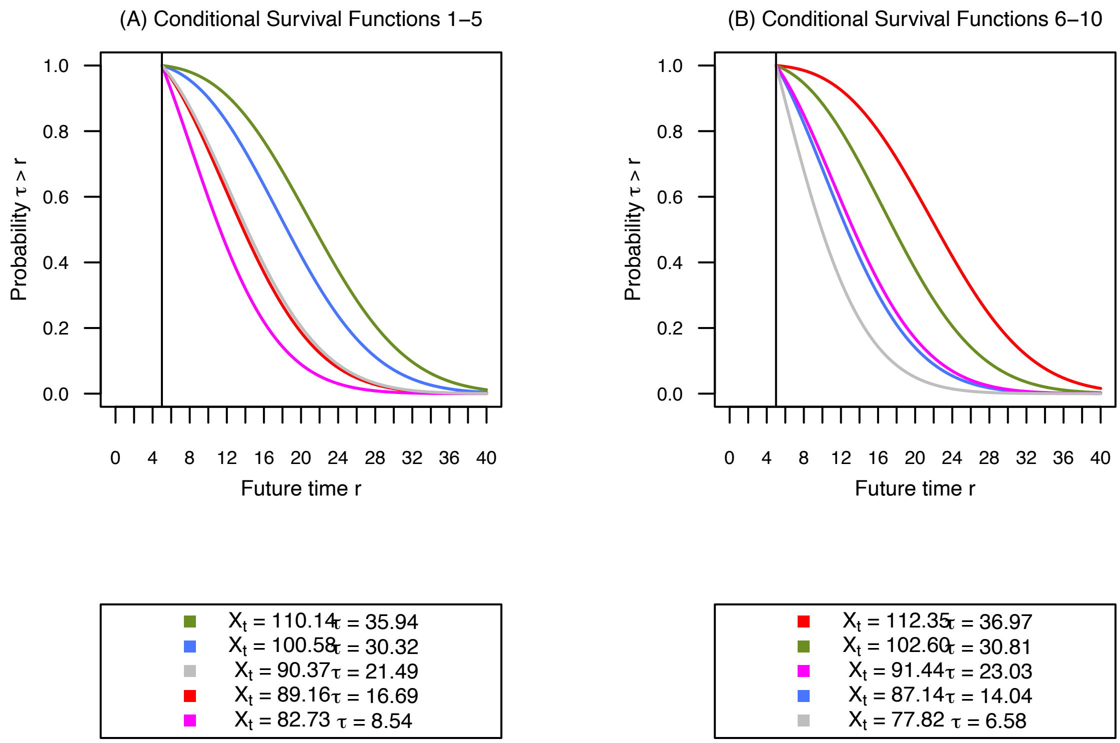

Figure 4.7. Individual conditional survival functions for the 10 subjects.

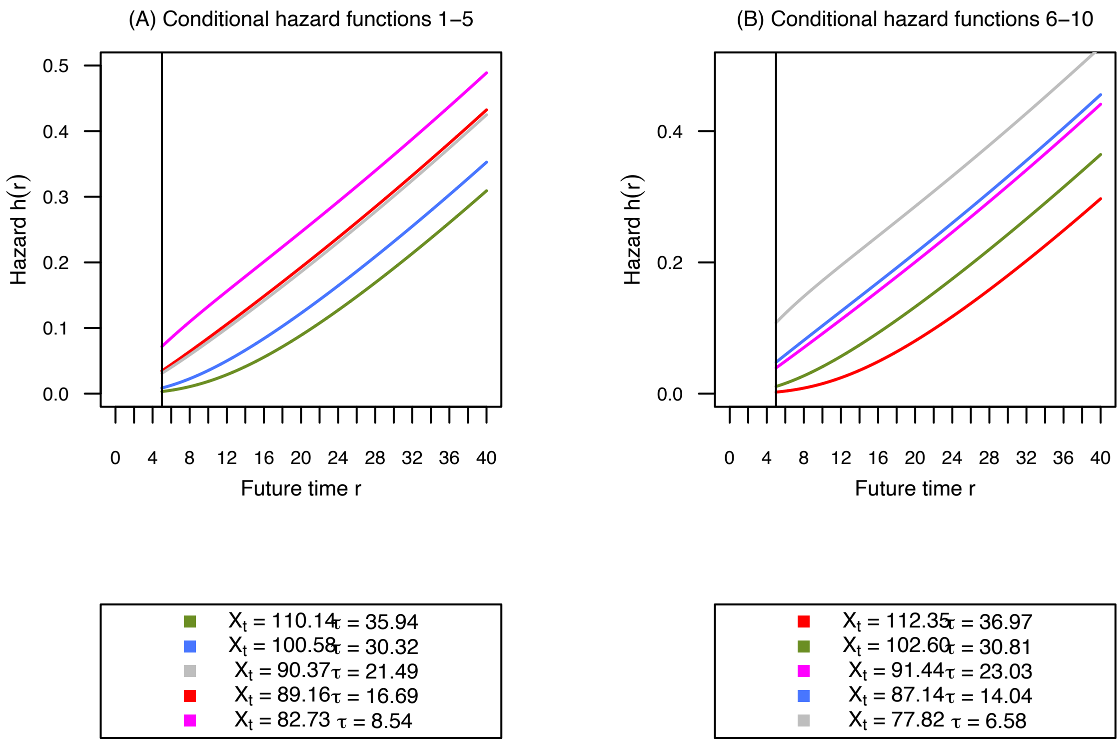

Figure 4.8. Individual conditional hazard functions for the 10 subjects.

Figure 4.9 shows the Kaplan-Meier empirical survival function in this example, together with 95% confidence intervals calculated by the exponential Greenwood method, and the 10 individual conditional survival functions calculated from the new model.

There is general agreement between the conditional survival functions and the intervals, but the conditional survival functions are more pessimistic for subjects with low covariates at observation time $t = 5$.

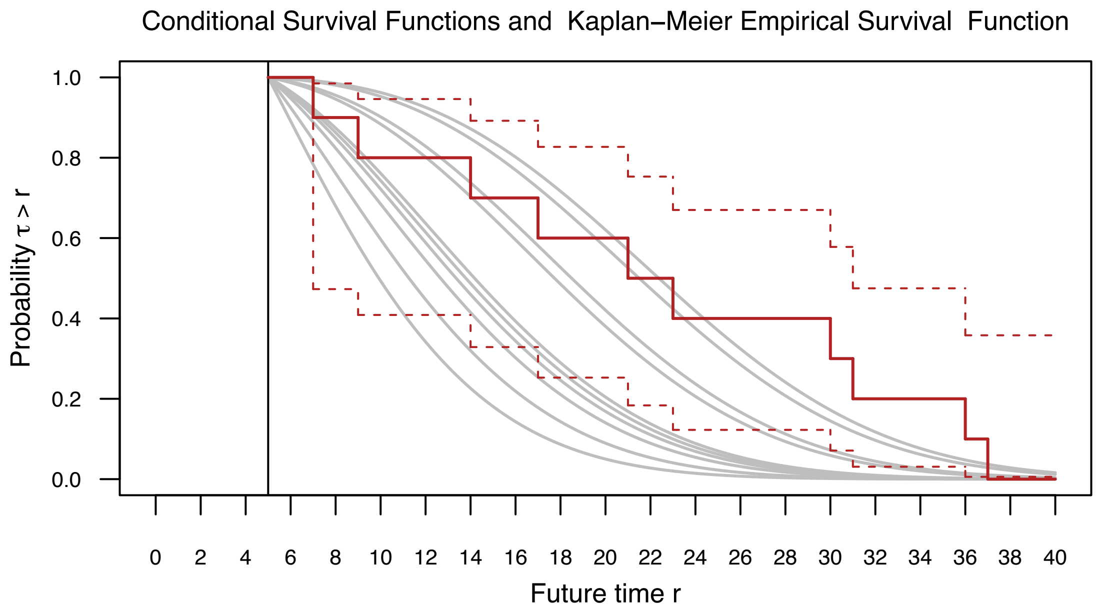

Figure 4.9. Kaplan-Meier empirical estimator and individual conditional survival functions for the 10 subjects.

For this simulation of conditional survival and hazard functions in the new model with one covariate the partial likelihood estimator of the parameter β does not exist. Figure 4.10 shows the dependence of the partial likelihood function on β using standardized covariate X_t for improved numerical stability. Cox model cannot be used in this simulation, but at observation time t the new model predicts individual hazard functions for future times $r \geq t$.

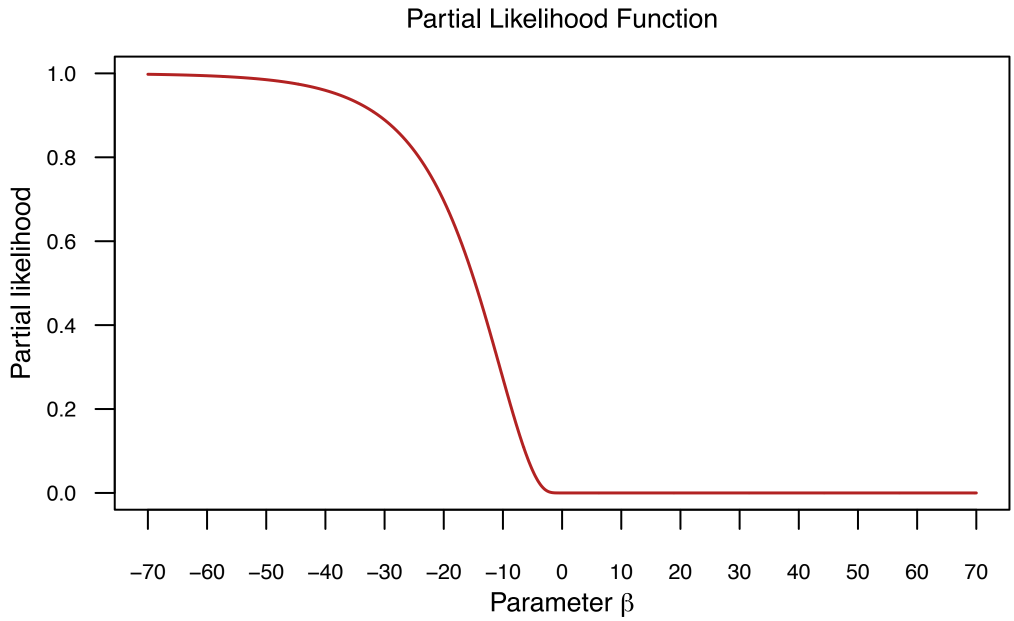

Figure 4.10. Partial likelihood function in the simulation of 10 subjects.

5 Filtering for a Random Time

5.1 Introduction

This chapter describes the stochastic filtering methodology that I propose to use to present a second family of models of random times and their conditional distributions given randomly arriving information.

Arriving information is modeled as health markers or biomarkers. A biomarker is an objective sign of medical state which can be measured accurately and reproducibly. Biomarkers Definitions Working Group [4] defines a biomarker as an objectively measured indicator of life processes or response to treatment. But even with accuracy, biomarkers are often measured with error and it is unclear that the resulting biases are sufficiently considered in the medical research literature Brakenholff et al. [6]. Examples of measurement errors include blood pressure, body chemistry, exposure to pollutants, or exposure to nutrients. Other examples of measurement errors involve functionality in relation to disability, symptoms of post-traumatic shock disorder, and symptoms of choronic inflamation.

Measurement error may arise from inaccurate instruments, discrete observation of continuous processes, lack of compliance with measurement protocols, measurement of external ambient rather than individual exposure (such as pollution), and increased use of "big data" databases such as routine care records and insurance claims, not originally intended for research purposes Brakenholff et al. [7]. One way to represent measurement errors is to introduce latent or unobservable medical states, and observable biomarkers that represent the true medical states with some measurement error.

To connect the two families of models I will briefly restate the model introduced in Chapter 3 using new notation that makes a clear distinction between observable and unobservable objects and is, therefore, more suitable for a filtering framework.

The first family of models, introduced in Chapter 3, involves an unobservable random variable

$$\theta = \int_0^\infty b(u) dW_u \tag{5.1}$$

and an observable process

$$\xi_t = \int_0^t b(u) dW_u \tag{5.2}$$

Then

$$\begin{aligned} m_t &= \mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right) \\ &= \mathbb{E}\left(\theta \middle| \mathcal{F}_t^{W}\right) \\ &= \mathbb{E}\left[\xi_t + \int_t^\infty b(u) dW_u \middle| \mathcal{F}_t^{W}\right] \\ &= \xi_t \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} \gamma_t &= \mathrm{var}\left(\theta \middle| \mathcal{F}_t^{\xi}\right) \\ &= \mathrm{var}\left(\theta \middle| \mathcal{F}_t^{W}\right) \\ &= \mathrm{var}\left[\xi_t + \int_t^\infty b(u) dW_u \middle| \mathcal{F}_t^{W}\right] \\ &= \int_t^\infty b^2(u) du \end{aligned} \tag{5.4}$$

To introduce the second family of models, consider an unobservable Ito process

$$\theta_t \quad = \quad \theta_0 + \int_0^t (a_0 + a_1\theta_u)\, du + \int_0^t b_1\, dW_{1,u} + \int_0^t b_2\, dW_{2,u} \tag{5.5}$$

and an observable Ito process

$$\xi_t \quad = \quad \xi_0 + \int_0^t (A_0 + A_1\theta_u)\, du + \int_0^t B_2\, dW_{2,u} \tag{5.6}$$

When the Ito process θ_t converges to a limiting random variable θ, this framework allows a definition of a random time $\tau = \psi(\theta)$, where ψ is a positive increasing function. The observable Ito process ξ_t represents the randomly arriving information, and we are interested in computing the conditional distribution

$$\mathbb{P}\left(\tau \leq y \middle| \mathcal{F}_t^{\xi}\right) \quad = \quad \mathbb{P}\left[\theta \leq \psi^{-1}(y) \middle| \mathcal{F}_t^{\xi}\right] \tag{5.7}$$

and its associated conditional survival, conditional hazard, conditional mean residual life, and conditional mean tail life functions.

The problem of calculating the conditional distribution $\mathbb{P}\left(\theta_t \leq y \middle| \mathcal{F}_t^{\xi}\right)$ is called a filtering problem, and the problem of calculating the conditional distribution $\mathbb{P}\left(\theta_t \leq y \middle| \mathcal{F}_s^{\xi}\right)$ where $s < t$ is called an extrapolation problem.

The filtering and extrapolation problems have solutions, but this modeling of a random time and its conditional distribution given randomly arriving information is new. This modeling is particularly suitable for practical applications with multi-dimensional information processes, and also allows the introduction of Bayesian analysis by the inclusion of a prior distribution of the initial state of the unobservable Ito process that generates the random survival time.

5.2 A Heuristic Derivation of a Filter of an Unobservable Random Variable

We have an unobservable, normal random variable θ and a discrete-time observation process

$$\xi_{t+1} = A_0(t,\xi) + A_1(t,\xi)\theta + B(t,\xi)\epsilon_{t+1} \tag{5.8}$$

The error term ϵ_{t+1} and θ are assumed to be independent, and the error term is standard normal. The prior distribution of θ is the conditional distribution $\mathbb{P}\left(\theta \leq y \middle| \xi_0\right)$ assumed to be $\mathbb{N}(\mu,\sigma)$ with known mean μ and variance σ^2. The notation for $A_0(t,\xi), A_1(t,\xi)$ and $B(t,\xi)$ is short for $A_0(t,\xi_0,\xi_1,\cdots,\xi_t), A_1(t,\xi_0,\xi_1,\cdots,\xi_t)$, and $B(t,\xi_0,\xi_1,\cdots,\xi_t)$.

As in the model in Chapter 3, because the unobservable random variable θ is normal, it generates a random time τ through $\tau = \psi(\theta)$ where $\psi = F^{-1} \circ \Phi$ with Φ denoting the normal distribution function with mean μ and variance σ^2, and F is the (strictly increasing) distribution function that we desire for the random time τ, given ξ_0.

We can redefine the function A_0 a bit and rewrite Equation (5.8) as

$$\xi_{t+1} - \xi_t = A_0(t,\xi) + A_1(t,\xi)\theta + B(t,\xi)\epsilon_{t+1} \tag{5.9}$$

To obtain an observation process in continuous time, with the error term represented by an increment of a Wiener process, we can rewrite Equation (5.9) for time increment Δt and make explicit the variance of the error term to be Δt

$$\xi_{t+\Delta t} - \xi_t = A_0(t,\xi)\Delta t + A_1(t,\xi)\theta\Delta t + B(t,\xi)\sqrt{\Delta t}\epsilon_{t+\Delta t} \tag{5.10}$$

Denote $m_t = \mathbb{E}\left(\theta \middle| \xi_0, \xi_{\Delta t}, \xi_{2\Delta t}, \ldots, \xi_t\right)$ and $\gamma_t = \mathrm{var}\left(\theta \middle| \xi_0, \xi_{\Delta t}, \xi_{2\Delta t}, \ldots, \xi_t\right)$. From the conditional distribution of a multivariate normal given another multivariate normal we get the formulas

$$\begin{aligned} \Delta m_t &= \frac{A_1(t,\xi)\gamma_t \Delta t}{B^2(t,\xi)\Delta t + \left[A_1(t,\xi)\Delta t\right]^2 \gamma_t}\left[\Delta \xi_t - A_0(t,\xi)\Delta t - A_1(t,\xi) m_t \Delta t\right] \\ \Delta \gamma_t &= -\frac{A_1^2(t,\xi)\gamma_t^2}{B^2(t,\xi)}\Delta t \end{aligned} \tag{5.11}$$

where $\Delta \xi_t = \xi_{t+\Delta t} - \xi_t$ and $\Delta m_t = m_{t+\Delta t} - m_t$. Dividing the numerator and the denominator of the fraction on the right side of the first equation in Equation (5.11) by Δt and making Δt go to zero we get the continuous-time observation process

$$d\xi_t = \left[A_0(t,\xi) + A_1(t,\xi)\theta\right] dt + B(t,\xi)\, dW_t \tag{5.12}$$

where $A_0(t,\xi), A_1(t,\xi), B(t,\xi)$ are functions of time and the history of the process ξ_t up to time t. From Equation (5.11) we get the filter

$$\begin{aligned} dm_t &= \frac{A_1 \gamma_t}{B^2}\left[d\xi_t - (A_0 + A_1 m_t)\, dt\right] \\ d\gamma_t &= -\frac{A_1^2 \gamma_t^2}{B^2} dt \end{aligned} \tag{5.13}$$

where, for simplicity, I omitted the arguments of the functions A_0, A_1, B.

To solve the filtering equations, start with the second line of Equation (5.13) and rewrite it in the form $\frac{d\gamma_t}{\gamma_t^2} = -\frac{A_1^2 dt}{B^2}$. Integrating both sides and using the initial condition $\gamma_0 = \sigma^2$ we get

$$\gamma_t = \frac{\sigma^2}{1+\sigma^2 \int_0^t \left(\frac{A_1}{B}\right)^2 du} \tag{5.14}$$

To continue the solution, use the two-dimensional Ito's formula to compute $d\frac{m_t}{\gamma_t}$

$$\begin{aligned} d\frac{m_t}{\gamma_t} &= \frac{1}{\gamma_t} dm_t - \frac{m_t}{\gamma_t^2} d\gamma_t \\ &= \frac{A_1}{B} d\overline{W}_t + \frac{m_t}{\gamma_t^2} \frac{A_1^2 \gamma_t^2}{B^2} dt \\ &= \frac{A_1}{B^2} (d\xi_t - A_0 dt) \end{aligned} \tag{5.15}$$

Integrating both sides, and using the initial conditions $m_0 = \mu$ and $\gamma_0 = \sigma^2$ yields

$$\frac{m_t}{\gamma_t} - \frac{\mu}{\sigma^2} = \int_0^t \frac{A_1}{B^2} (d\xi_u - A_0 du) \tag{5.16}$$

We get the solution of the filtering equations is

$$\begin{aligned} m_t &= \frac{\mu + \sigma^2 \int_0^t \frac{A_1}{B^2} (d\xi_u - A_0 du)}{1+\sigma^2 \int_0^t \left(\frac{A_1}{B}\right)^2 du} \\ \gamma_t &= \frac{\sigma^2}{1+\sigma^2 \int_0^t \left(\frac{A_1}{B}\right)^2 du} \end{aligned} \tag{5.17}$$

5.2.1 Relation to Bayesian Analysis and Conjugate Distributions

Equation (5.17) is an extension of the derivation of a formula for a posterior conjugate distribution when the prior and the likelihood are normal. Let the conditional distribution of data X given random variable θ be normal unknown mean θ and known variance B^2. Assume a normal prior for θ with known mean μ and known variance σ^2. Those parameters of the prior distribution are frequently called hyper-parameters. Our objective is to derive the posterior distribution of θ given an observation of a single value X. After that, we will extend the posterior distribution to an observation of a finite sample $X_1, \ldots, X_n$.

The random variables X and θ have a joint normal distribution given by $f(X,\theta) = f(X|\theta)f(\theta)$. We want to determine $f(\theta|X)$, and there is a shortcut in doing it. Recall that for joint normal distribution of X and θ, the conditional mean is

$$\begin{aligned} \mathbb{E}(X|\theta) &= \mathbb{E}(X) + \frac{\mathrm{cov}(X,\theta)}{\mathrm{var}(\theta)}\Big[\theta - \mathbb{E}(\theta)\Big] \\ &= \theta \end{aligned} \tag{5.18}$$

for every θ. Therefore, $\mathrm{cov}(X,\theta) = \mathrm{var}(\theta) = \sigma^2$. And $\mathbb{E}(X) = \mathbb{E}(\theta) = \mu$.

In addition,

$$\begin{aligned} \mathrm{var}(X|\theta) &= \mathrm{var}(X) - \frac{\mathrm{cov}^2(X,\theta)}{\mathrm{var}(\theta)} \\ &= \mathrm{var}(X) - \mathrm{var}(\theta) \end{aligned} \tag{5.19}$$

Therefore, $\mathrm{var}(X) = \mathrm{var}(X|\theta) + \mathrm{var}(\theta) = B^2 + \sigma^2$.

Applying those same two relationships, we can compute

$$\begin{aligned}
\mathbb{E}(\theta|X) &= \mathbb{E}(\theta) + \frac{\mathrm{cov}(X,\theta)}{\mathrm{var}(X)}\Big[X - \mathbb{E}(X)\Big] \\
&= \mu + \frac{\sigma^2}{B^2+\sigma^2}(X-\mu) \\
&= \frac{B^2}{B^2+\sigma^2}\mu + \frac{\sigma^2}{B^2+\sigma^2}X \\
&= \frac{B^2\mu+\sigma^2 X}{B^2+\sigma^2} \\
&= \frac{\dfrac{\mu}{\sigma^2}+\dfrac{X}{B^2}}{\dfrac{1}{\sigma^2}+\dfrac{1}{B^2}} \qquad (5.20)\\
\mathrm{var}(\theta|X) &= \mathrm{var}(\theta) - \frac{\mathrm{cov}^2(X,\theta)}{\mathrm{var}(X)} \\
&= \sigma^2 - \frac{\sigma^4}{B^2+\sigma^2} \\
&= \frac{B^2\sigma^2}{B^2+\sigma^2} \\
&= \frac{1}{\dfrac{1}{\sigma^2}+\dfrac{1}{B^2}}
\end{aligned}$$

We can use Equation (5.20) for a quick derivation of the posterior distribution of θ given observations $X_1, \ldots, X_n$. We know that $\overline{X}$ is a sufficient statistic for θ and that $\overline{X}|\theta \sim \mathbb{N}\left(\theta, \frac{B^2}{n}\right)$. Therefore,

$$
\begin{aligned}
\mathbb{E}(\theta|X_1,\ldots,X_n) &= \frac{\frac{\mu}{\sigma^2}+\frac{n\overline{X}}{B^2}}{\frac{1}{\sigma^2}+\frac{n}{B^2}} = \frac{\frac{\mu}{\sigma^2}+\frac{\sum_{i=1}^{n} X_i}{B^2}}{\frac{1}{\sigma^2}+\frac{n}{B^2}} \\
\mathrm{var}(\theta|X_1,\ldots,X_n) &= \frac{1}{\frac{1}{\sigma^2}+\frac{n}{B^2}}
\end{aligned} \tag{5.21}
$$

We will turn now to a heuristic derivation of Equation (5.17). To simplify things a bit, assume that $A_0 = 0$ and $A_1 = 1$, and replace dt by a finite (non-infinitesimal) time interval Δt. Also, assume that B is constant and write $\Delta\xi_t$ instead of $d\xi_t$ and ΔW_t instead of dW_t. Then, the observation equation, Equation (5.12), becomes

$$
\Delta\xi_t = \theta\Delta t + B\Delta W_t \tag{5.22}
$$

The conditional distribution of $\Delta\xi_t$ given $\theta\Delta t$ is normal with mean $\theta\Delta t$ and variance $B^2\Delta t$, and the prior distribution of $\theta\Delta t$ is normal with mean $\mu\Delta t$ and variance $\sigma^2(\Delta t)^2$.

Applying Equation (5.20) we get

$$
\begin{aligned}
\mathbb{E}(\theta\Delta t|\Delta\xi_1,\cdots,\Delta\xi_n) &= \frac{\dfrac{\mu\Delta t}{\sigma^2(\Delta t)^2}+\dfrac{\sum_{i=1}^{n}\Delta\xi_i}{B^2\Delta t}}{\dfrac{1}{\sigma^2(\Delta t)^2}+\dfrac{n}{B^2\Delta t}} = \frac{\dfrac{\mu}{\sigma^2}+\dfrac{\sum_{i=1}^{n}\Delta\xi_i}{B^2}}{\dfrac{1}{\sigma^2}+\dfrac{n\Delta t}{B^2}}\Delta t \\
&= \frac{\dfrac{\mu}{\sigma^2}+\dfrac{\xi_t-\xi_0}{B^2}}{\dfrac{1}{\sigma^2}+\dfrac{t}{B^2}}\Delta t \\
\mathrm{var}(\theta\Delta t|\Delta\xi_1,\ldots,\Delta\xi_n) &= \frac{1}{\dfrac{1}{\sigma^2(\Delta t)^2}+\dfrac{n}{B^2\Delta t}} = \frac{1}{\dfrac{1}{\sigma^2}+\dfrac{t}{B^2}}(\Delta t)^2
\end{aligned}
\tag{5.23}
$$

Therefore

$$
\begin{aligned}
\mathbb{E}(\theta|\Delta\xi_1,\ldots,\Delta\xi_n) &= \frac{\dfrac{\mu}{\sigma^2}+\dfrac{\xi_t-\xi_0}{B^2}}{\dfrac{1}{\sigma^2}+\dfrac{t}{B^2}} \\
\mathrm{var}(\theta|\Delta\xi_1,\ldots,\Delta\xi_n) &= \frac{1}{\dfrac{1}{\sigma^2}+\dfrac{t}{B^2}}
\end{aligned}
\tag{5.24}
$$

which is the same as Equation (5.17) when $A_0 = 0, A_1 = 1$ and B is constant.

5.3 Model of a Survival Time with Filtering

I will use this example to illustrate the computation of a conditional survival function is $A_0 = 0$, $A_1 = 1$, and $B = \frac{1}{\sqrt{2\delta}}\mathrm{e}^{-\delta t}$, we have then

$$
\begin{aligned}
\xi_t &= \xi_0 + \theta t + \frac{1}{\sqrt{2\delta}} \int_0^t e^{-\delta u}\, dW_u \\
m_t &= \frac{\mu + 2\delta\sigma^2 \int_0^t e^{2\delta u}\, d\xi_u}{1 + \sigma^2\left(e^{2\delta t} - 1\right)} \\
\gamma_t &= \frac{\sigma^2}{1 + \sigma^2\left(e^{2\delta t} - 1\right)}
\end{aligned}
\tag{5.25}
$$

If $\theta \big| \xi_0 \sim \mathbb{N}(0,1)$ then

$$
\begin{aligned}
m_t &= 2\delta \int_0^t e^{-2\delta(t-u)}\, d\xi_u \\
\gamma_t &= e^{-2\delta t}
\end{aligned}
\tag{5.26}
$$

For a numerical example choose $\delta = 0.02$ and generate a standard normal variable θ and an approximate Wiener process W_t with time increment $dt = 0.01$ and $n = 8000$ time steps that will cover the time interval from $t = 0$ to $t = 80$ years. Panel (A) of Figure 5.1 shows a simulated observation process ξ_t, Panel (B) shows the conditional mean m_t, and Panel (C) shows the square root of the conditional variance γ_t. The gray line in Panel (B) shows the generated value of the unobservable random variable θ.

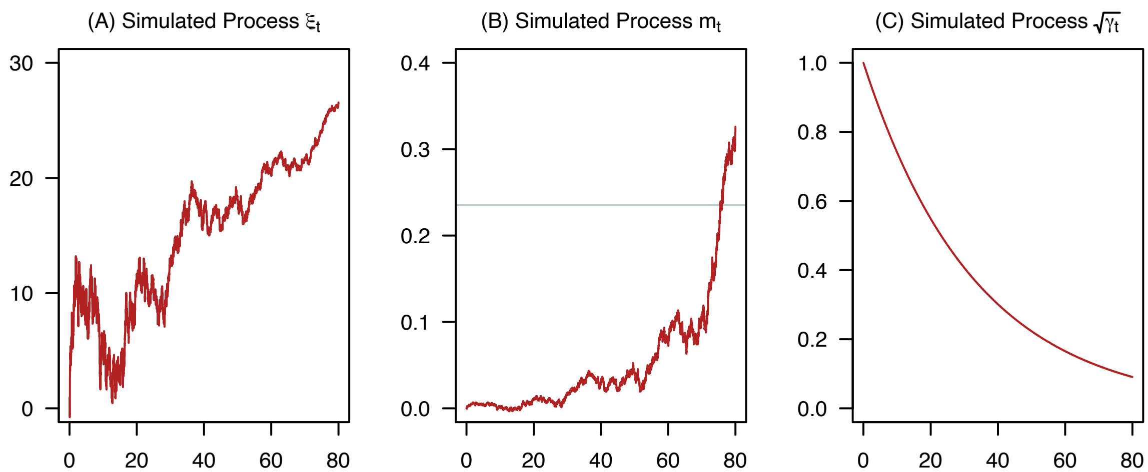

Figure 5.1. Observation process ξ_t and filter m_t, γ_t. The process ξ_t, the conditional mean m_t, and the conditional variance γ_t are generated by Equation (5.25) with parameter values $\mu = 0$, $\sigma = 1$ and $\delta = 0.02$. The realized value of θ is $\theta = 0.2352$.

A Weibull distribution function F with shape parameter $k = 5$ and scale parameter $\lambda = 85$ has mean $\lambda\Gamma\left(1+\frac{1}{k}\right) = 78.04$ and standard deviation
$\lambda\sqrt{\Gamma\left(1+\frac{2}{k}\right)-\left[\Gamma\left(1+\frac{1}{k}\right)\right]^2} = 17.88.$

The same transformation that I used in Chapter 3, $\tau = \psi(\theta) = F^{-1} \circ \Phi(\theta)$, where θ is a standard normal variable, defines a random time (survival time) with distribution function F.

The conditional distribution of the random time τ given the history $\mathcal{F}_t^{\xi}$ of the observation process ξ_t is

$$\begin{aligned}\mathbb{P}\left(\tau \le y \middle| \mathcal{F}_t^{\xi}\right) &= \mathbb{P}\left[\psi(\theta) \le y \middle| \mathcal{F}_t^{\xi}\right] \\ &= \mathbb{P}\left[\theta \le \psi^{-1}(y) \middle| \mathcal{F}_t^{\xi}\right] \\ &= \Phi\left[\frac{\psi^{-1}(y) - m_t}{\sqrt{\gamma_t}}\right]\end{aligned} \tag{5.27}$$

The corresponding conditional survival function is

$$\mathbb{P}\left(\tau > y \middle| \mathcal{F}_t^{\xi}\right) = \Phi\left[\frac{m_t - \psi^{-1}(y)}{\sqrt{\gamma_t}}\right] \tag{5.28}$$

At age t, the conditional survival function given survival to time t and the history of the observation process ξ_t is

$$S\left(y \middle| \mathcal{F}_t^{\xi}, \tau > t\right) = \mathbb{P}\left(\tau > y \middle| \mathcal{F}_t^{\xi}, \tau > t\right) = \frac{\Phi\left[\dfrac{m_t - \psi^{-1}(y)}{\sqrt{\gamma_t}}\right]}{\Phi\left[\dfrac{m_t - \psi^{-1}(t)}{\sqrt{\gamma_t}}\right]} \quad \text{for} \quad y \ge t \tag{5.29}$$

In this simulation the random variable θ has the realization $\theta = 0.2352$ which gives the realization of the random time $\tau = \psi(0.2352) = 83.21$,

For illustration, consider two ages, $t = 40$ and $t = 60$. The conditional mean m_t at age $t = 40$ is $m_{40} = 0.0317$, and conditional standard deviation $\sqrt{\gamma_t}$ at age $t = 40$ is $\sqrt{\gamma_{40}} = 0.3012$. The conditional mean m_t at age $t = 60$ is $m_{60} = 0.0746$, and conditional standard deviation $\sqrt{\gamma_t}$ at age $t = 60$ is $\sqrt{\gamma_{60}} = 0.1653$.

Figure 5.2 shows the conditional survival functions at age $t = 40$ and $t = 60$ obtained by substituting those realized values of m_t and γ_t into Equation (5.29).

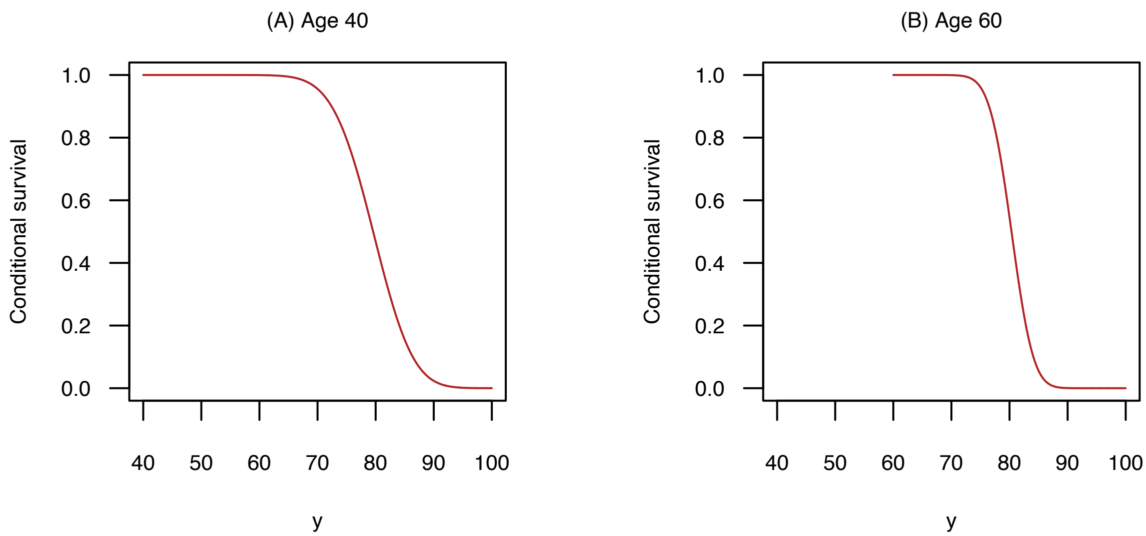

Figure 5.2. Conditional survival function at age 40 and age 60. The conditional survival functions were computed using Equation (5.29) where m_t and γ_t come from Equation (5.26) with parameter value $\delta = 0.02$.

5.4 Model Modification to Allow Fluctuations Around True Value

Consider again the model in Section 4.6 with the extra condition that the health marker can be observed only with a measurement error, or alternatively, that the process that generates the survival time is the true medical condition, and the health marker is a noisy observation of the true medical condition.

Accordingly, specify the true medical condition an an unobservable process θ_t, and the health marker as an observable process ξ_t, as follows

$$
\begin{aligned}
d\theta_t &= -\frac{1}{2}b^2 \mathrm{e}^{-2\delta t}\,dt + b\,\mathrm{e}^{-\delta t}\,dW_{1,t} \\
d\xi_t &= \kappa(\theta_t - \xi_t)\,dt + B\,dW_{2,t}
\end{aligned} \tag{5.30}
$$

There is a conditional prior on θ_0 given $\mathcal{F}_0^\xi$ given by $\mathbb{N}\left(\mu, \sigma^2\right)$

$$
\mathbb{P}\left(\theta_0 \leq r \middle| \mathcal{F}_0^\xi\right) = \Phi\left(\frac{r-\mu}{\sigma}\right) \tag{5.31}
$$

The filtering equations are

$$
\begin{aligned}
dm_t &= -\frac{1}{2}b^2 \mathrm{e}^{-2\delta t}\,dt + \frac{\kappa\gamma_t}{B^2}\left[d\xi_t - \kappa(m_t - \xi_t)\,dt\right] \\
d\gamma_t &= \left(b^2 \mathrm{e}^{-2\delta t} - \frac{\kappa^2\gamma_t^2}{B^2}\right)dt
\end{aligned} \tag{5.32}
$$

with the initial conditions $m_0 = m$ and $\gamma_0 = \sigma^2$; The unobservable process θ_t converges to a limiting random variable θ, whose conditional distribution given $\mathcal{F}_0^\xi$ is

$$
\theta \sim \mathbb{N}\left(m - \frac{b^2}{4\delta}, \frac{b^2}{2\delta}\right) \tag{5.33}
$$

The random survival time τ is defined

$$
\tau = \psi(\theta) = F^{-1} \circ \Phi\left[\frac{\theta - \left(m - \frac{b^2}{4\delta}\right)}{b}\sqrt{2\delta}\right] \tag{5.34}
$$

That gives survival time τ conditional distribution F given $\mathcal{F}_0^\xi$. The next task is to compute the conditional distribution of τ given $\mathcal{F}_t^\xi$ for $t > 0$. We have

$$
\begin{aligned}
\mathbb{P}\left(\tau \leq r \middle| \mathcal{F}_t^{\xi}\right) &= \mathbb{P}\left[\psi(\theta) \leq r \middle| \mathcal{F}_t^{\xi}\right] \\
&= \mathbb{P}\left[\theta \leq \psi^{-1}(r) \middle| \mathcal{F}_t^{\xi}\right]
\end{aligned} \tag{5.35}
$$

The conditional distribution of θ given $\mathcal{F}_t^{\xi}$ is normal with moments that can be computed from m_t and γ_t.

The model can be easily extended to a multidimensional observable process $\xi_t = (\xi_{1,t}, \ldots, \xi_{n,t})'$.

To solve the filtering model in Equation (5.32) denote by f_t a non-vanishing, continuously differentiable deterministic function of time, and apply Ito's formula to the ratio $\dfrac{m_t}{f_t}$

$$
\begin{aligned}
d\left(\frac{m_t}{f_t}\right) &= \frac{dm_t}{f_t} - \frac{m_t df_t}{f_t^2} \\
&= \frac{1}{f_t}\left\{-\frac{1}{2}\eta_t dt + \frac{\kappa\gamma_t}{B^2}\left[d\xi_t - \kappa(m_t - \xi_t)\,dt\right]\right\} - \frac{m_t df_t}{f_t^2} \\
&= \frac{1}{f_t}\left[-\frac{1}{2}\eta_t dt + \frac{\kappa\gamma_t}{B^2}(d\xi_t + \kappa\xi_t dt)\right] - \left(\frac{\kappa^2\gamma_t}{B^2}dt + \frac{df_t}{f_t}\right)\frac{m_t}{f_t}
\end{aligned} \tag{5.36}
$$

where $\eta_t = b^2\,\mathrm{e}^{-2\delta t}$. If we can choose f_t to be such that $\dfrac{\kappa^2\gamma_t}{B^2}dt + \dfrac{df_t}{f_t} = 0$ then

$$
d\left(\frac{m_t}{f_t}\right) = \frac{1}{f_t}\left[-\frac{1}{2}\eta_t dt + \frac{\kappa\gamma_t}{B^2}(d\xi_t + \kappa\xi_t dt)\right] \tag{5.37}
$$

Integrating both sides and rearranging terms

$$
\frac{m_t}{f_t} = \frac{m_0}{f_0} - \frac{1}{2}\int_0^t \frac{\eta_u}{f_u}du + \frac{\kappa}{B^2}\int_0^t \frac{\gamma_u}{f_u}(d\xi_u + \kappa\xi_u du) \tag{5.38}
$$

To find the function f_t write

$$\frac{d\log f_t}{dt} = -\frac{\kappa^2\gamma_t}{B^2} \tag{5.39}$$

Combining Equation (5.39) with the second line of Equation (5.32) gives

$$\frac{d\log f_t}{dt} = \frac{d\log\gamma_t}{dt} - \frac{\eta_t}{\gamma_t} \tag{5.40}$$

Integrating both sides and rearranging terms

$$\log\frac{f_t}{\gamma_t} - \log\frac{f_0}{\gamma_0} = -\int_0^t \frac{\eta_u}{\gamma_u}du \tag{5.41}$$

Choosing $f_0 = \gamma_0$ we get

$$f_t = \gamma_t \exp\left(-\int_0^t \frac{\eta_u}{\gamma_u}du\right) \tag{5.42}$$

Substitution into Equation (5.43) yields

$$\begin{aligned} m_t &= \gamma_t \exp\left(-\int_0^t \frac{\eta_u}{\gamma_u}du\right)\left[\frac{m}{\sigma^2} - \frac{1}{2}\int_0^t \frac{\eta_u}{\gamma_u}\exp\left(\int_0^u \frac{\eta_s}{\gamma_s}ds\right)du\right. \\ &+ \left.\frac{\kappa}{B^2}\int_0^t \exp\left(\int_0^u \frac{\eta_s}{\gamma_s}ds\right)(d\xi_u + \kappa\xi_u du)\right] \end{aligned} \tag{5.43}$$

The function γ_t satisfies a Riccati ordinary differential equation which I solve in Section 5.5.

5.5 Ornstein-Uhlenbeck Observation Process

Consider a model of an unobservable Ito process θ_t

$$\theta_t = \theta_0 + \int_0^t b_1(u)dW_{1,u} \tag{5.44}$$

and an observable Ito process ξ_t

$$\xi_t = \xi_0 + \int_0^t \left[A_0(u) + A_1(u)\theta_t\right]du + \int_0^t B_2(u)dW_{2,u} \tag{5.45}$$

Assume that the function $b_1(t)$ is such that $\int_0^\infty b_1^2(t)dt < \infty$. Let θ be the limit of the process θ_t when t goes to infinity, and define the random time $\tau = \psi(\theta)$. This allows us to combine the model introduced in Chapter 3 with a filtering model introduced in this chapter. The difference between this formulation and the models in the preceding two examples is that now we observe the process θ_t with an error, rather than observe the random variable θ with an error.

Denote $m_t = \mathbb{E}\left(\theta_t \middle| \mathcal{F}_t^\xi\right)$ and $\gamma_t = \text{var}\left(\theta_t \middle| \mathcal{F}_t^\xi\right)$. Again, the difference is that now m_t and γ_t are conditional moments of θ_t rather than θ.

As I will describe in Section 5.6.2, the filtering equations in this case are

$$\begin{aligned} dm_t &= \frac{A_1(t)\gamma_t}{B_2^2(t)}\left\{d\xi_t - \left[A_0(t) + A_1(t)m_t\right]dt\right\} \\ \frac{d\gamma_t}{dt} &= b_1^2(t) - \frac{A_1^2(t)}{B_2^2(t)}\gamma_t^2 \end{aligned} \tag{5.46}$$

It is worth noting that the difference between Equation (5.46), where θ_t is an Ito process, and Equation (5.13), where θ is a random variable, is the term $b_1^2(t)$ accounting for the dynamic nature of θ_t.

The conditional mean $\mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ is equal to m_t because

$$\begin{aligned}
\mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right) &= \mathbb{E}\left(\theta_t \middle| \mathcal{F}_t^{\xi}\right) + \mathbb{E}\left(\int_t^{\infty} b_1(u) dW_u \middle| \mathcal{F}_t^{\xi}\right) \\
&= m_t + \mathbb{E}\left(\int_t^{\infty} b_1(u) dW_u\right) \\
&= m_t
\end{aligned} \tag{5.47}$$

The conditional variance γ_t is a solution of a Riccati ordinary differential equation in the second line of Equation (5.46). I will show the solution of this Riccati equation for the case when $b_1(t) = b\exp(-ht)$ where b, h and B_2 are positive constants and A_1 is a constant.

Let

$$\begin{aligned}
P(t) &= b_1^2(t) \\
R &= -\frac{A_1^2}{B_2^2}
\end{aligned} \tag{5.48}$$

Then, the Riccati equation becomes $\gamma_t' = P(t) + R\gamma_t^2$. It is known that the substitution $\gamma_t = -\frac{\delta_t'}{R\delta_t}$ transforms a Riccati equation into a second order linear ordinary differential equation

$$\delta_t'' + P(t)R\delta_t = 0 \tag{5.49}$$

Denote $s = k\exp(-ht)$ where k is a positive constant to be specified later, and $\eta_s = \delta_t$. Then $\dfrac{ds}{dt} = -hs$ and

$$\begin{aligned}\delta_t' &= -hs\eta_s' \\ \delta_t'' &= \Big(-h\eta_s' - hs\eta_s''\Big)(-hs) \\ &= h^2 s\eta_s' + h^2 s^2 \eta_s''\end{aligned} \tag{5.50}$$

Denote $g = \dfrac{A_1^2}{B_2^2} b^2$, choose $k = \dfrac{\sqrt{g}}{h}$, then $g = k^2h^2$ and $g\exp(-2ht) = h^2k^2\exp(-2ht) = h^2s^2$, and we get

$$s^2\eta_s'' + s\eta_s' - s^2\eta_s = 0 \tag{5.51}$$

Equation (5.51) is a second-order linear differential equation called modified Bessel equation of order zero, with two linearly independent solutions $I_0(s)$ and $K_0(s)$, called, respectively, modified Bessel function of the first kind and modified Bessel function of the second kind.

Accounting for the substitution $s = \dfrac{\sqrt{g}}{h}\exp(-ht)$ and $\delta_t = \eta_s$ we get that the general solution of the differential equation for δ_t is

$$\delta_t = C_1 I_0\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right] + C_2 K_0\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right] \tag{5.52}$$

Making use of the formulas

$$\begin{aligned}I_0'(x) &= I_1(x) \\ K_0'(x) &= -K_1(x)\end{aligned} \tag{5.53}$$

the corresponding derivative is

$$\delta_t' = -C_1 \frac{A_1 b}{B_2} \exp(-ht) I_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + C_2 \frac{A_1 b}{B_2} \exp(-ht) K_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] \tag{5.54}$$

Reversing another substitution we have

$$\begin{aligned}
\gamma_t &= \frac{B_2^2}{A_1^2} \frac{\delta_t'}{\delta_t} \\
&= \frac{B_2^2}{A_1^2} \frac{-C_1 \frac{A_1 b}{B_2} \exp(-ht) I_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + C_2 \frac{A_1 b}{B_2} \exp(-ht) K_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]}{C_1 I_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + C_2 K_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]} \\
&= \frac{B_2 b}{A_1} \frac{-C_1 I_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + C_2 K_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]}{C_1 I_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + C_2 K_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]} \exp(-ht) \\
&= \frac{B_2 b}{A_1} \frac{-C I_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + K_1 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]}{C I_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right] + K_0 \left[\frac{A_1 b}{B_2 h} \exp(-ht) \right]} \exp(-ht)
\end{aligned} \tag{5.55}$$

To determine the constant C, assume that $\gamma_0 = \mathrm{var}\left(\theta_0 \middle| \mathcal{F}_0^{\xi}\right)$ is known. Then we can write

$$\gamma_0 = \frac{B_2 b}{A_1} \frac{-C I_1 \left(\frac{A_1 b}{B_2 h} \right) + K_1 \left(\frac{A_1 b}{B_2 h} \right)}{C I_0 \left(\frac{A_1 b}{B_2 h} \right) + K_0 \left(\frac{A_1 b}{B_2 h} \right)} \tag{5.56}$$

It follows that

$$C = \frac{K_1\left(\frac{A_1 b}{B_2 h}\right) - \frac{A_1\gamma_0}{B_2 b} K_0\left(\frac{A_1 b}{B_2 h}\right)}{I_1\left(\frac{A_1 b}{B_2 h}\right) + \frac{A_1\gamma_0}{B_2 b} I_0\left(\frac{A_1 b}{B_2 h}\right)} \tag{5.57}$$

Next, $\theta = \theta_t + \int_t^\infty b_1(u) dW_{1,u}$, and therefore

$$\begin{aligned} \mathrm{var}\left(\theta \middle| \mathcal{F}_t^\xi\right) &= \mathrm{var}\left(\theta_t \middle| \mathcal{F}_t^\xi\right) + \mathrm{var}\left[\int_t^\infty b_1(u) dW_{1,u}\right] \\ &= \gamma_t + \int_t^\infty b_1^2(u) du \end{aligned} \tag{5.58}$$

It follows that

$$\begin{aligned} \mathrm{var}\left(\theta \middle| \mathcal{F}_t^\xi\right) &= \frac{Bb}{A_1} \frac{-CI_1\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right] + K_1\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right]}{CI_0\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right] + K_0\left[\frac{A_1 b}{B_2 h}\exp(-ht)\right]} \exp(-ht) \\ &+ \frac{b^2 \exp(-2ht)}{2h} \end{aligned} \tag{5.59}$$

The conditional distribution of θ given $\mathcal{F}_t^\xi$ is normal with mean m_t and variance given in Equation (5.59).

Figure 5.3 shows the evolution of the conditional variance γ_t of the stochastic process θ_t and conditional variance $\text{var}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ of the random variable θ.

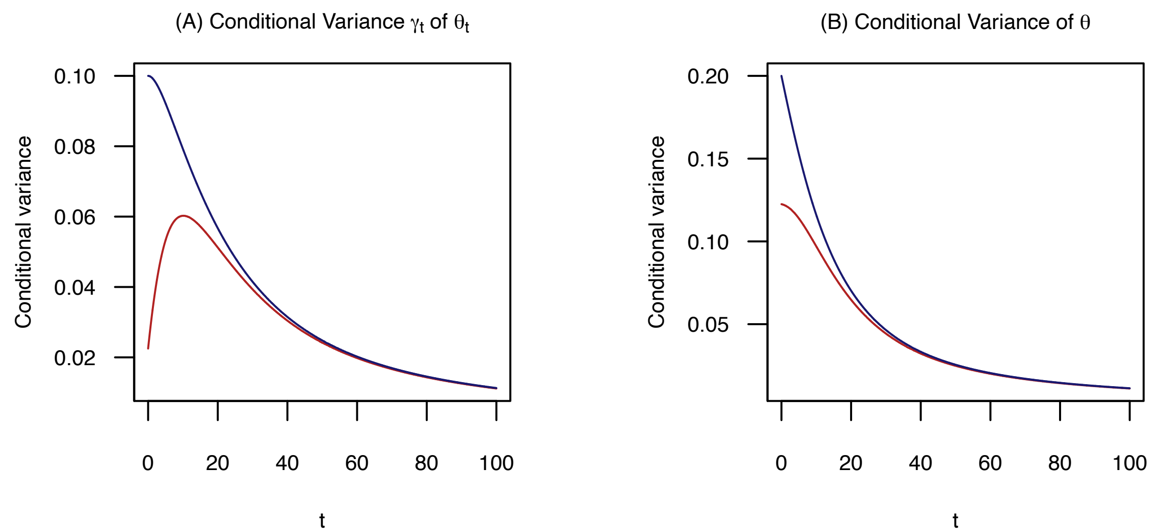

Figure 5.3. Panel (A) shows conditional variance γ_t of the stochastic process θ_t and Panel (B) shows conditional variance $\text{var}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ of the random variable θ. In both panels, the blue line corresponds to $\gamma_0 = 0.1$ and the red line corresponds to $\gamma_0 = 0.0225$. The value of the other parameters are $b = 0.1$, $h = 0.05$, $A_1 = 1$, and $B_2 = 1$.

To interpret the plot of γ_t when $\gamma_0 = 0.0225$, consider a simpler case of Equation (5.46) where $b_1(t) = 1$, $A_0(t) = 0$, $A_1(t) = 1$, and $B_2(t) = 1$. In that case, the differential equation for conditional variance γ_t is $\gamma_t' = 1 - \gamma_t^2$. The solution depends on whether the initial value γ_0 is less than 1 or greater than 1. If $\gamma_0 < 1$, then the solution is $\gamma_t = \tanh(t + \operatorname{atanh}\gamma_0)$, which is an increasing function. If $\gamma_0 > 1$, then the solution is $\gamma_t = \dfrac{\gamma_0 + 1 + (\gamma_0 - 1)\exp(-2t)}{\gamma_0 + 1 - (\gamma_0 - 1)\exp(-2t)}$, which is a decreasing function. Allowing for the more complicated situation presented in Figure 5.3, this helps explain why the conditional variance γ_t initially increases and then decreases for a small initial value of γ_0.

As a special case of the preceding discussion, consider an Ornstein-Uhlenbeck observation process

$$d\xi_t = \kappa(\theta_t - \xi_t)\,dt + B_2\,dW_{2,t} \tag{5.60}$$

where κ and B_2 are positive constants. In this case $A_0 = -\kappa\xi_t$ and $A_1 = \kappa$. The observation process ξ_t fluctuates around the unobservable process θ_t.

The first line of Equation (5.46) becomes

$$dm_t = \frac{\kappa\gamma_t}{B_2^2}\Big[d\xi_t - \kappa(m_t - \xi_t)dt\Big] \tag{5.61}$$

The only change in the second line of Equation (5.46) is $A_1 = \kappa$.

Equation (5.61) is recursive, and we can solve the corresponding finite-difference equation numerically using the values of γ_t from Equation (5.55). The finite-difference equation is

$$m_t = \left(1 - \frac{\kappa^2\gamma_{t-\Delta t}\Delta t}{B_2^2}\right)m_{t-\Delta t} - \frac{\kappa\gamma_{t-\Delta t}(1-\kappa\Delta t)}{B_2^2}\xi_{t-\Delta t} + \frac{\kappa\gamma_{t-\Delta t}}{B_2^2}\xi_t \tag{5.62}$$

Figure 5.4 shows a simulated path of the unobservable process θ_t and the corresponding biomarker ξ_t, and the conditional mean process m_t of the unobservable process given the observed histoary of the biomarker. Panel (A) illustrates the convergence of the unobservable process to a random variable, and the mean-reverting nature of the biomarker around the unobservable process. Panel (B) shows that the quality of estimation is good. The conditional mean m_t of θ_t was computed using the finite-difference Equation (5.62).

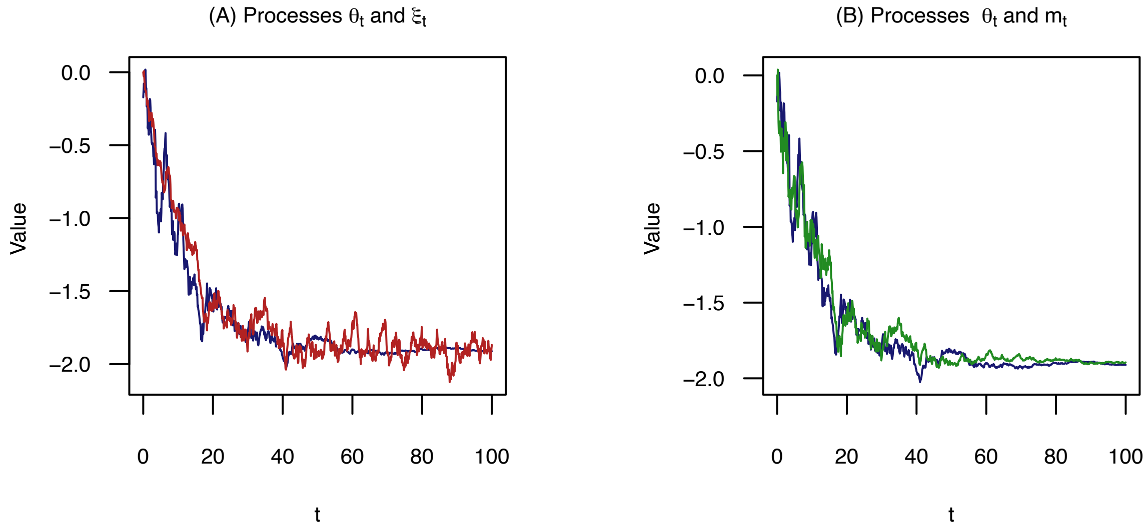

Figure 5.4. Panel (A) shows a simulated path of the unobservable process θ_t (blue line) and the corresponding path of the biomarker ξ_t (red line). Panel (B) shows a simulated path of the unobservable process θ_t (blue line) and the conditional mean m_t of the unobservable process given the history of the biomarker (green line). Parameter values are: $h = 0.05$, $b = \sqrt{2h}$, $\kappa = 0.5$, $B_2 = 0.1$, $\xi_0 = 0$, $m_0 = 0$, $\gamma_0 = 0.1$, $\theta_0 \sim \mathbb{N}(m_0, \gamma_0)$.

5.6 A General Filtering Framework

This section describes a rigorous justification for the filtering equations which I described heuristically, it follows the approach of Liptser and Shiryaev [27].

5.6.1 Optimal Filtering Equation

Let $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$ be a filtered probability space. We will use the generic notation $(g_t, \mathcal{F}_t)$ for a stochastic process g_t that is adapted to the filtration $\mathcal{F}_t$.

In this setup there are three stochastic processes:

1. An unobservable process $(\theta_t, \mathcal{F}_t)$.

2. An observable process $(\xi_t, \mathcal{F}_t)$.

3. An estimated process $(h_t, \mathcal{F}_t)$.

We assume that the observable process $(\xi_t, \mathcal{F}_t)$ is an Ito process

$$\xi_t = \xi_0 + \int_0^t A_u du + \int_0^t B_u dW_u \tag{5.63}$$

where $(A_t, \mathcal{F}_t)$ is integrable and the integrand in the Ito integral is $\left(B_t, \mathcal{F}_t^{\xi}\right)$. Stated differently, cumulative observation noise is conditionally normal, given the history of the process ξ_t.

We also assume that $(h_t, \mathcal{F}_t)$ has the representation

$$h_t = h_0 + \int_0^t H_u du + X_t \tag{5.64}$$

where $(H_t, \mathcal{F}_t)$ is integrable and $(X_t, \mathcal{F}_t)$ is a square-integrable martingale.

Introduce the following additional notation:

1. The predictable quadratic covariation process $\langle X, W \rangle_t$ such that the process $X_t W_t - \langle X, W \rangle_t$ is an $\mathcal{F}_t$-martingale.

2. The process $(D_t, \mathcal{F}_t)$ such that

$$D_t = \frac{d\langle X, W \rangle_t}{dt} \tag{5.65}$$

3. For any integrable process $(g_t, \mathcal{F}_t)$ denote

$$\pi_t(g) = \mathbb{E}\left(g_t \middle| \mathcal{F}_t^{\xi}\right) \tag{5.66}$$

Then we have the following representation of the estimate $\pi_t(h)$

$$d\pi_t(h) = \pi_t(H)\,dt + \frac{\pi_t(D)\,B_t + \pi_t(hA) - \pi_t(h)\pi_t(A)}{B_t^2}\left[d\xi_t - \pi_t(A)\,dt\right] \tag{5.67}$$

Furthermore, define

$$\overline{W}_t = \int_0^t \frac{d\xi_u - \pi_u(A)\,du}{B_u} \tag{5.68}$$

The process $\left(\overline{W}_t, \mathcal{F}_t^{\xi}\right)$ is a Wiener process and for every $t \geq 0$ we have $\mathcal{F}_t^{\overline{W}} = \mathcal{F}_t^{\xi}$, that is, the Wiener process $\overline{W}_t$ carries the same information as the observation process ξ_t. The Wiener process $\overline{W}_t$ is called an innovation process.

We can write the representation of the estimate $\pi_t(h)$ in the form

$$d\pi_t(h) = \pi_t(H)\,dt + \frac{\pi_t(D)\,B_t + \pi_t(hA) - \pi_t(h)\pi_t(A)}{B_t}\,d\overline{W}_t \tag{5.69}$$

5.6.2 Application of the General Filtering Framework

Let the unobservable process θ_t be the Ito process

$$d\theta_t = a_1(t)\theta_t dt + b_1(t)\,dW_{1,t} + b_2(t)\,dW_{2,t} \tag{5.70}$$

and let the observable process be

$$d\xi_t = A_1(t)\theta_t dt + B_2(t)\,dW_{2,t} \tag{5.71}$$

By assumption, the Wiener processes $W_{1,t}$ and $W_{2,t}$ are independent. The preceding formulation allows the noise in the unobservable process θ_t to be correlated with observation noise.

The first part of the filtering problem is the calculation of $m_t = \mathbb{E}\left(\theta_t \middle| \mathcal{F}_t^{\xi}\right)$.

Set $h_t = \theta_t$, then

$$\begin{aligned}
A_t &= A_1(t)\theta_t \\
B_t &= B_1(t) \\
W_t &= W_{2,t} \\
H_t &= a_1(t)\theta_t \\
X_t &= \int_0^t b_1(u)\,dW_{1,u} + \int_0^t b_2(u)\,W_{2,u}
\end{aligned} \tag{5.72}$$

We know that the process

$$\left[\int_0^t b_1(u)\,dW_{1,u} + \int_0^t b_2(u)\,dW_{2,u}\right] W_{2,u} - \int_0^t b_2(u)\,du \tag{5.73}$$

is an $\mathcal{F}_t$-martingale, therefore

$$\langle X, W \rangle_t = \int_0^t b_2\,du \tag{5.74}$$

and $D_t = b_2(t)$. Denote

$$\begin{aligned} m_t &= \pi_t(h) \\ &= \pi_t(\theta) \\ &= \mathbb{E}\left(\theta_t \middle| \mathcal{F}_t^{\xi}\right) \\ \gamma_t &= \mathbb{E}\left[(\theta_t - m_t)^2 \middle| \mathcal{F}_t^{\xi}\right] \\ &= \mathbb{E}\left(\theta_t^2 \middle| \mathcal{F}_t^{\xi}\right) - m_t^2 \end{aligned} \tag{5.75}$$

Then

$$\begin{aligned} \pi_t(h) &= m_t \\ \pi_t(A) &= A_1(t)\, m_t \\ \pi_t(hA) &= \pi_t\left(A_1\theta^2\right) \\ &= A_1(t)\,\mathbb{E}\left(\theta_t^2 \middle| \mathcal{F}_t^{\xi}\right) \\ &= A_1(t)\left(\gamma_t + m_t^2\right) \\ \pi_t(H) &= a_1(t)\, m_t \\ \pi_t(D) &= b_2(t) \end{aligned} \tag{5.76}$$

From Equation (5.67) we get the filtering equation for the conditional mean m_t

$$\begin{aligned} dm_t &= a_1(t)\, m_t dt + \frac{b_2(t) B_2(t) + A_1(t)\left(\gamma_t + m_t^2\right) - m_t A_1(t)\, m_t}{B_2^2(t)} \left[d\xi_t - A_1(t)\, m_t dt\right] \\ &= a_1(t)\, m_t dt + \frac{b_2(t) B_2(t) + A_1(t)\gamma_t}{B_2^2(t)} \left[d\xi_t - A_1(t)\, m_t dt\right] \end{aligned} \tag{5.77}$$

We can also write the filtering equation for the conditional mean in terms of the innovation process $\overline{W}_t$

$$dm_t \quad = \quad a_1(t)\, m_t dt + \frac{b_2(t)\, B_2(t) + A_1(t)\,\gamma_t}{B_2(t)} d\overline{W}_t \tag{5.78}$$

If we modify the stochastic differential representation of the unobservable process θ_t in Equation (5.70) by adding a free term $a_0(t)$ to the drift, and similarly modify the drift of the observation equation in Equation (5.71) by adding to the drift a free term $A_0(t)$, the first filtering equation becomes

$$dm_t \quad = \quad \Big[a_0(t) + a_1(t)\, m_t\Big] dt + \frac{b_2(t)\, B_2(t) + A_1(t)\,\gamma_t}{B_2^2(t)} \left\{ d\xi_t - \Big[A_0(t) + A_1(t)\, m_t\Big] dt \right\} \tag{5.79}$$

The second part of the filtering problem is the calculation of $\pi_t\left(\theta^2\right)$. Apply Ito's formula to the function $f(\theta) = \theta^2$ and the representation of θ_t in Equation (5.70) to get

$$d\theta_t^2 \quad = \quad \Big[2a_1(t)\,\theta_t^2 + b_1^2(t) + b_2^2(t)\Big] dt + 2b_1(t)\,\theta_t dW_{1,t} + 2b_2(t)\,\theta_t dW_{2,t} \tag{5.80}$$

Set $h_t = \theta^2$, then

$$
\begin{aligned}
A_t &= A_1(t)\theta_t \\
B_t &= B_1(t) \\
W_t &= W_{2,t} \\
H_t &= 2a_1(t)\theta_t^2 + b_1^2(t) + b_2^2(t) \\
X_t &= 2\int_0^t b_1(u)\theta_u dW_{1,u} + 2\int_0^t b_2(u)\theta_u W_{2,u}
\end{aligned}
\tag{5.81}
$$

We now get $D_t = 2b_2(t)\theta_t$. Furthermore, now

$$
\begin{aligned}
\pi_t(h) &= \gamma_t + m_t^2 \\
\pi_t(A) &= A_1(t)m_t \\
\pi_t(hA) &= A_1(t)\pi_t\left(\theta^3\right) \\
\pi_t(H) &= 2a_1(t)\left(\gamma_t + m_t^2\right) + b_1^2(t) + b_2^2(t) \\
\pi_t(D) &= 2b_2(t)m_t
\end{aligned}
\tag{5.82}
$$

Substituting Equation (5.82) into Equation (5.67) delivers

$$
\begin{aligned}
d\pi_t(h) &= \left[2a_1(t)\left(\gamma_t + m_t^2\right) + b_1^2(t) + b_2^2(t)\right]dt \\
&+ \frac{2b_2(t)B_2(t)m_t + \pi_t(hA) - \pi_t(h)\pi_t(A)}{B_2^2(t)}\left[d\xi_t - A_1(t)m_t dt\right]
\end{aligned}
\tag{5.83}
$$

Continuing to transform Equation (5.83) write

$$\begin{aligned}\pi_t(hA) - \pi_t(h)\pi_t(A) &= A_1(t)\pi_t\left(\theta^3\right) - A_1(t)m_t\left(\gamma_t + m_t^2\right) \\ &= A_1(t)\left[\pi_t\left(\theta^3\right) - m_t\left(\gamma_t + m_t^2\right)\right] \\ &= 2A_1(t)m_t\gamma_t\end{aligned} \tag{5.84}$$

The last line of Equation (5.84) follows from the calculation of the third moment of a normal distribution.

Differentiate the moment generating function $M(r) = \exp\left(\mu + \frac{1}{2}\sigma^2 r^2\right)$ of $\mathbb{N}\left(\mu, \sigma^2\right)$ to get $M'''(r) = \left[3\sigma^2 + \left(\mu + \sigma^2 r\right)^2\right]\left(\mu + \sigma^2 r\right)\exp\left(\mu + \frac{1}{2}\sigma^2 r^2\right)$, therefore, the third moment is $M'''(0) = \mu\left(3\sigma^2 + \mu^2\right)$, which we can rewrite $M'''(0) = \mu\left(\sigma^2 + \mu^2\right) + 2\mu\sigma^2$, and it follows that $\pi_t\left(\theta^3\right) = m_t\left(\gamma_t + m_t^2\right) + 2m_t\gamma_t$.

Recalling that $\pi_t(h) = \gamma_t + m_t^2$, and using the form with the innovation process, Equation (5.83) becomes

$$\begin{aligned}d\left(\gamma_t + m_t^2\right) &= \left[2a_1(t)\left(\gamma_t + m_t^2\right) + b_1^2(t) + b_2^2(t)\right]dt \\ &+ \frac{2b_2(t)B_2(t)m_t + 2A_1(t)m_t\gamma_t}{B_2(t)}\,d\overline{W}_t\end{aligned} \tag{5.85}$$

Applying Ito's formula to the function $f(m) = m^2$ and the representation of m_t in Equation (5.78)

$$\begin{aligned}dm_t^2 &= \left\{2a_1(t)m_t^2 + \left[\frac{b_2(t)B_2(t) + A_1(t)\gamma_t}{B_2(t)}\right]^2\right\}dt \\ &+ 2\frac{b_2(t)B_2(t) + A_1(t)\gamma_t}{B_2(t)}m_t d\overline{W}_t\end{aligned} \tag{5.86}$$

Subtracting Equation (5.86) from Equation (5.85) we get the filtering equation for the conditional variance γ_t

$$d\gamma_t = \left\{ 2a_1(t)\gamma_t + b_1^2(t) + b_2^2(t) - \left[\frac{b_2(t)B_2(t) + A_1(t)\gamma_t}{B_2(t)} \right]^2 \right\} dt \qquad (5.87)$$

6 Multivariate Filtering Models

6.1 General Setting

In this section we consider models with multi-dimensional observation process and either multi-dimensional or one-dimensional unobservable random variable or stochastic process.

When the unobservable is a multi-dimensional random variable, the model is suitable for the analysis of competing risks. When the unobservable is one-dimensional, the model is an alternative to, or an improvement on, the Cox proportional hazards model.

More specifically, consider a m-dimensional unobservable process $\theta_t = \begin{pmatrix} \theta_{1,t} \\ \vdots \\ \theta_{m,t} \end{pmatrix}$ and a n-dimensional observable process $\xi_t = \begin{pmatrix} \xi_{1,t} \\ \vdots \\ \xi_{n,t} \end{pmatrix}$ that are solutions of the stochastic differential equations

$$\begin{aligned} d\theta_t &= (a_0 + a_1\theta_t)\,dt + b_1\,dW_{1,t} + b_2\,dW_{2,t} \\ d\xi_t &= (A_0 + A_1\theta_t)\,dt + B_1\,dW_{1,t} + B_2\,dW_{2,t} \end{aligned} \tag{6.1}$$

In Equation (6.1) $W_{1,t}$ is a m-dimensional Wiener process, $W_{2,t}$ is a n-dimensional Wiener process, the processes $W_{1,t}$ and $W_{2,t}$ are independent, a_0 is a m-dimensional vector, a_1 is a $m \times m$ matrix, b_1 is a $m \times m$ matrix, b_2 is a $m \times n$ matrix, A_0 is a n-dimensional vector, A_1 is a $n \times m$ matrix, B_1 is a $n \times m$ matrix, and B_2 is a $n \times n$ matrix.

All the proposed models of the conditional distribution of a random time are of this general form.

The filtering equations for this general setup are

$$
\begin{aligned}
dm_t &= (a_0 + a_1 m_t)\,dt + \left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)\left(B_1 B_1' + B_2 B_2'\right)^{-1} \\
&\times \left[d\xi_t - (A_0 + A_1 m_t)\,dt\right] \\
d\gamma_t &= \Big[b_1 b_1' + b_2 b_2' + a_1 \gamma_t + \gamma_t a_1' \\
&- \left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)\left(B_1 B_1' + B_2 B_2'\right)^{-1}\left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)'\Big]dt
\end{aligned}
\tag{6.2}
$$

6.1.1 Several Unobservable Random Variables and Several Observable Processes

Consider the model

$$
d\xi_t = (A_0 + A_1 \theta)\,dt + B\,dW_t \tag{6.3}
$$

where the unobservable θ is a m-dimensional random vector and the observable ξ_t is a n-dimensional stochastic process.

The filtering equations are

$$
\begin{aligned}
dm_t &= \gamma_t A_1' (BB')^{-1} \left[d\xi_t - (A_0 + A_1 m_t)\, dt \right] \\
\frac{d\gamma_t}{dt} &= -\gamma_t A_1' (BB')^{-1} A_1 \gamma_t
\end{aligned}
\tag{6.4}
$$

Equation (6.4) has an explicit solution

$$
\begin{aligned}
m_t &= \left[I + \Gamma \int_0^t A_1' (BB')^{-1} A_1\, du \right]^{-1} \left[\mu + \Gamma \int_0^t A_1' (BB')^{-1} (d\xi_u - A_0\, du) \right] \\
\gamma_t &= \left[I + \Gamma \int_0^t A_1' (BB')^{-1} A_1\, du \right]^{-1} \Gamma
\end{aligned}
\tag{6.5}
$$

A vector mean-reverting Ornstein-Uhlenbeck process fits this observation model

$$
d\xi_t = k(\nu - \theta)\, dt + B\, dW_t \tag{6.6}
$$

where k is a $m \times m$ matrix and ν is a m-dimensional vector.

One possible application of this model is a situation of competing risks, where there are multiple causes of the event of interest. The random variables $\theta_1, \ldots, \theta_m$ generate the corresponding random times $\tau_1, \ldots, \tau_m$, with the event of interest occurring at the random time $\tau = \min(\tau_1, \ldots, \tau_m)$. This points to a future direction in which this model can be taken.

6.2 One Unobservable Random Variable and Several Observable Processes

The model in this section can serve as an alternative to, and possible improvement on, the Cox proportional hazards model. The proposed model does not require the assumption of proportional hazards, naturally allows the covariates to evolve stochastically over time, and allows us to write the likelihood function from which the parameters of the model can be estimated.

In this model we have one generating random variable θ and a vector of observable processes $\xi_t = \begin{pmatrix} \xi_{1,t} \\ \vdots \\ \xi_{n,t} \end{pmatrix}$. With independent observation noises the model is

$$d\begin{pmatrix} \xi_{1,t} \\ \vdots \\ \xi_{n,t} \end{pmatrix} = \left[\begin{pmatrix} A_{01} \\ \vdots \\ A_{0n} \end{pmatrix} + \begin{pmatrix} A_{11} \\ \vdots \\ A_{1n} \end{pmatrix}\theta\right] dt + \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_n \end{pmatrix} d\begin{pmatrix} W_{1,t} \\ \vdots \\ W_{n,t} \end{pmatrix} \tag{6.7}$$

This is a model of the general form given in Equation (6.1).

To illustrate the application of this model we can think of the unobservable random variable θ as the true condition of a patient's health, and think of the observable processes $\xi_{1,t},\ldots,\xi_{n,t}$ as diastolic blood pressure, level of high-density lipoprotien (HDL) cholesterol, fasting blood sugar, and level of hemoglobin.

From Equation (6.2) we get

$$\begin{aligned} dm_t &= \gamma_t A_1' \left(B_1 B_1'\right)^{-1} \left[d\xi_t - (A_0 + A_1 m_t)\, dt\right] \\ \frac{d\gamma_t}{dt} &= -\gamma_t A_1' \left(B_1 B_1'\right)^{-1} A_1 \gamma_t \end{aligned} \tag{6.8}$$

Further manipulation of the first line of Equation (6.8) yields

$$
\begin{aligned}
dm_t &= \gamma_t A_1' \left(B_1 B_1'\right)^{-1} \left[d\xi_t - (A_0 + A_1 m_t)\, dt\right] \\
&= \gamma_t \begin{pmatrix} A_{11} & \dots & A_{1n} \end{pmatrix} \begin{pmatrix} \frac{1}{B_1^2} & & \\ & \ddots & \\ & & \frac{1}{B_n^2} \end{pmatrix} \begin{pmatrix} d\xi_{1,t} - (A_{01} + A_{11} m_t)\, dt \\ \vdots \\ d\xi_{n,t} - (A_{0n} + A_{1n} m_t)\, dt \end{pmatrix} \\
&= \gamma_t \sum_{i=1}^{n} \frac{d\xi_{i,t} - (A_{0i} + A_{1i} m_t)\, dt}{B_i^2}
\end{aligned} \tag{6.9}
$$

Similarly, for the second line of Equation (6.8)

$$
\frac{d\gamma_t}{dt} = -\left(\sum_{i=1}^{n} \frac{A_{1i}}{B_i^2}\right) \gamma_t^2 \tag{6.10}
$$

We can derive an explicit solution of the filter

$$
\begin{aligned}
m_t &= \frac{\mu + \sigma^2 \displaystyle\sum_{i=1}^{n} \frac{A_{1i}}{B_i^2} (\xi_{i,t} - A_{0i} t)}{1 + \sigma^2 \displaystyle\sum_{i=1}^{n} \left(\frac{A_{1i}}{B_i}\right)^2 t} \\
\gamma_t &= \frac{\sigma^2}{1 + \sigma^2 \displaystyle\sum_{i=1}^{n} \left(\frac{A_{1i}}{B_i}\right)^2 t}
\end{aligned} \tag{6.11}
$$

where $\mu = \mathbb{E}\left(\theta \middle| \mathcal{F}_0^{\xi}\right)$ and $\sigma^2 = \mathrm{var}\left(\theta \middle| \mathcal{F}_0^{\xi}\right)$.

6.3 Simulation of Model with One Unobservable Random Variable and Several Observable Processes

6.3.1 Derivation of Maximum Likelihood Estimators of Model Parameters

To estimate the parameters A_{0i}, A_{1i}, and B_i, we will omit the subscript i for economy of notation. We have

$$\xi_t = \xi_0 + (A_0 + A_1\theta)\,t + BW_t \tag{6.12}$$

To estimate model parameters, we consider a sample of patients for whom we observe their realizations of ξ_t and the time of death τ. Because $\theta = \psi^{-1}(\tau)$, we equivalently observe the realized value of θ for each patient in the sample. Therefore, given an observation of θ

$$\begin{aligned} \mathbb{E}\left(\xi_t\middle|\theta\right) &= \xi_0 + (A_0 + A_1\theta)\,t \\ \operatorname{var}\left(\xi_t\middle|\theta\right) &= B^2 t \\ \operatorname{cov}\left(\xi_t, \xi_s\middle|\theta\right) &= B^2 s \end{aligned} \tag{6.13}$$

It follows that

$$
\begin{aligned}
\mathbb{E}\left(\xi_t \middle| \xi_s, \theta\right) &= \mathbb{E}\left(\xi_t \middle| \theta\right) + \frac{\operatorname{cov}\left(\xi_t, \xi_s \middle| \theta\right)}{\operatorname{var}\left(\xi_s \middle| \theta\right)}\left[\xi_s - \mathbb{E}\left(\xi_s \middle| \theta\right)\right] \\
&= \xi_s + (A_0 + A_1\theta)(t-s) \\
\operatorname{var}\left(\xi_t \middle| \xi_s, \theta\right) &= \operatorname{var}\left(\xi_t \middle| \theta\right) - \frac{\operatorname{cov}^2\left(\xi_t, \xi_s \middle| \theta\right)}{\operatorname{var}(\xi_s)} \\
&= B^2(t-s)
\end{aligned}
\tag{6.14}
$$

Suppose ξ_0 is known and we observe the process ξ_t at times $t_1 < t_2 < \cdots < t_m$. Denote the observed values $x_1, x_2, \ldots, x_m$, then we have

$$
f(x_1, x_2, \ldots, x_m) = f(x_1) f\left(x_2 \middle| x_1\right) f\left(x_3 \middle| x_1, x_2\right) \cdots f\left(x_m \middle| x_1, x_2, \ldots, x_{m-1}\right) \tag{6.15}
$$

Accounting for the fact that the process ξ_t has the Markov property, we can simplify Equation (6.15) to the form

$$
f(x_1, x_2, \ldots, x_m) = f(x_1) f\left(x_2 \middle| x_1\right) f\left(x_3 \middle| x_2\right) \cdots f\left(x_m \middle| x_{m-1}\right) \tag{6.16}
$$

Combining equation 6.14 and equation 6.16 we get the relevant part of the log-likelihood function for patients i from $i = 1$ to $i = 20$, and a representative health marker

$$
\begin{aligned}
\ell &= -\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\log\left[B^2\left(t_j - t_{j-1}\right)\right] \\
&- \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{m_i}\frac{\left[\left(x_{i,j} - x_{i,j-1}\right) - (A_0 + A_1\theta_i)\left(t_j - t_{j-1}\right)\right]^2}{B^2\left(t_j - t_{j-1}\right)}
\end{aligned}
\tag{6.17}
$$

where $p = 20$ is the number of patients in the cross validation set, m_i is the number of observations for patient i, and by definition, $t_0 = 0$. The number of observations of each one of the four health markers is different for different patients because they have different survival times.

Differentiating the relevant part of log-likelihood in equation 6.17 with respect to A_0 and A_1, and equating the derivative to zero, we get the first order conditions for maximizing log-likelihood with respect to those two parameters.

$$
\begin{aligned}
A_0\sum_{i=1}^{p} t_{m_i} + A_1\sum_{i=1}^{p}\theta_i t_{m_i} &= \sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right) \\
A_0\sum_{i=1}^{p}\theta_i t_{m_i} + A_1\sum_{i=1}^{p}\theta_i^2 t_{m_i} &= \sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right)\theta_i
\end{aligned}
\tag{6.18}
$$

The first order conditions in equation 6.18 resemble the normal equations in a weighted linear regression. The solutions are

$$
\begin{aligned}
\widehat{A_0} &= \frac{\left[\sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right)\right]\left(\sum_{i=1}^{p}\theta_i^2 t_{m_i}\right) - \left[\sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right)\theta_i\right]\left(\sum_{i=1}^{p}\theta_i t_{m_i}\right)}{\left(\sum_{i=1}^{p} t_{m_i}\right)\left(\sum_{i=1}^{p}\theta_i^2 t_{m_i}\right) - \left(\sum_{i=1}^{p}\theta_i t_{m_i}\right)^2} \\
\widehat{A_1} &= \frac{\left[\sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right)\theta_i\right]\left(\sum_{i=1}^{p} t_{m_i}\right) - \left[\sum_{i=1}^{p}\left(x_{i,t_{m_i}} - x_{i,0}\right)\right]\left(\sum_{i=1}^{p}\theta_i t_{m_i}\right)}{\left(\sum_{i=1}^{p} t_{m_i}\right)\left(\sum_{i=1}^{p}\theta_i^2 t_{m_i}\right) - \left(\sum_{i=1}^{p}\theta_i t_{m_i}\right)^2}
\end{aligned}
\tag{6.19}
$$

It is important to recognize that the maximum likelihood estimators of the parameters A_0 and A_1 do not require continuous observation of the health marker process ξ_t of any patient, as the estimators in equation 6.19 depends only on the initial and final observations of the process ξ_t represented by $x_{i,0}$ and $x_{i,t_{m_i}}$ respectively.

Similarly, differentiating equation 6.17 with respect ot B^2, we get the estimator

$$
\widehat{B^2} = \left(\sum_{i=1}^{p} m_i\right)^{-1} \sum_{i=1}^{p}\sum_{j=1}^{m_i} \frac{\left[\left(x_{i,j} - x_{i,j-1}\right) - \left(\widehat{A_0} + \widehat{A_1}\theta_i\right)\left(t_j - t_{j-1}\right)\right]^2}{t_j - t_{j-1}} \tag{6.20}
$$

Unfortunately, the maximum likelihood estimator in equation 6.20 requires frequent observation of the health marker processes of all patients, and may not be well-suited for practical application. To deal with issue, I experimented with a simplified version of this estimator which requires observations only 10 times a year. As the simulations results in Table 6.5 show, the simplified estimator for B^2 delivers very good results.

6.3.2 Simulation Results

To fix ideas in the following simulation, I chose four health markers: Diastolic blood pressure, high-density lipoprotein (HDL) cholesterol, fasting blood sugar level, and blood hemoglobin. The normal levels of those health markers are as follows. Health diastolic blood pressure is below 80 mm Hg, healthy level of high-density cholesterol is 60 mg/dL or higher, fasting blood sugar level is 99 mg/dL or lower, and healthy blood hemoglobin level is roughly between 13.2 and 16.5 gm/dL. For some of those markers, there are difference between men and women, which I disregard in this simulation. All the results in this simulation are for illustration only, I do not claim that the simulated levels of those health markers have the simulated effect on longevity.

Table 6.1 shows the assumed initial levels of the four health markers in a simulated group of twenty patients. The last column of the table shows the survival times of the patients, which have been sampled with a Weibull distribution with shape parameter equal to 4, and scale parameter equal to 10.

Table 6.1. Patient data for simulation. Columns (2) through (5) show initial levels of four biomarkers. Column (6) shows survival times drawn from a Weibull distribution with shape parameter $k = 4$, and scale parmater $\lambda = 10$.

Patient No.	Initial Diastolic Blood Pressure	Initial HDL Choesterol	Initial Fasting Blood Sugar	Initial Level of Hemoglobin	Survival Time
1	85	55	99	14.8	7.86
2	82	56	98	14.9	8.02
3	90	50	110	12.7	6.46
4	76	62	92	14.9	10.23
5	87	54	101	14.4	8.49
6	77	60	94	14.9	10.12
7	72	63	90	15.3	11.43
8	83	57	95	14.5	9.00
9	73	64	91	15.0	10.46
10	75	59	92	15.0	8.78
11	86	54	105	14.0	7.69
12	87	52	110	13.8	7.76
13	88	52	112	13.4	6.79
14	81	55	99	14.5	8.25
15	83	51	111	13.5	8.53
16	82	53	109	14.1	8.43
17	72	61	90	15.0	10.91
18	79	56	98	14.6	8.32
19	90	50	121	12.0	5.75
20	70	64	90	15.3	12.57

The assumed values of the model parameters are: $A_0 = \begin{pmatrix} -0.675 \\ -0.625 \\ -0.500 \\ -0.125 \end{pmatrix}$, $A_1 = \begin{pmatrix} 0.675 \\ 0.625 \\ 0.500 \\ 0.125 \end{pmatrix}$,

$B = \begin{pmatrix} 0.20 \\ 0.18 \\ 0.15 \\ 0.04 \end{pmatrix}$.

Next, recalling that $\theta = \psi^{-1}(\tau) = \Phi^{-1}\left[F(\tau)\right]$, where F is the Weibull distribution used in the simulation and Φ is the standard normal distribution function. I calculated the values of θ corresponding to the 20 survival times in Table 6.1. Table 6.2 shows the 20 survival times and corresponding values of the generating random variable θ.

Table 6.2. Survival times and corresponding values of generating Variable

Patient No.	Survival Time τ	Generating Variable θ	Patient No.	Survival Time τ	Generating Variable θ
1	7.86	-0.4753	11	7.69	-0.5385
2	8.02	-0.4147	12	6.79	-0.5125
3	6.46	-0.9952	13	8.25	-0.8724
4	10.23	0.4276	14	8.53	-0.3298
5	8.49	-0.2399	15	8.25	-0.2248
6	10.12	0.3844	16	8.43	-0.2624
7	11.43	0.9099	17	10.91	0.6983
8	9.00	-0.0473	18	8.32	-0.3036
9	10.46	0.5185	19	5.75	-1.2616
10	8.78	-0.1306	20	12.57	1.3893

Figure 6.1 through Figure 6.4 show a realization of the stochastic evolution of the four health markers in the sample of the 20 patients. The length of the lines corresponds to the survival time of the patients. Each patient has an assigned color that is consistent through the four markers.

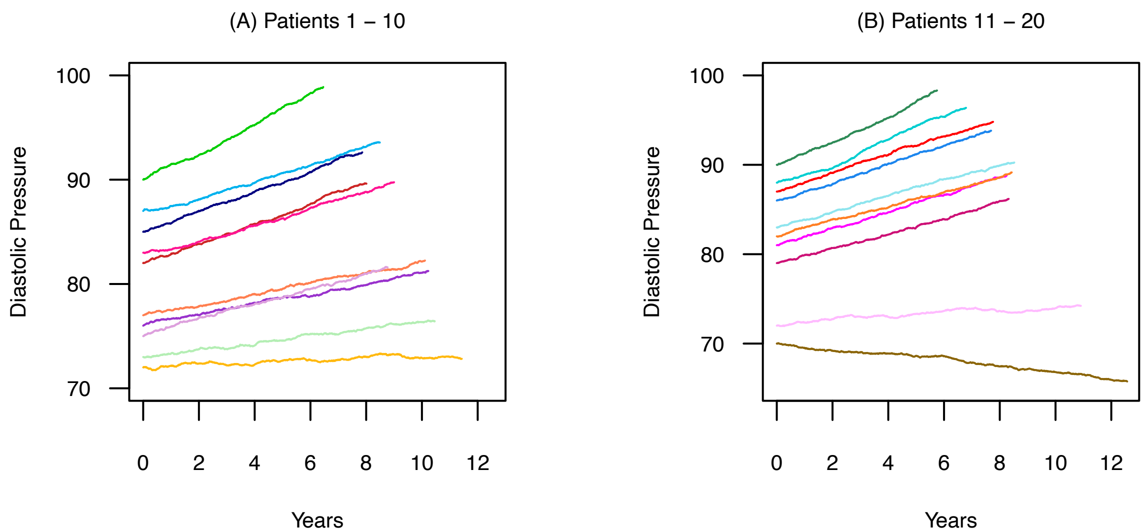

Figure 6.1. Evolution of diastolic blood pressure in 20 patients.

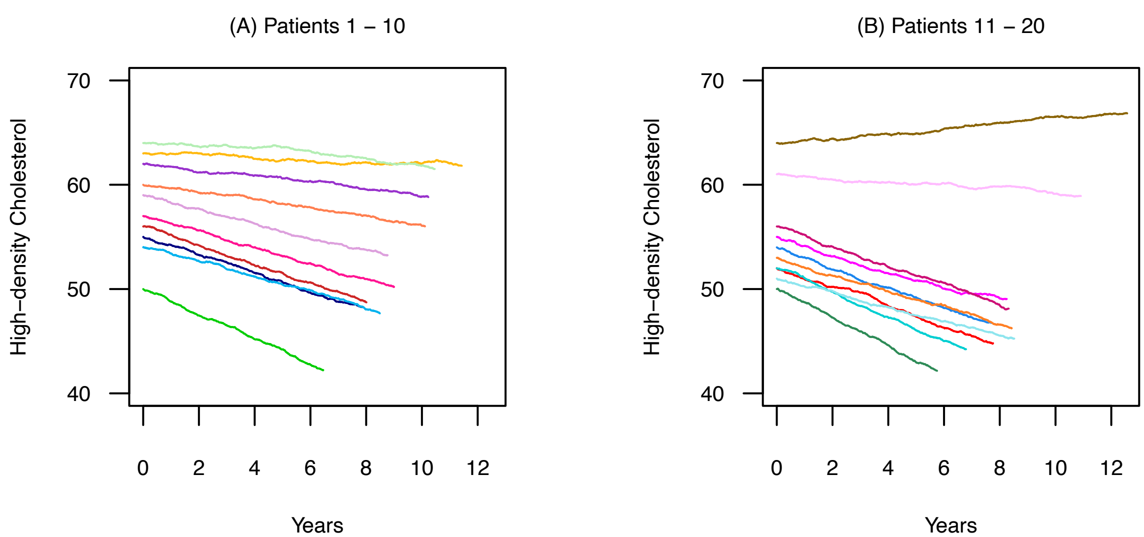

Figure 6.2. Evolution of high-density lipoprotein cholesterol in 20 patients.

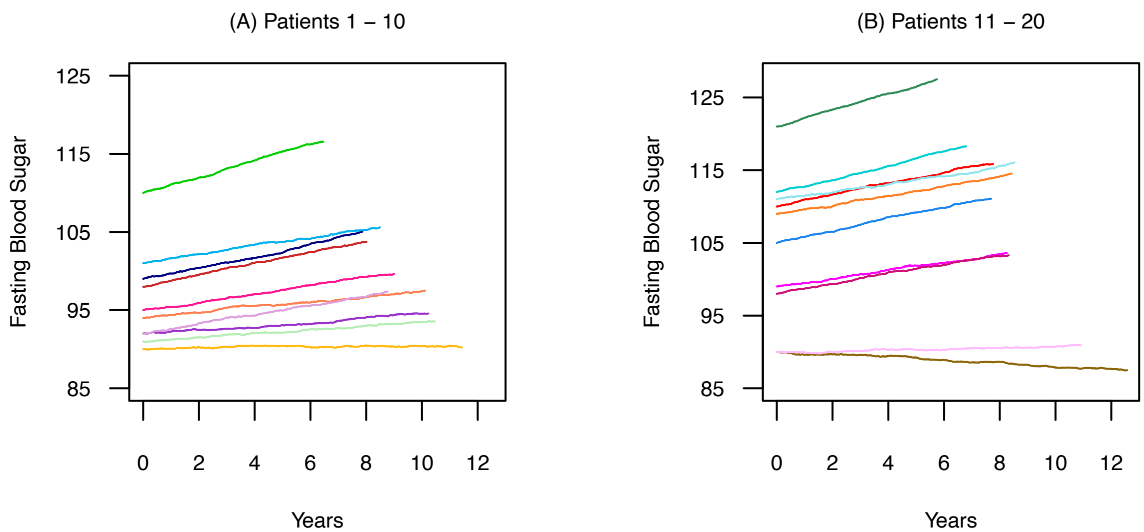

Figure 6.3. Evolution of fasting blood sugar in 20 patients.

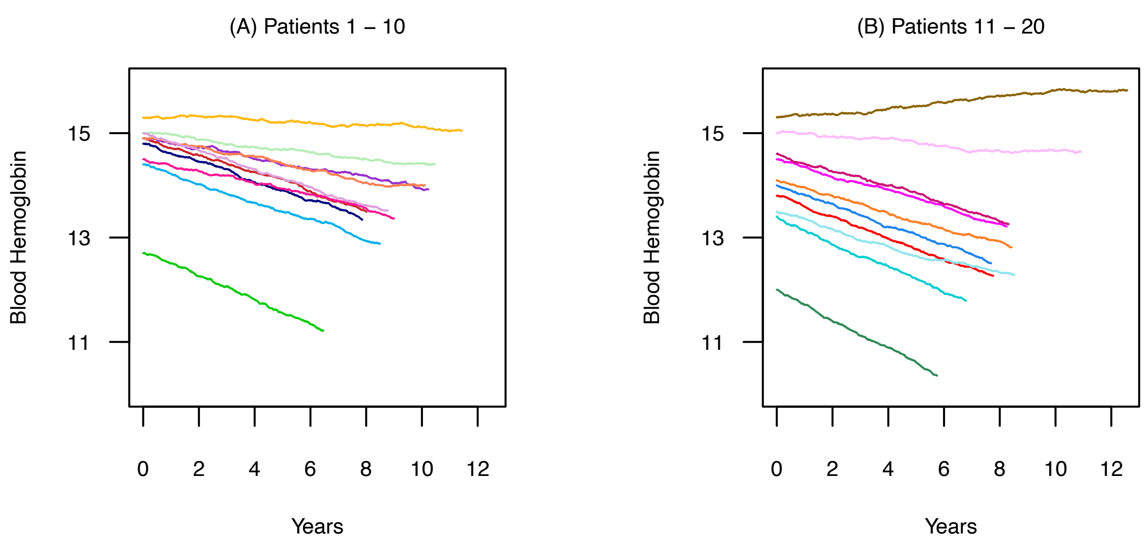

Figure 6.4. Evolution of blood hemoglobin level in 20 patients.

Table 6.3 through Table 6.5 display the results of estimating the parameters A_{0i}, A_{1i}, and B_i in the context of 20-fold leave-one-out cross-validation analysis in which we systematically leave out each one of the patients in turn and estimate the parameters from the data of the remaining 19 patients, using Equation (6.19) and the simplified version of Equation (6.20), as discussed immediately following that equation. The 20 estimates of each one of the 12 parameters are close both to their true values and to each other.

Table 6.3. Estimation of Parameter Vector A_0

Leave Out	$\widehat{A}_{01}$	$\widehat{A}_{02}$	$\widehat{A}_{03}$	$\widehat{A}_{04}$	Leave Out	$\widehat{A}_{01}$	$\widehat{A}_{02}$	$\widehat{A}_{03}$	$\widehat{A}_{04}$
True	-0.6750	-0.6250	-0.5000	-0.1250	True	-0.6750	-0.6250	-0.5000	-0.1250
1	-0.6792	-0.6142	-0.4895	-0.1288	11	-0.6800	-0.6161	-0.4865	-0.1275
2	-0.6805	-0.6163	-0.4878	-0.1288	12	-0.6803	-0.6140	-0.4883	-0.1288
3	-0.6795	-0.6121	-0.4889	-0.1290	13	-0.6800	-0.6140	-0.4893	-0.1285
4	-0.6778	-0.6147	-0.4888	-0.1296	14	-0.6805	-0.6188	-0.4945	-0.1294
5	-0.6718	-0.6170	-0.4908	-0.1277	15	-0.6769	-0.6180	-0.4900	-0.1295
6	-0.6826	-0.6149	-0.4931	-0.1277	16	-0.6781	-0.6127	-0.4878	-0.1292
7	-0.6733	-0.6134	-0.4863	-0.1284	17	-0.6796	-0.6143	-0.4932	-0.1296
8	-0.6792	-0.6111	-0.4900	-0.1286	18	-0.6799	-0.6066	-0.4900	-0.1290
9	-0.6771	-0.6081	-0.4894	-0.1293	19	-0.6801	-0.6147	-0.4896	-0.1286
10	-0.6796	-0.6183	-0.4885	-0.1294	20	-0.6822	-0.6143	-0.4875	-0.1292

Table 6.4. Estimation of Parameter Vector A_1

Leave Out No.	$\widehat{A}_{11}$	$\widehat{A}_{12}$	$\widehat{A}_{13}$	$\widehat{A}_{14}$	Leave Out No.	$\widehat{A}_{11}$	$\widehat{A}_{12}$	$\widehat{A}_{13}$	$\widehat{A}_{14}$
True	0.6750	0.6250	0.5000	0.1250	True	0.6750	0.6250	0.5000	0.1250
1	0.6678	0.6068	0.5078	0.1202	11	0.6677	0.6045	0.5082	0.1201
2	0.6668	0.6027	0.5090	0.1203	12	0.6673	0.6045	0.5094	0.1198
3	0.6634	0.6062	0.5084	0.1218	13	0.6687	0.6047	0.5106	0.1197
4	0.6739	0.6008	0.5073	0.1216	14	0.6649	0.6078	0.5130	0.1207
5	0.6684	0.6047	0.5112	0.1197	15	0.6661	0.6062	0.5096	0.1206
6	0.6719	0.6042	0.5112	0.1209	16	0.6668	0.6036	0.5085	0.1205
7	0.6674	0.6084	0.5065	0.1214	17	0.6661	0.6027	0.5025	0.1195
8	0.6665	0.6035	0.5092	0.1204	18	0.6672	0.5995	0.5098	0.1204
9	0.6667	0.5986	0.5096	0.1200	19	0.6743	0.6066	0.5105	0.1193
10	0.6669	0.6048	0.5084	0.1199	20	0.6381	0.6002	0.5116	0.1205

Table 6.5. Estimation of Parameter Vector B

Leave Out No.	$\widehat{B}_1$	$\widehat{B}_2$	$\widehat{B}_3$	$\widehat{B}_4$	Leave Out No.	$\widehat{B}_1$	$\widehat{B}_2$	$\widehat{B}_3$	$\widehat{B}_4$
True	0.2000	0.1800	0.1500	0.0400	True	0.2000	0.1800	0.1500	0.0400
1	0.1993	0.1743	0.1441	0.0401	11	0.1983	0.1744	0.1435	0.0402
2	0.1983	0.1746	0.1440	0.0401	12	0.1982	0.1741	0.1446	0.0402
3	0.1984	0.1745	0.1446	0.0398	13	0.1987	0.1744	0.1445	0.0402
4	0.1977	0.1746	0.1436	0.0396	14	0.1994	0.1737	0.1440	0.0401
5	0.1988	0.1751	0.1434	0.0402	15	0.1992	0.1752	0.1442	0.0400
6	0.1975	0.1754	0.1433	0.0400	16	0.1976	0.1731	0.1444	0.0402
7	0.1973	0.1733	0.1449	0.0400	17	0.1983	0.1745	0.1441	0.0401
8	0.1980	0.1751	0.1447	0.0400	18	0.1986	0.1750	0.1434	0.0402
9	0.1991	0.1737	0.1431	0.0402	19	0.1986	0.1747	0.1439	0.0402
10	0.1992	0.1741	0.1439	0.0402	20	0.1989	0.1760	0.1441	0.0399

Figure 6.5 displays the evolution of the 20 conditional means $m_t = \mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ where ξ_t represents the $n = 4$ health markers in the simulation, using the 20-fold leave-one-out estimation of the filter parameters. The values of m_t were computed from the first line of Equation (6.11). The colors assigned to the patients are consistent throughout and corresponds to the colors in Figure 6.1 through Figure 6.4 of the evolution of the four health markers. We can see that after a burn-in period of about one year, the conditional means converge, for each patient, to values that are close to the corresponding true values of the generating random variable θ.

Figure 6.6 shows the conditional standard deviation $\sqrt{\gamma_t}$ computed from the second line of Equation (6.11). Panel (A) shows the evolution of the conditional standard deviation of the shortest-lived patient, and Panel (B) shows the evolution of the conditional standard deviation of the longest-lived patient. The figure shows only those two patients because all the 20 conditional standard deviations are close, and mostly differ in their evolution times. It is also important to note that the smooth evolution of the conditional standard deviation comes from the normal distribution of the generating variable θ, so that conditional standard deviation is only function of time and does not depend on the observed history of the health markers. Finally, the conditional standard deviation declines rapidly to fairly small values, further supporting the accuracy of the estimated conditional means.

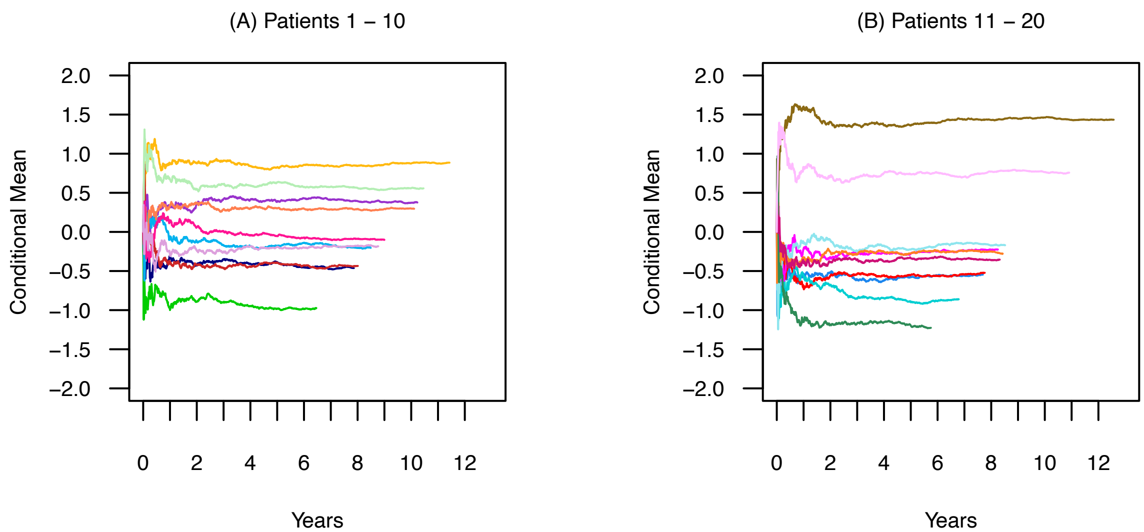

Figure 6.5. Conditional mean $m_t = \mathbb{E}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ of 20 patients.

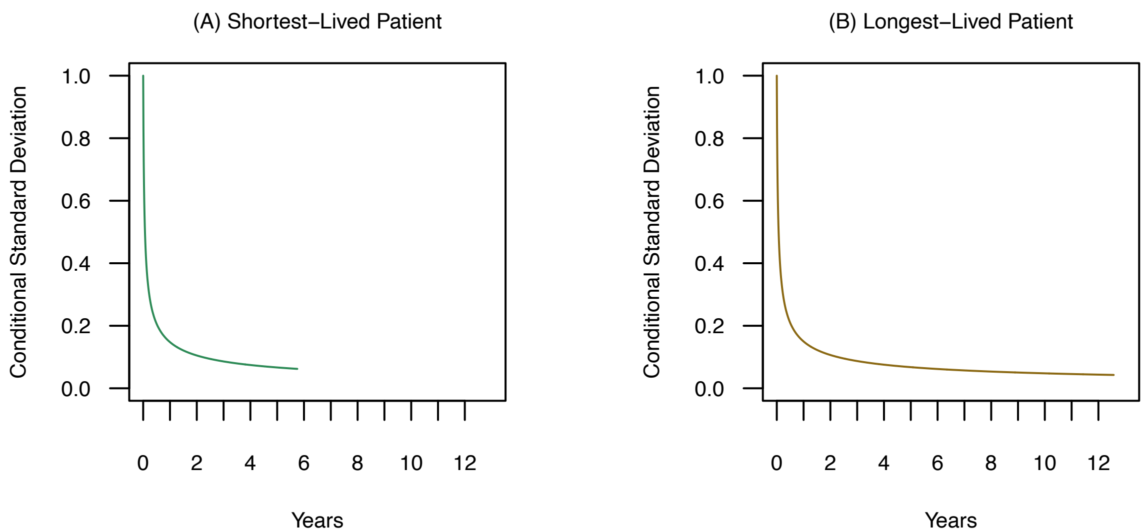

Figure 6.6. Square root of conditional variance $\gamma_t = \mathrm{var}\left(\theta \middle| \mathcal{F}_t^{\xi}\right)$ of two patients.

Table 6.6 shows the 20 values of the conditional mean m_t for t equal to $\tau_i - 2$, that is, two years before the survival time of each patient. The purpose of calculating those numbers is to determine, for each patient, the probability that the patient survives six months or longer beyond his or her actual survival time τ_i. The formula for calculating this probability comes from the following equation

$$\begin{aligned} \mathbb{P}\left(\tau > y \middle| \mathcal{F}_t^{\xi}\right) &= \mathbb{P}\left[\theta > \psi^{-1}(y) \middle| \mathcal{F}_t^{\xi}\right] \\ &= \Phi\left[\frac{m_t - \psi^{-1}(y)}{\sqrt{\gamma_t}}\right] \end{aligned} \tag{6.21}$$

where $t = \tau_i - 2$ and $y = \tau_i + 0.5$. The simulation was repeated 100 times and average probabilities are shown in the last column of Table 6.6. None of the average probabilities is more than 5.5 percent.

In addition, Figure 6.7 shows the histograms of those probabilities for each one of the 20 patients. For each patient, the frequency of such probability of survival being between 0 and 5 percent is at least 80 out of 100. We can see from those results that the probabilities of survival of six months or longer beyond the actual survival time are small for all patients.

Table 6.6. Conditional Mean and Standard Deviation, and Average Probability of Survival 6 Months Beyond Time τ

Leave Out Patient No.	Generating Variable θ	Cond Mean 2 Yr Before Death	Cond Std Dev 2 Yr Before Death	Average Prob of Survival 6 Mos After τ
1	-0.4753	-0.4951	0.0605	0.0177
2	-0.4157	-0.3701	0.0601	0.0140
3	-0.9952	-0.9958	0.0695	0.0347
4	0.4276	0.2937	0.0512	0.0031
5	-0.2399	-0.2768	0.0579	0.0092
6	0.3844	0.4852	0.0518	0.0057
7	0.9099	0.9557	0.0480	0.0018
8	-0.0473	0.0433	0.0556	0.0072
9	0.5185	0.4861	0.0508	0.0020
10	-0.1306	-0.1181	0.0566	0.0066
11	-0.5385	-0.4237	0.0615	0.0241
12	-0.5125	-0.4601	0.0614	0.0239
13	-0.8724	-0.9386	0.0676	0.0408
14	-0.3298	-0.2303	0.0587	0.0137
15	-0.2248	-0.2420	0.0578	0.0155
16	-0.2624	-0.2673	0.0582	0.0086
17	0.6983	0.7039	0.0494	0.0013
18	-0.3036	-0.3702	0.0585	0.0155
19	-1.2616	-1.3934	0.0766	0.0544
20	1.3893	1.3127	0.0450	0.0018

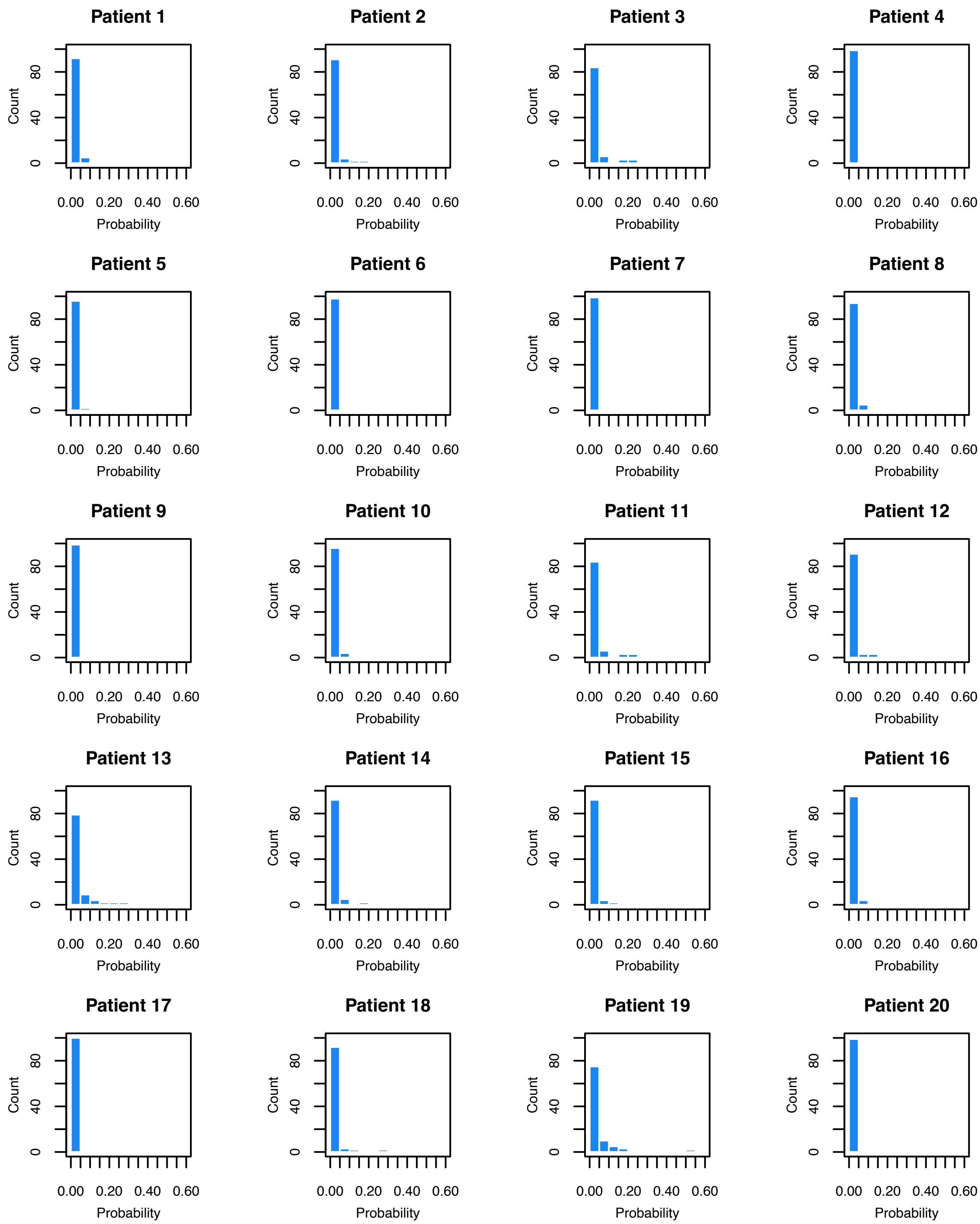

Figure 6.7. Histogram of probability of survival 6 months beyond time τ.

7 Forecasting Investment Performance

7.1 Introduction

This chapter describes a financial application of the models developed in the preceding chapters. Working with real data for the period 1928 to 2021, I estimate the distribution of weekly log-returns to the S & P 500 stock market index. As it is clear that the normal distribution doest not fit log-return data, I show the results of working first with an asymmetric Laplace distribution, and second, with a noraml-Laplace distribution. The latter is a convolution of a normal distribution and an asymmetic Laplace distribution.

After arriving at a suitable approximate distribution of log-returns, I apply the model developed in the preceding chapters to a problem of aggregating forecasts made by analysts with a normally distributed error into a group forecast that corresponds to the fitted normal-Laplace distribution of log-returns.

7.2 Standard & Poor 500 Stock Index Weekly Data

Description of Data

- Data source: Yahoo Finance
- Starting date: December 30, 1927
- Ending date: December 31, 2021
- Starting index value: 17.66
- Ending index value: 4766.18

- Range of index values: 4.41, 4766.18
- Sample size of log-returns: 4903
- Range of weekly log-returns: -0.205370, 0.149385

Sample moments of weekly log-returns

- Mean: 0.001142
- Standard deviation: 0.024958
- Skewness: -0.610736
- Excess kurtosis: 6.645062

Figure 7.1 shows the S & P 500 weekly closing prices, the natural log of weekly closing prices, and weekly log price returns (dividends not included).

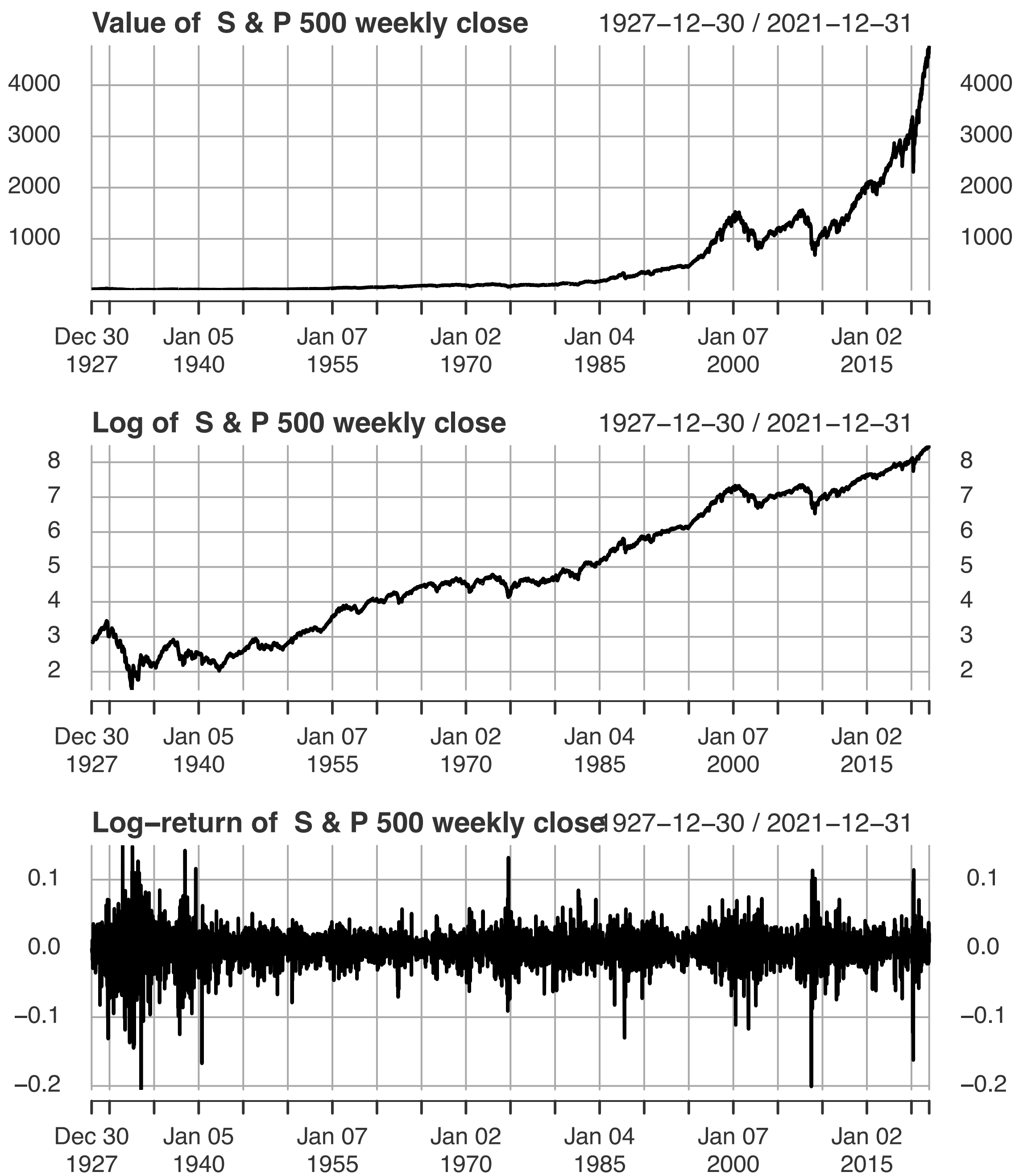

Figure 7.1. S & P 500 weekly closing price, weekly log of closing price, and weekly log price return.

7.3 Fitting a Normal Distribution to Log Returns

Maximum likelihood estimation produces the result

$$\begin{aligned} \widehat{\mu} &= 0.001142 \\ \widehat{\sigma} &= 0.024956 \end{aligned} \tag{7.1}$$

Log-likelihood of the fitted parameters is 11138.19.

AIC of the fitted parameters is -22272.38.

Figure 7.2 shows the Q-Q plot for the fitted normal distribution.

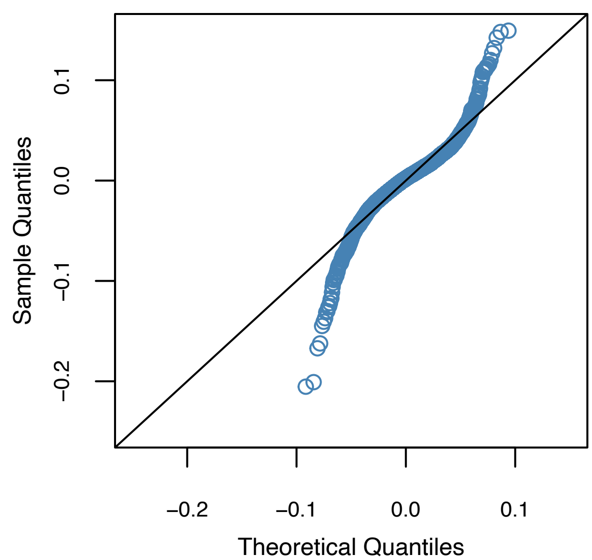

Figure 7.2. Q-Q plot for fitted normal distribution.

Next, Figure 7.3 shows logarithmic tail plots with the sample (blue circles) and the fitted normal distribution (black line). The logarithm is base 10 for more intuitive interpretation, that is, a difference of 1 means a probability ratio of 10, a difference of 2 means a probability ratio of 100, and so on.

For meaningful examination, we need to consider the left tail and the right tail separately, with the examination of the right tail being based on the complementary distribution or survival function.

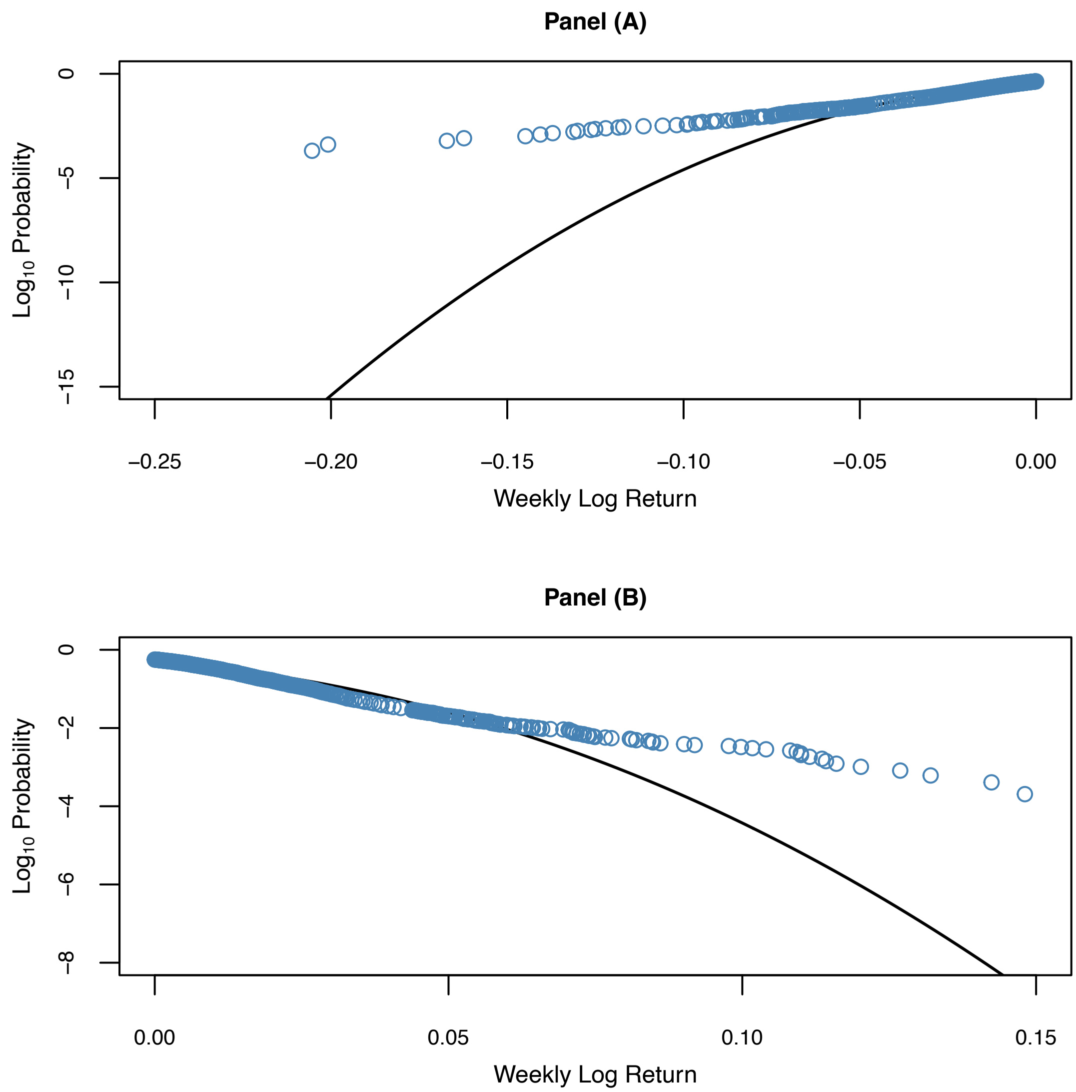

Figure 7.3. Goodness-of-fit tail plots for fitted normal distribution. Panel (A) shows left tail of empirical and fitted distribution functions and Panel (B) shows right tail of empirical and fitted complementary distribution functions.

The χ^2 test used here for goodness-of-fit follows the method of Vose (Vose [40]). The sample of size $N = 4903$ is divided into the integer part of $(4N)^{\frac{2}{5}}$ groups (here 52), with 51 groups of equal size (here 94) and the 52nd group picking up the rest of the sample. It is important to keep in mind that the χ^2 test depends on the definition of groups.

The resulting χ^2 statistic is 635.116 with $52 - 2 - 1 = 49$ degrees of freedom and corresponding p-value of 6.488×10^{-103}. Table 7.1 and Table 7.2 break down the calculation of the χ^2 statistic.

The interpretation of the numbers in the Group column is as follows. Denote the number in row $1 \leq i \leq 52$ by g_i. Then Group 1 is $(-\infty, g_1]$, Group 52 is (g_{52}, ∞), and every other Group i is $(g_{i-1}, g_i]$.

Table 7.1. χ^2 Test for Normmal.

	Group	Observed	Expected
1	-0.0594	94	37.364
2	-0.0444	94	129.576
3	-0.0362	94	163.591
4	-0.0300	94	187.932
5	-0.0265	94	138.444
6	-0.0233	94	146.236
7	-0.0208	94	128.672
8	-0.0187	94	112.355
9	-0.0168	94	111.243
10	-0.0152	94	103.663
11	-0.0137	94	96.641
12	-0.0122	94	98.957
13	-0.0107	94	100.268
14	-0.0095	94	84.436
15	-0.0083	94	92.441
16	-0.0071	94	87.089
17	-0.0059	94	85.444
18	-0.0048	94	87.339
19	-0.0037	94	83.494
20	-0.0027	94	77.954
21	-0.0017	94	77.508
22	-0.0006	94	83.275
23	0.0002	94	66.839
24	0.0011	94	66.907
25	0.0020	94	71.547
26	0.0030	94	73.971

Table 7.2. χ^2 Test for Normal (Cont.)

	Group	Observed	Expected
27	0.0037	94	61.575
28	0.0046	94	69.343
29	0.0054	94	59.266
30	0.0061	94	56.431
31	0.0069	94	55.282
32	0.0077	94	64.646
33	0.0085	94	63.614
34	0.0095	94	68.319
35	0.0104	94	68.180
36	0.0113	94	69.228
37	0.0120	94	47.375
38	0.0132	94	82.222
39	0.0141	94	65.800
40	0.0151	94	63.519
41	0.0161	94	67.784
42	0.0172	94	72.975
43	0.0187	94	95.398
44	0.0204	94	101.826
45	0.0222	94	100.122
46	0.0242	94	106.732
47	0.0267	94	119.702
48	0.0291	94	107.764
49	0.0322	94	120.288
50	0.0378	94	175.032
51	0.0485	94	206.157
52	0.0485	109	141.235

Using a Lilliefors-corrected (Lilliefors [26]) Kolmogorov-Smirnov test in the R package KScorrect, when the parameters of the normal distribution are estimated from the sample, we get p-value of 2×10^{-4}.

Using a Stephens-corrected (Stephens [35]) Anderson-Darling test in the R package nortest, when the parameters of the normal distribution are estimated from the sample, we get p-value of $< 2.2 \times 10^{-16}$.

Using a Stephens-corrected (Stephens [35]) Cramer-von Mises test in the R package nortest, when the parameters of the normal distribution are estimated from the sample, we get p-value of 7.37×10^{-10}.

Using a Lilliefors-Stephens-Dallal-Wilkinson-corrected (Lilliefors [26], Stephens [34], Dallal and Wilkinson [14]) Kolmogorov-Smirnov test in the R package nortest, when the parameters of the normal distribution are estimated from the sample, we get p-value of $< 2.2 \times 10^{-16}$.

For additional details of the various corrections see Thode [38].

Using a Braun-corrected (Braun [8]) Anderson-Darling test in the R package goftest, when the parameters of the normal distribution are estimated from the sample, we get p-value of 0.3914.

Using a Braun-corrected (Braun [8]) Cramer-von Mises test in the R package goftest, when the parameters of the normal distribution are estimated from the sample, we get p-value of 0.2673.

Those results illustrate the low power of some of the available analytical goodness-of-fit tests when the parameters of the fitted distribution are estimated from the sample, and the importance of graphical methods in assessing goodness of fit. See for example Casella and Berger [10], Klugman et al. [25], Delignette-Muller and Dutang [15], and references cited there.

7.4 Asymmetric Laplace and Double Pareto Distributions

7.4.1 Asymmetric Laplace Distribution

Univariate Laplace distribution is the distribution of a random variable

$$X = k + U - V \tag{7.2}$$

where k is a constant, U has exponential distribution with rate β, V has an exponential distribution with rate α, and U and V are independent.When $\alpha \neq \beta$ the Laplace distribution is called asymmetric.

Univariate, asymmetric Laplace distribution has three parameters: Rate parameters $\alpha > 0$ and $\beta > 0$, and location parameter k. Its density function is

$$f(x) = \frac{\alpha\beta}{\alpha+\beta} \times \begin{cases} \exp\Big(-\alpha|x-k|\Big) & \text{if} \quad x \leq k \\ \exp\Big(-\beta|x-k|\Big) & \text{if} \quad x > k \end{cases} \tag{7.3}$$

or alternatively

$$f(x) = \frac{\alpha\beta}{\alpha+\beta} \times \begin{cases} \exp\Big[\alpha(x-k)\Big] & \text{if} \quad x \leq k \\ \exp\Big[-\beta(x-k)\Big] & \text{if} \quad x > k \end{cases} \tag{7.4}$$

The corresponding distribution function is

$$F(x) = \begin{cases} \frac{\beta}{\alpha+\beta}\exp\Big[\alpha(x-k)\Big] & \text{if} \quad x \leq k \\ 1 - \frac{\alpha}{\alpha+\beta}\exp\Big[-\beta(x-k)\Big] & \text{if} \quad x > k \end{cases} \tag{7.5}$$

If random variable X has asymmetric Laplace distribution then

$$
\begin{aligned}
\mathbb{E}(X) &= k + \frac{1}{\beta} - \frac{1}{\alpha} \\
\mathrm{var}(X) &= \frac{1}{\alpha^2} + \frac{1}{\beta^2} \\
\mathrm{skew}(X) &= 2\frac{\alpha^3 - \beta^3}{(\alpha^2 + \beta^2)^{\frac{3}{2}}} \\
\mathrm{ex.kurt}(X) &= 6\frac{\alpha^4 + \beta^4}{(\alpha^2 + \beta^2)^2}
\end{aligned}
\tag{7.6}
$$

where $\mathrm{skew}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$, and $\mathrm{ex.kurt}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$. It can be shown that for the asymmetric Laplace distribution $-2 < \mathrm{skew}(X) < 2$ and $3 < \mathrm{ex.kurt}(X) < 6$.

Maximum Likelihood Estimation of Parameters

The log-likelihood function of an i.i.d. sample $x = (x_1, x_2, \ldots, x_n)$ from an asymmetric Laplace distribution is

$$
\ell(\alpha, \beta, k; x) = n \log \frac{\alpha\beta}{\alpha + \beta} - \alpha \sum_{i=1}^{n} (x_i - k)^- - \beta \sum_{i=1}^{n} (x_i - k)^+ \tag{7.7}
$$

where

$$
\begin{aligned}
(x-k)^- &= \begin{cases} |x-k| & \text{if} \quad x \le k \\ 0 & \text{if} \quad x > k \end{cases} \\
(x-k)^+ &= \begin{cases} 0 & \text{if} \quad x \le k \\ |x-k| & \text{if} \quad x > k \end{cases}
\end{aligned} \tag{7.8}
$$

Assume for a moment that we know the values of the parameters α and β. Then the estimation of k consists of minimizing the function

$$
M(k;x) = \alpha \sum_{i=1}^{n} (x_i - k)^- + \beta \sum_{i=1}^{n} (x_i - k)^+ \tag{7.9}
$$

Without loss of generality, we can assume that the x_i are order statistics, that is, they are arranged in non-decreasing order. Then the function $M(k;x)$ is piecewise linear and continuous (polygonal line), which is decreasing for all $k \le x_1$ and increasing for all $k > x_n$. Therefore, $M(k;x)$ attains a global minimum at one of the sample points x_i.

For example, when $k \le x_1$ we have $M(k;x) = \beta \sum_{i=1}^{n} (x_i - k)$, when $x_1 < k \le x_2$ we have $M(k;x) = \alpha(k - x_1) + \beta \sum_{i=2}^{n} (x_i - k)$, and when $k > x_n$ we have $M(k;x) = \alpha \sum_{i=1}^{n} (k - x_i)$.

Assume now for a moment that we know k and want to derive the maximum likelihood estimators of α and β. Differentiating the log-likelihood function in Equation (7.7) with respect to α, and separately, with respect to β, and equating the derivatives to zero, we get

$$
\begin{aligned}
\frac{\partial \ell}{\partial \alpha} &= \frac{n}{\alpha} - \frac{n}{\alpha+\beta} - \sum_{i=1}^{n} (x_i - k)^- &= 0 \\
\frac{\partial \ell}{\partial \beta} &= \frac{n}{\beta} - \frac{n}{\alpha+\beta} - \sum_{i=1}^{n} (x_i - k)^+ &= 0
\end{aligned} \tag{7.10}
$$

Multiply the first line in Equation (**??**) by α and the second line by β and rearrabge terms to get

$$
\begin{aligned}
\frac{n\beta}{\alpha+\beta} &= \alpha \sum_{i=1}^{n} (x_i - k)^- \\
\frac{n\alpha}{\alpha+\beta} &= \beta \sum_{i=1}^{n} (x_i - k)^+
\end{aligned} \tag{7.11}
$$

Assuming that the distribution is two-sided, that is, neither sum above is zero, gives

$$
\frac{\beta}{\alpha} = \frac{\alpha}{\beta} \frac{\sum_{i=1}^{n} (x_i - k)^-}{\sum_{i=1}^{n} (x_i - k)^+} \tag{7.12}
$$

and

$$
\frac{\beta}{\alpha} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - k)^-}{\sum_{i=1}^{n} (x_i - k)^+}} \tag{7.13}
$$

Denote $p = \sqrt{\sum_{i=1}^{n} (x_i - k)^-}$ and $q = \sqrt{\sum_{i=1}^{n} (x_i - k)^+}$. Then using Equation (7.11) and Equation (7.13) we get the estimators

$$\begin{aligned} \widehat{\alpha} &= \frac{n}{p(p+q)} \\ \widehat{\beta} &= \frac{n}{q(p+q)} \end{aligned} \tag{7.14}$$

The strategy for obtaining the estimators is to try $\widehat{k} = x_i$ for each $1 \le i \le n$ and calculate the corresponding $\widehat{\alpha}$ and $\widehat{\beta}$ to get the largest value of the log-likrlihood function.

7.4.2 Double Pareto Distribution

If log-returns to the S & P 500 index have an asymmetric Laplace distribution, then the index itself has a double Pareto distribution.

Let $Y = Y_0 \exp(X)$ and denote $Y_0 \exp(k) = K$.

The resulting double Pareto density function is

$$g(y) = \frac{\alpha\beta}{\alpha+\beta} \times \begin{cases} \dfrac{y^{\alpha-1}}{K^{\alpha}} & \text{if} \quad y \le K \\ \dfrac{K^{\beta}}{y^{\beta+1}} & \text{if} \quad y > K \end{cases} \tag{7.15}$$

The corresponding distribution function is

$$G(y) = \begin{cases} \dfrac{\beta}{\alpha+\beta}\left(\dfrac{y}{K}\right)^{\alpha} & \text{if} \quad y \leq K \\ 1 - \dfrac{\alpha}{\alpha+\beta}\left(\dfrac{K}{y}\right)^{\beta} & \text{if} \quad y > K \end{cases} \tag{7.16}$$

If random variable Y has double Pareto distribution then

$$\begin{aligned} \mathbb{E}(Y) &= \frac{\alpha\beta}{(\alpha+1)(\beta-1)}K & \text{if} \quad \beta > 1 \\ \text{var}(Y) &= \frac{\alpha\beta}{(\alpha+1)^2(\beta-1)^2}\left[\frac{(\alpha+1)^2(\beta-1)^2}{(\alpha+2)(\beta-2)} - \alpha\beta\right]K^2 & \text{if} \quad \beta > 2 \end{aligned} \tag{7.17}$$

7.5 Normal-Laplace Distribution

Reed and Jorgensen [31] and Reed [30] introduced a normal Laplace distribution, a four-parameter distribution with location parameter μ, scale parameter $\sigma > 0$, and two rate parameters $\alpha, \beta > 0$. By definition, a random variable X with a normal-Laplace distribution has the representation

$$X = \mu + \sigma Z + U - V \tag{7.18}$$

where Z is standard normal, U is exponential with rate parameter β, and V is exponential with rate parameter α. The normal-Laplace density function is

$$f(x) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{x-\mu}{\sigma}\right)\left[h\left(\alpha\sigma + \frac{x-\mu}{\sigma}\right) + h\left(\beta\sigma - \frac{x-\mu}{\sigma}\right)\right] \tag{7.19}$$

where $\phi(z)$ is standard normal density function, $\Phi(z)$ is the standard normal distribution function, and $h(z) = \frac{1-\Phi(z)}{\phi(z)}$ is one over the standard normal hazard function.

The normal Laplace distribution function is

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) + \phi\left(\frac{x-\mu}{\sigma}\right)\left[\frac{\beta}{\alpha+\beta}h\left(\alpha\sigma + \frac{x-\mu}{\sigma}\right) - \frac{\alpha}{\alpha+\beta}h\left(\beta\sigma - \frac{x-\mu}{\sigma}\right)\right] \quad (7.20)$$

The first five cumulants of the normal-Laplace distribution are

$$\begin{aligned} \kappa_1 &= \mu + \frac{1}{\beta} - \frac{1}{\alpha} \\ \kappa_2 &= \sigma^2 + \frac{1}{\alpha^2} + \frac{1}{\beta^2} \\ \kappa_3 &= \frac{2}{\beta^3} - \frac{2}{\alpha^3} \\ \kappa_4 &= \frac{6}{\alpha^4} + \frac{6}{\beta^4} \\ \kappa_5 &= \frac{24}{\beta^5} - \frac{24}{\alpha^5} \end{aligned} \quad (7.21)$$

The cumulant formulas can be used to compute estimators of mean, variance, skewness, and kurtosis. Care, however, needs to be taken with the computation to avoid nummerical instability, and products of the form ϕh have to calculated using log transform.

7.6 Fitting a Normal Laplace Distribution to Log Returns

Maximum likelihood estimation produces the result

$$\begin{aligned} \widehat{\mu} &= 0.005640 \\ \widehat{\sigma} &= 0.005521 \\ \widehat{\alpha} &= 53.806293 \\ \widehat{\beta} &= 70.989143 \end{aligned} \tag{7.22}$$

Log-likelihood of the fitted parameters is 11648.30.

AIC of the fitted parameters is -23288.60.

Figure 7.4 shows the Q-Q plot for the fitted normal Laplace distribution.

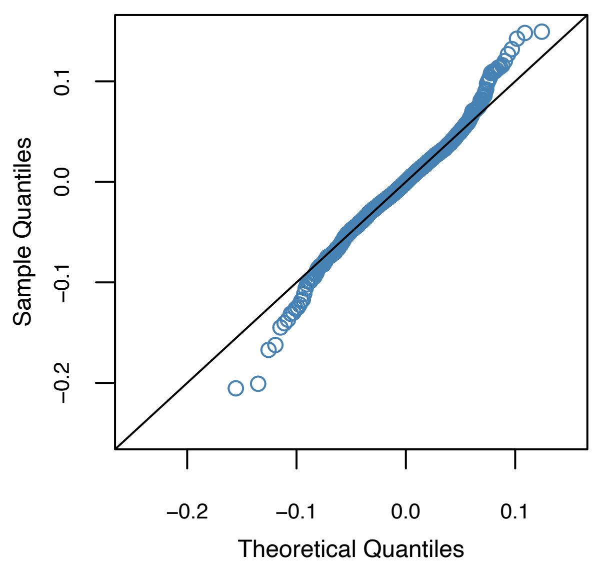

Figure 7.4. Q-Q plot for fitted normal Laplace distribution.

Next, Figure 7.5 shows logarithmic tail plots with the sample (blue circles) and the fitted normal Laplace distribution (black line). The logarithm is base 10 for more intuitive interpretation, that is, a difference of 1 means a probability ratio of 10, a difference of 2 means a probability ratio of 100, and so on.

For meaningful examination, we need to consider the left tail and the right tail separately, with the examination of the right tail being based on the complementary distribution or survival function.

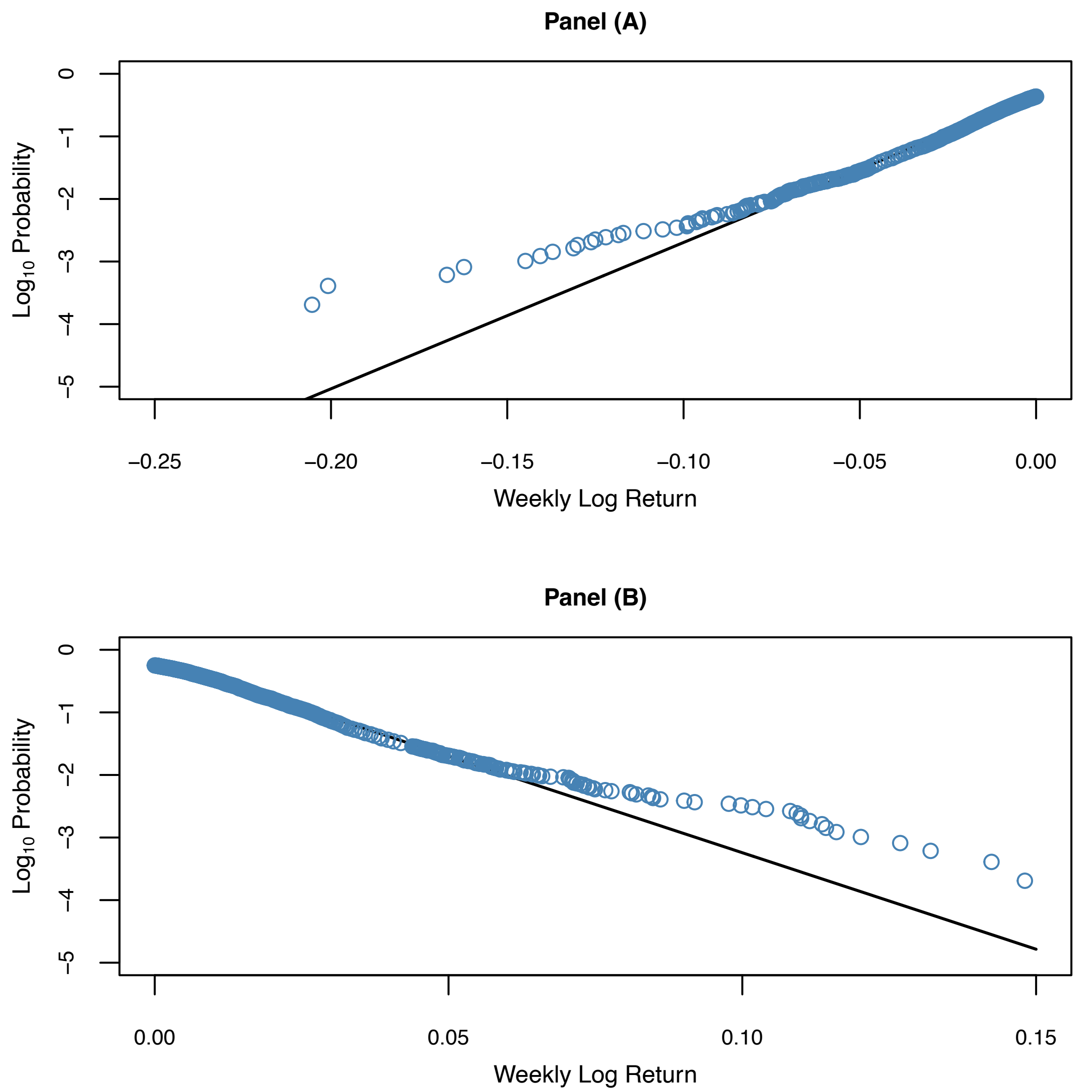

Figure 7.5. Goodness-of-fit tail plots for fitted normal Laplace distribution. Panel (A) shows left tail of empirical and fitted distribution functions and Panel (B) shows right tail of empirical and fitted complementary distribution functions.

The χ^2 test used here for goodness-of-fit follows the method of Vose (Vose [40]). The sample of size $N = 4903$ is divided into the integer part of $(4N)^{\frac{2}{5}}$ groups (here 52), with 51 groups of equal size (here 94) and the 52nd group picking up the rest of the sample. It is important to keep in mind that the χ^2 test depends on the definition of groups.

The resulting χ^2 statistic is 70.747 with $52-4-1 = 47$ degrees of freedom and corresponding p-value of 0.014. Table 7.3 and Table 7.4 break down the calculation of the χ^2 statistic.

The interpretation of the numbers in the Group column is as follows. Denote the number in row $1 \leq i \leq 52$ by g_i. Then Group 1 is $(-\infty, g_1]$, Group 52 is (g_{52}, ∞), and every other Group i is $(g_{i-1}, g_i]$.

Table 7.3. χ^2 Test for Normmal Laplace.

	Group	Observed	Expected
1	-0.0594	94	87.986
2	-0.0444	94	109.533
3	-0.0362	94	109.728
4	-0.0300	94	120.174
5	-0.0265	94	89.587
6	-0.0233	94	97.644
7	-0.0208	94	89.485
8	-0.0187	94	81.473
9	-0.0168	94	84.198
10	-0.0152	94	82.007
11	-0.0137	94	79.858
12	-0.0122	94	85.500
13	-0.0107	94	90.824
14	-0.0095	94	80.020
15	-0.0083	94	91.589
16	-0.0071	94	90.340
17	-0.0059	94	92.672
18	-0.0048	94	99.049
19	-0.0037	94	98.900
20	-0.0027	94	96.108
21	-0.0017	94	99.154
22	-0.0006	94	110.424
23	0.0002	94	91.405
24	0.0011	94	93.789
25	0.0020	94	102.572
26	0.0030	94	108.158

Table 7.4. χ^2 Test for Normal Laplace (Cont.)

	Group	Observed	Expected
27	0.0037	94	91.344
28	0.0046	94	103.895
29	0.0054	94	89.314
30	0.0061	94	85.183
31	0.0069	94	83.308
32	0.0077	94	96.886
33	0.0085	94	94.394
34	0.0095	94	99.893
35	0.0104	94	97.737
36	0.0113	94	96.846
37	0.0120	94	64.708
38	0.0132	94	108.983
39	0.0141	94	83.976
40	0.0151	94	78.204
41	0.0161	94	80.285
42	0.0172	94	82.754
43	0.0187	94	102.493
44	0.0204	94	102.511
45	0.0222	94	94.218
46	0.0242	94	93.736
47	0.0267	94	97.564
48	0.0291	94	81.616
49	0.0322	94	84.875
50	0.0378	94	113.527
51	0.0485	94	123.839
52	0.0485	109	108.731

The R package KScorrect for a Lilliefors-corrected (Braun [8]) Kolmogorov-Smirnov test, when the parameters of the fitted distribution are estimated from the sample, does not support the normal Laplace distribution.

Using a Braun-corrected (Braun [8]) Anderson-Darling test in the R package goftest, when the parameters of the normal Laplace distribution are estimated from the sample, we get p-value of 0.2757.

Using a Braun-corrected (Braun [8]) Cramer-von Mises test in the R package goftest, when the parameters of the normal Laplace distribution are estimated from the sample, we get p-value of 0.9204.

Those results illustrate the low power of some of the available analytical goodness-of-fit tests when the parameters of the fitted distribution are estimated from the sample, and the importance of graphical methods in assessing goodness of fit. See for example Casella and Berger [10], Klugman et al. [25], Delignette-Muller and Dutang [15], and references cited there.

7.7 Generalized Hyperbolic Distribution

The generalized hyperbolic distribution was introduced by Barndorff-Nielsen [2] to describe size distribution of sand particles in deposits created by the wind.

A normal mean-variance mixture is a random variable X that has the representation

$$X = \mu + \gamma Y + \sigma \sqrt{Y} Z$$

where Z is standard normal random variable, Y is non-negative random variable whose distribution is called mixing distribution, Z and Y are independent, μ, γ are constants, and σ is a positive constant.

The distribution of X is called generalized hyperbolic (GH) when the mixing distribution is generalized inverse Gaussian (GIG) with the three-parameter density

$$g(y) = \frac{\left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}}}{2K_\lambda\left(\sqrt{\psi\chi}\right)} y^{\lambda-1} \exp\left[-\frac{1}{2}\left(\psi y + \frac{\chi}{y}\right)\right]$$

The parameters λ, ψ, χ satisfy one of the three cases:

1. $\lambda < 0,\ \psi \geq 0,\ \chi > 0.$

2. $\lambda = 0,\ \psi > 0,\ \chi > 0.$

3. $\lambda > 0,\ \psi > 0,\ \chi \geq 0.$

Special cases of the generalized inverse Gaussian distribution include:

1. Inverse Gaussian distribution when $\lambda = -\frac{1}{2}$.

2. Gamma distribution when $\lambda > 0$ and $\chi = 0$.

3. Inverse gamma distribution when $\lambda < 0$ and $\psi = 0$.

When $\psi > 0$ and $\chi > 0$ the normmal mean-variance mixture variable X has the six-parameter generalized hyperbolic density

$$f(x) = \frac{\left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}}\left[\psi + \left(\frac{\gamma}{\sigma}\right)^2\right]^{\frac{1}{2}-\lambda}}{\sqrt{2\pi\sigma^2}K_\lambda\left(\sqrt{\psi\chi}\right)} \frac{K_{\lambda-\frac{1}{2}}\left\{\sqrt{\left[\chi^2 + \left(\frac{x-\mu}{\sigma}\right)^2\right]\left[\psi + \left(\frac{\gamma}{\sigma}\right)^2\right]}\right\}}{\left\{\sqrt{\left[\chi^2 + \left(\frac{x-\mu}{\sigma}\right)^2\right]\left[\psi + \left(\frac{\gamma}{\sigma}\right)^2\right]}\right\}^{\frac{1}{2}-\lambda}} \exp\left(\gamma\frac{x-\mu}{\sigma}\right)$$

K is modified Bessel function of the second kind.

Call this the $(\lambda, \psi, \chi, \mu, \sigma, \gamma)$ parametrization.

This parametrization has a downside of an identification problem: For any $k > 0$ the densities $f(x;\lambda,\psi,\chi,\mu,\sigma,\gamma)$ and $f\left(x;\lambda,k\psi,\frac{\chi}{k},\mu,k\sigma,k\gamma\right)$ are the same.

For parameter estimation, we need to reduce the number of parameters from six to five. A way that has nummerical advantages for estimation is to require that the mean of the GIG mixing distribution of the random variable Y be one

$$\mathbb{E}(Y) = \sqrt{\frac{\chi}{\psi}}\frac{K_{\lambda+1}\left(\sqrt{\psi\chi}\right)}{K_{\lambda}\left(\sqrt{\psi\chi}\right)} = 1$$

Denote $\alpha = \sqrt{\psi\chi}$, then

$$\psi = \alpha\frac{K_{\lambda+1}(\alpha)}{K_{\lambda}(\alpha)}$$

$$\chi = \frac{\alpha^2}{\psi} = \alpha\frac{K_{\lambda}(\alpha)}{K_{\lambda+1}(\alpha)}$$

7.8 Fitting a Generalized Hyperbolic Distribution to Log Returns

Maximum likelihood estimation using the $(\lambda,\alpha,\mu,\sigma,\gamma)$ parametrization produces the result

$$\begin{aligned}\widehat{\lambda} &= -1.224968\\ \widehat{\alpha} &= 0.4686233\\ \widehat{\mu} &= 0.004847\\ \widehat{\sigma} &= 0.024415\\ \widehat{\gamma} &= -0.003705\end{aligned} \tag{7.23}$$

from which we can also compute the estimators

$$\begin{aligned} \widehat{\psi} &= 0.198124 \\ \widehat{\chi} &= 1.108435 \end{aligned} \tag{7.24}$$

Log-likelihood of the fitted parameters is 11672.97.

AIC of the fitted parameters is -23335.94.

Figure 7.6 shows the Q-Q plot for the fitted generalized hyperbolic distribution.

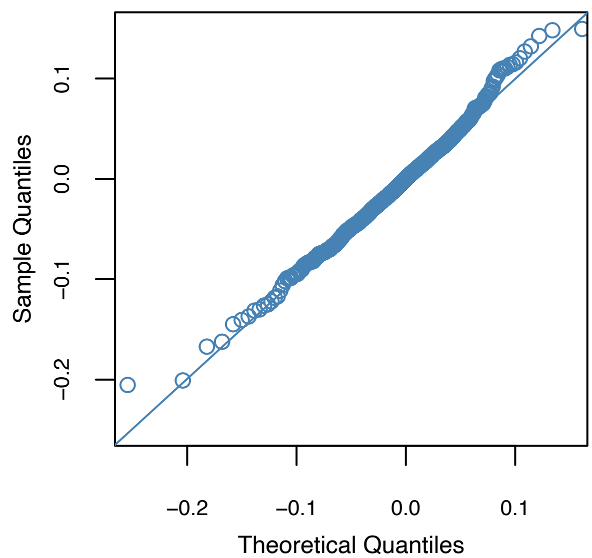

Figure 7.6. Q-Q plot for fitted generalized hyperbolic distribution.

Next, Figure 7.7 shows logarithmic tail plots with the sample (blue circles) and the fitted generalized hyperbolic distribution (black line). The logarithm is base 10 for more intuitive interpretation, that is, a difference of 1 means a probability ratio of 10, a difference of 2 means a probability ratio of 100, and so on.

For a meaningful examination, we need to consider the left tail and the right tail separately, with the examination of the right tail being based on the complementary distribution or survival function.

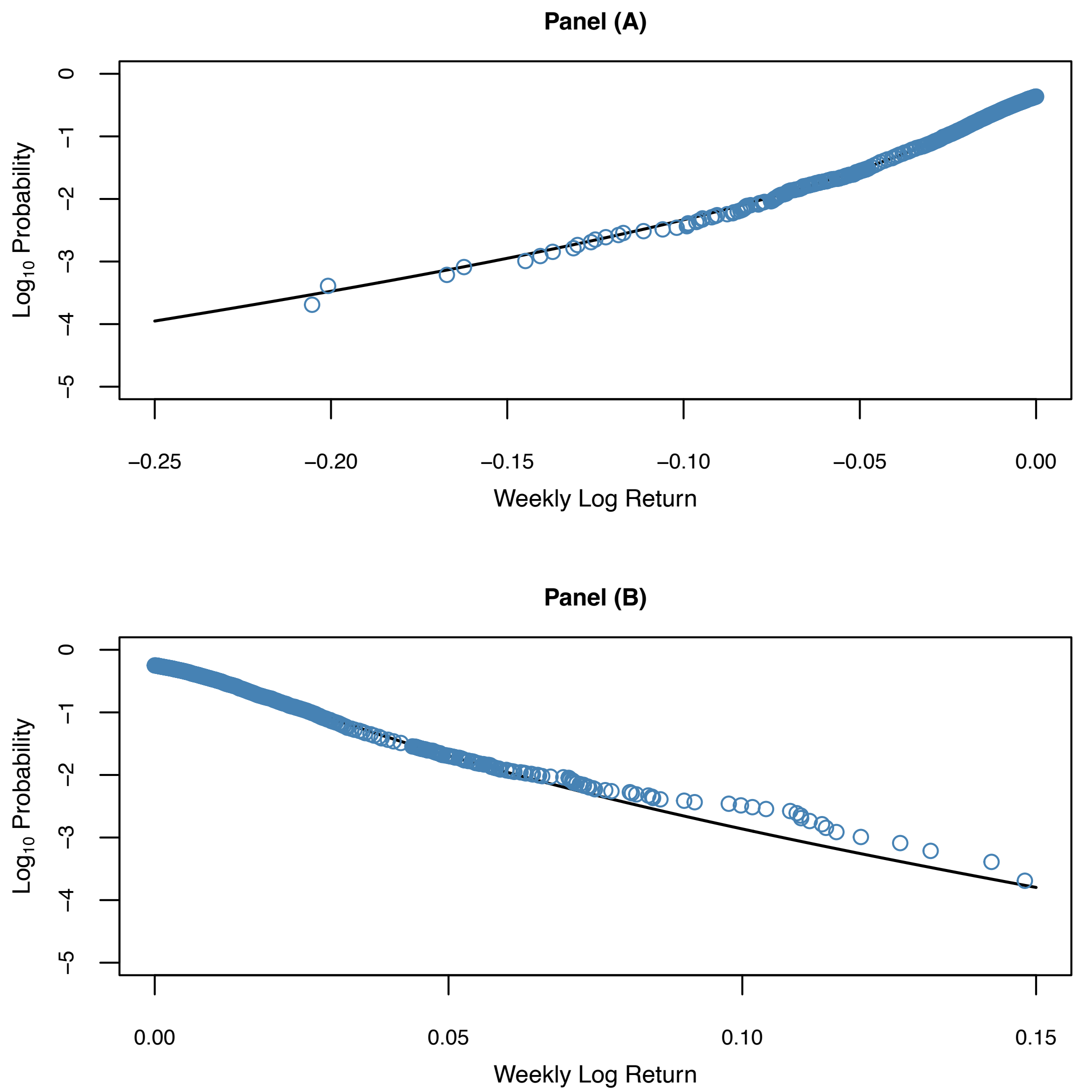

Figure 7.7. Goodness-of-fit tail plots for fitted generalized hyperbolic distribution. Panel (A) shows left tail of empirical and fitted distribution functions and Panel (B) shows right tail of empirical and fitted complementary distribution functions.

The χ^2 test used here for goodness-of-fit follows the method of Vose (Vose [40]). The sample of size $N = 4903$ is divided into the integer part of $(4N)^{\frac{2}{5}}$ groups (here 52), with 51 groups of equal size (here 94) and the 52nd group picking up the rest of the sample. It is important to keep in mind that the χ^2 test depends on the definition of groups.

The resulting χ^2 statistic is 60.250 with $52-5-1 = 46$ degrees of freedom and corresponding p-value of 0.077. Table 7.5 and Table 7.6 break down the calculation of the χ^2 statistic.

The interpretation of the numbers in the Group column is as follows. Denote the number in row $1 \leq i \leq 52$ by g_i. Then Group 1 is $(-\infty, g_1]$, Group 52 is (g_{52}, ∞), and every other Group i is $(g_{i-1}, g_i]$.

Table 7.5. χ^2 Test for Generalized Hyperbolic

	Group	Observed	Expected
1	-0.0594	94	96.384
2	-0.0444	94	89.512
3	-0.0362	94	91.592
4	-0.0300	94	105.310
5	-0.0265	94	82.197
6	-0.0233	94	93.056
7	-0.0208	94	88.208
8	-0.0187	94	82.486
9	-0.0168	94	87.137
10	-0.0152	94	86.426
11	-0.0137	94	85.352
12	-0.0122	94	92.368
13	-0.0107	94	98.872
14	-0.0095	94	87.457
15	-0.0083	94	100.179
16	-0.0071	94	98.600
17	-0.0059	94	100.655
18	-0.0048	94	106.784
19	-0.0037	94	105.608
20	-0.0027	94	101.532
21	-0.0017	94	103.551
22	-0.0006	94	113.869
23	0.0002	94	93.127
24	0.0011	94	94.548
25	0.0020	94	102.326
26	0.0030	94	106.815

Table 7.6. χ^2 Test for Generalized Hyperbolic (Cont.)

	Group	Observed	Expected
27	0.0037	94	89.464
28	0.0046	94	101.076
29	0.0054	94	86.438
30	0.0061	94	82.160
31	0.0069	94	80.189
32	0.0077	94	93.184
33	0.0085	94	90.850
34	0.0095	94	96.369
35	0.0104	94	94.670
36	0.0113	94	94.324
37	0.0120	94	63.376
38	0.0132	94	107.487
39	0.0141	94	83.533
40	0.0151	94	78.383
41	0.0161	94	81.069
42	0.0172	94	84.172
43	0.0187	94	104.998
44	0.0204	94	105.607
45	0.0222	94	97.251
46	0.0242	94	96.509
47	0.0267	94	99.639
48	0.0291	94	82.236
49	0.0322	94	83.960
50	0.0378	94	109.024
51	0.0485	94	114.071
52	0.0485	109	109.013

The R package `KScorrect` for a Lilliefors-corrected (Braun [8]) Kolmogorov-Smirnov test, when the parameters of the fitted distribution are estimated from the sample, does not support the generalized hyperbolic distribution.

Using a Braun-corrected (Braun [8]) Anderson-Darling test in the R package goftest, when the parameters of the generalized hyperbolic distribution are estimated from the sample, we get p-value of 0.9846.

Using a Braun-corrected (Braun [8]) Cramer-von Mises test in the R package goftest, when the parameters of the generalized hyperbolic distribution are estimated from the sample, we get p-value of 0.9860.

Those results should be interpreted with caution because of the low power of some of the available analytical goodness-of-fit tests when the parameters of the fitted distribution are estimated from the sample, but combined with the χ^2 test, the Q-Q plot, the log tail plots, and the relative AIC values, we should not reject the generalized hyperbolic distribution for a long history of weekly log returns.

Tail behavior of the generalized hyperbolic distribution was investigated by von Hammerstein E. A. [39].

Whereas the normal distribution has light tails, and the normal Laplace distribution has exponential tails, the generalized hyperbolic distribution has semi-heavy tails

$$\begin{aligned} f(x) &\approx C(-x)^{\lambda-1}\exp(ax) \quad \text{as} \quad x \to -\infty \\ f(x) &\approx Cx^{\lambda-1}\exp(-bx) \quad \text{as} \quad x \to \infty \end{aligned}$$

where a, b, C are positive constants determined by the parameters of the distribution. Thus, the tails resemble the right tail of a gamma distribution.

Because estimated λ is negative, and $|x| < 1$, the tails are heavier than exponential tails, but less heavy than power tails (tails of a power distribution).

7.9 A Forecasting Model

Many investors, including institutional portfolio managers, think of stock log returns as being approximately normally distributed. That manifests itself when they talk about a large negative return, such as -0.15, and note that it represents six standard deviations (normal probability 10^{-9} and should occur only once every billion weeks or 20 million years.

We can use the filtering model for the conditional distribution of survival times to develop a model that will deliver dynamic conditional distributions of weekly log returns, based on the generalized hyperbolic distribution, from either individual or group forecasts of investors who think in terms of normal distributions.

One obstacle we need to overcome in the modeling is the fact that stock returns are observable, and therefore, weekly return forecasts must converge at the end of the week to the actual log return. I overcame this obstacle by modeling dynamic forecast noise as a Brownian bridge process.

Brownian bridge is a Wiener process restricted to be zero at a specified, deterministic, future time T. Heuristically, a possible description of a Brownian bridge is an Ito process that satisfies the stochastic differential equation

$$dZ_t \quad = \quad \frac{1}{T-t}(0-Z_t)\,dt + dW_t \tag{7.25}$$

In words, Brownian bridge is a mean-reverting Ito process which fluctuates around zero, and whose speed of adjustment goes to infinity as time t goes to the future time T. Rewriting Equation 7.25 we get the defining equation of a Brownian bridge

$$dZ_t = -\frac{Z_t}{T-t}dt + dW_t \tag{7.26}$$

Equation 7.26 is easy to solve. Denote $Y_t = \int_0^t \frac{dW_u}{T-u}$ and apply Ito's formula to the process $Z_t = (T-t)Y_t$. Define the deterministic function $f(Y,t) = (T-t)Y$, then

$$\begin{aligned}
dY_t &= 0 \times dt + \frac{1}{T-t}dW_t \\
df(Y_t,t) &= \left[\frac{\partial f}{\partial t}(Y_t,t) + 0 \times \frac{\partial f}{\partial Y}(Y_t,t) + \frac{1}{2}\frac{\partial^2 f}{\partial Y^2}(Y_t,t)\left(\frac{1}{T-t}\right)^2\right]dt \\
&+ \frac{\partial f}{\partial Y}(Y_t,t)\frac{1}{T-t}dW_t \qquad (7.27) \\
&= -Y_t + 0 \times (T-t) + \frac{1}{2} \times 0 \times \left(\frac{1}{T-t}\right)^2 + (T-t)\frac{1}{T-t}dW_t \\
&= -\frac{Z_t}{T-t}dt + dW_t
\end{aligned}$$

Therefore, the Brownian bridge has the representation

$$Z_t = (T-t)\int_0^t \frac{dW_u}{T-u} \tag{7.28}$$

Figure 7.8 shows a simulated path of Brownian bridge.

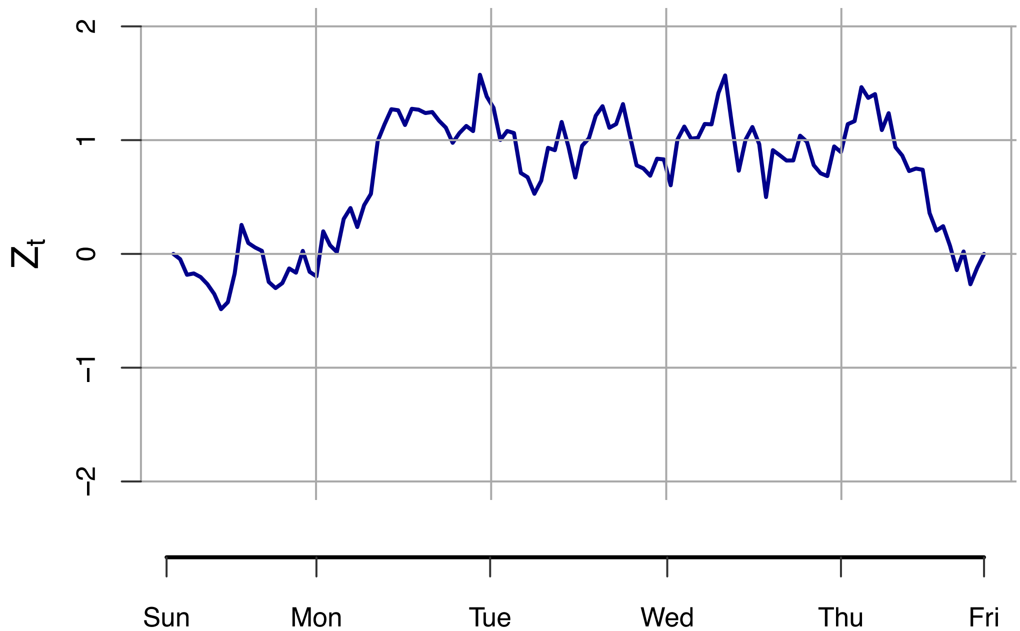

Figure 7.8. Simulated Brownian bridge process. Brownian bridge is a Wiener process constrained to be zero at a specified, non-random, future time.

I model the observable processes of the investors as

$$d\xi_{i,t} \quad = \quad \theta_{n+1}\,dt + S_{i,i}\,dZ_{i,t} \tag{7.29}$$

where θ_{n+1} is the normally distributed generating random variable for the weekly log return $\tau = F^{-1}\left[\Phi\left(\theta_{n+1}\right)\right]$, $1 \leq i \leq n$, n is the number of investors, and the Brownian bridge process $Z_{i,t}$ is

$$dZ_{i,t} \quad = \quad -\frac{Z_{i,t}}{T-t}dt + \frac{1}{S_{i,i}}\sum_{j=1}^{i} B_{i,j}\,dW_{j,t} \tag{7.30}$$

where $S_{i,i}^2 = \sum_{j=1}^{i} B_{1,i,j}^2$. That relies on the fact that if $W_{1,t}, \ldots, W_{n,t}$ are independent Wiener processes and $\alpha_1, \ldots, \alpha_n$ are constants such that $\sum_{i=1}^{n} \alpha_i^2 = 1$ then the linear combination $\sum_{i=1}^{n} \alpha_i W_{i,t}$ is also a Wiener process.
The distribution function F is the fitted generalized hyperbolic distribution of log returns, and Φ is a suitable normal distribution.

The Brownian bridge processes are unobservable, and for consistent notation I will use for them the symbols $\theta_{i,t}$. Thus we have a model with $n+1$ unobservable processes $\theta_{1,t}, \theta_{2,t}, \ldots, \theta_{n,t}, \theta_{n+1,t}$, where the process $\theta_{n+1,t}$ has constant paths because it is a random variable θ_{n+1}. We are interested in estimating the conditional distribution of θ_{n+1} given $\mathcal{F}_t^{\xi}$ for every time $0 < t < T$.

It follows that we have a partially observable process (θ_t, ξ_t) that satisfies the state stochastic differential equations

$$d\theta_t = a_{1,t}\theta_t dt + b_1 dW_{1,t} \tag{7.31}$$

where θ_t is a $(n+1) \times 1$ vector

$$\theta_t = \begin{pmatrix} \theta_{1,t} \\ \vdots \\ \theta_{n+1,t} \end{pmatrix} \tag{7.32}$$

$a_{1,t}$ is a $(n++1) \times (n+1)$ matrix

$$a_{1,t} = \begin{pmatrix} -\frac{1}{T-t} I_n & 0 \\ 0 & 0 \end{pmatrix} \tag{7.33}$$

b_1 is a $(n+1) \times n$ matrix

$$b_1 = \begin{pmatrix} S^{-1} B_1 \\ 0 \end{pmatrix} \tag{7.34}$$

and $W_{1,t}$ is a n-dimensional Wiener process

$$W_t = \begin{pmatrix} W_{1,1,t} \\ \vdots \\ W_{1,n+1,t} \end{pmatrix} \tag{7.35}$$

$$d\xi_t = A_{1,t} \theta_t dt + B_1 dW_{1,t} \tag{7.36}$$

where ξ_t is a $n \times 1$ vector

$$\xi_t = \begin{pmatrix} \xi_{1,t} \\ \vdots \\ \xi_{n,t} \end{pmatrix} \tag{7.37}$$

$A_{1,t}$ is a $n \times (n+1)$ matrix

$$A_{1,t} = \begin{pmatrix} -\frac{1}{T-t}S & 1 \end{pmatrix} \tag{7.38}$$

B_1 is a $n \times n$ matrix given by the lower triangular component of the Cholesky decomposition of the covariance matrix of the forecast markers $\Omega = B_1 B_1'$, and

$$S = \begin{pmatrix} S_{1,1} & 0 \\ \vdots & \vdots \\ 0 & S_{n,n} \end{pmatrix} \tag{7.39}$$

This falls within the general scheme of evolution of observable and unobservable processes given in Equation (6.1) and reproduced here for convenience

$$\begin{aligned} d\theta_t &= (a_0 + a_1\theta_t)\,dt + b_1\,dW_{1,t} + b_2\,dW_{2,t} \\ d\xi_t &= (A_0 + A_1\theta_t)\,dt + B_1\,dW_{1,t} + B_2\,dW_{2,t} \end{aligned} \tag{7.40}$$

The corresponding filtering equations are given in Equation (6.2) and reproduced below

$$\begin{aligned} dm_t &= (a_0 + a_1 m_t)\,dt + \left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)\left(B_1 B_1' + B_2 B_2'\right)^{-1} \\ &\times \left[d\xi_t - (A_0 + A_1 m_t)\,dt\right] \\ d\gamma_t &= \Big[b_1 b_1' + b_2 b_2' + a_1\gamma_t + \gamma_t a_1' \\ &- \left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)\left(B_1 B_1' + B_2 B_2'\right)^{-1}\left(b_1 B_1' + b_2 B_2' + \gamma_t A_1'\right)'\Big]dt \end{aligned} \tag{7.41}$$

7.10 Simulation of the Forecasting Model

The simulation setup has six investors with a positive-definite correlation matrix represented by its upper triangular part for greater readability

$$\rho = \begin{pmatrix} 1 & 0.7 & 0.4 & 0.3 & 0.5 & 0.2 \\ & 1 & 0.4 & 0.6 & 0.8 & 0.3 \\ & & 1 & 0.2 & 0.3 & 0.1 \\ & & & 1 & 0.5 & 0.2 \\ & & & & 1 & 0.6 \\ & & & & & 1 \end{pmatrix} \tag{7.42}$$

The full correlation matrix is

$$\rho = \begin{pmatrix} 1.0 & 0.7 & 0.4 & 0.3 & 0.5 & 0.2 \\ 0.7 & 1.0 & 0.4 & 0.6 & 0.8 & 0.3 \\ 0.4 & 0.4 & 1.0 & 0.2 & 0.3 & 0.1 \\ 0.3 & 0.6 & 0.2 & 1.0 & 0.5 & 0.2 \\ 0.5 & 0.8 & 0.3 & 0.5 & 1.0 & 0.6 \\ 0.2 & 0.3 & 0.1 & 0.2 & 0.6 & 1.0 \end{pmatrix} \tag{7.43}$$

Diagonal matrix of standard deviations is

$$S = \begin{pmatrix} 0.03 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.02 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.03 \end{pmatrix} \tag{7.44}$$

The resulting covariance matrix of is

$$\Omega = S\rho S = \begin{pmatrix} 0.00090 & 0.00105 & 0.00036 & 0.00036 & 0.00030 & 0.00018 \\ 0.00105 & 0.00250 & 0.00060 & 0.00120 & 0.00080 & 0.00045 \\ 0.00036 & 0.00060 & 0.00090 & 0.00024 & 0.00018 & 0.00009 \\ 0.00036 & 0.00120 & 0.00024 & 0.00160 & 0.00040 & 0.00024 \\ 0.00030 & 0.00080 & 0.00018 & 0.00040 & 0.00040 & 0.00036 \\ 0.00018 & 0.00045 & 0.00009 & 0.00024 & 0.00036 & 0.00090 \end{pmatrix} \tag{7.45}$$

Covariance matrix Ω is positive definite because all its eigenvalues are positive.

Cholesky decomposition of Ω gives us the lower triangular matrix B_1 such that $\Omega = B_1 B_1'$

$$B_1 = \begin{pmatrix} 0.030 & 0 & 0 & 0 & 0 & 0 \\ 0.035 & 0.035707 & 0 & 0 & 0 & 0 \\ 0.012 & 0.005041 & 0.027029 & 0 & 0 & 0 \\ 0.012 & 0.021844 & -0.000522 & 0.031282 & 0 & 0 \\ 0.010 & 0.012603 & -0.000131 & 0.000148 & 0.011880 & 0 \\ 0.006 & 0.006721 & -0.000588 & 0.000667 & 0.018107 & 0.022139 \end{pmatrix} \tag{7.46}$$

The sum of squares of each row of the matrix B_1 is a square of the corresponding diagonal entry of the matrix S

$$\sum_{j=1}^{i} B_{1,i,j}^2 = S_{i,i}^2 \tag{7.47}$$

The equations that generate the six individual forecasts are

$$d\xi_t = A_{1,t}\theta_t dt + B_1 dW_{1,t} \tag{7.48}$$

where

$$\xi_t = \begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \\ \xi_{3,t} \\ \xi_{4,t} \\ \xi_{5,t} \\ \xi_{6,t} \end{pmatrix}, \theta_t = \begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \\ \theta_{3,t} \\ \theta_{4,t} \\ \theta_{5,t} \\ \theta_{6,t} \\ \theta_{7,t} \end{pmatrix}, A_{1,t} = \begin{pmatrix} -\frac{S_{1,1}}{T-t} & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -\frac{S_{2,2}}{T-t} & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -\frac{S_{3,3}}{T-t} & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{S_{4,4}}{T-t} & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\frac{S_{5,5}}{T-t} & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -\frac{S_{6,6}}{T-t} & 1 \end{pmatrix},$$

$$W_{1,t} = \begin{pmatrix} W_{1,1,t} \\ W_{1,2,t} \\ W_{1,3,t} \\ W_{1,4,t} \\ W_{1,5,t} \\ W_{1,6,t} \end{pmatrix}$$

and the matrices B_1 and S as above. The Wiener proces $W_{1,t}$ is six-dimensional because $\theta_{7,t} = \theta_7$ is a random variable.

In this simulation, $\theta_{7,t} = 0.04$, which corresponds to a weekly return to the S & P 500, based on the fitted generalized hyperbolic distribution of $\tau = \psi(0.04) = F^{-1}\left[\Phi(0.04)\right] = 0.0340$, where F is the fitted generalized hyperbolic distribution with parameters shown below, and Φ is a normal distribution with mean 0.001 and standard deviation 0.025.

To recall, the parameters of the fitted generalized hypergeometric distribution are

$$
\begin{aligned}
\lambda &= -1.224968 \\
\alpha &= 0.4686233 \\
\mu &= 0.004847 \\
\sigma &= 0.024415 \\
\gamma &= -0.003705
\end{aligned}
\tag{7.49}
$$

The parameters of the normal distribution that characterizes the thinking of the investors can be chosen at will (as long as the standard deviation is positive), and in this simulation I chose them to be equal, respectively, to the mean and the standard deviation of the fitted generalized hyperbolic distribution above.

Table 7.7 shows the correspondence between values of the generating random variable θ_7 and the weekly log return to the S & P 500, based on the fitted generalized hyperbolic distribution of $\tau = \psi(\theta_7) = F^{-1}\left[\Phi(\theta_7)\right]$.

Table 7.7. Correspondence between selected values of the generating random variable θ_7 and the weekly log return τ.

θ_7	τ	θ_7	τ
-0.08	-0.1754	0.00	0.0019
-0.07	-0.1243	0.01	0.0088
-0.06	-0.0860	0.02	0.0159
-0.05	-0.0582	0.03	0.0241
-0.04	-0.0384	0.04	0.0340
-0.03	-0.0243	0.05	0.0467
-0.02	-0.0137	0.06	0.0632
-0.01	-0.0054	0.07	0.0847

Figure 7.9 through Figure 7.14 show the simulated paths of the six forecast markers. The time scale is Sunday for time 0, Monday for time 1, *etc.*, Friday for time 5. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

Figure 7.15 groups the paths of all six forecast markers for ease of comparison.

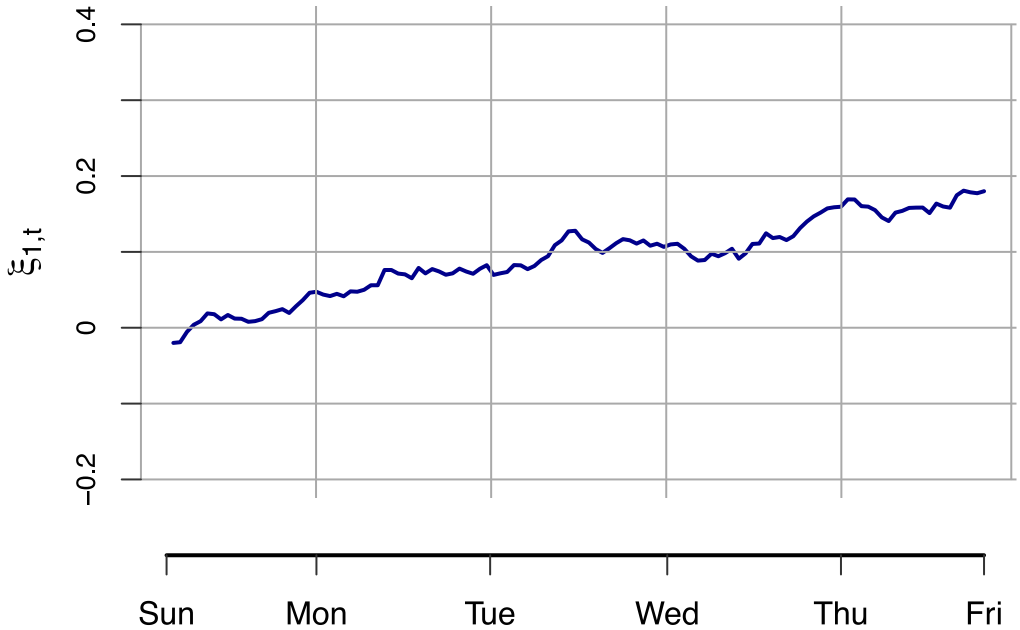

Figure 7.9. Simulated observed path of investor 1. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

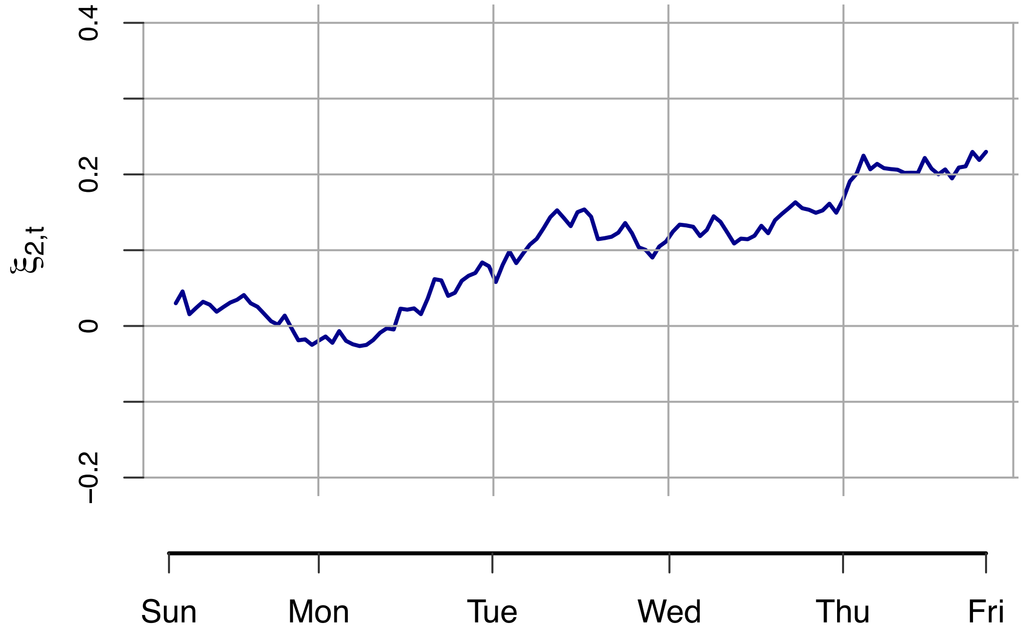

Figure 7.10. Simulated observed path of investor 2. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

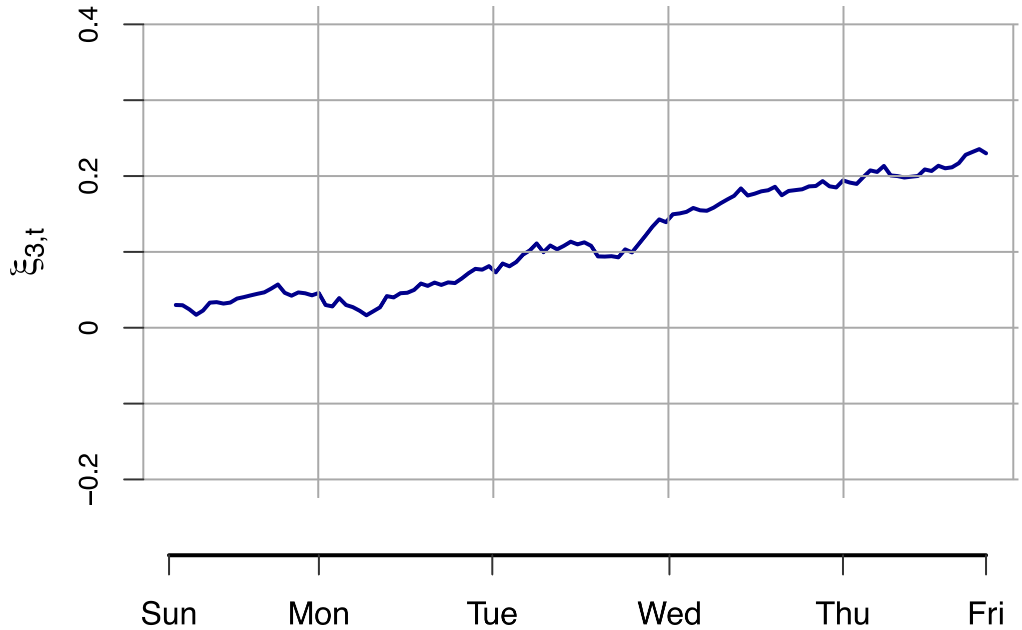

Figure 7.11. Simulated observed path of investor 3. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

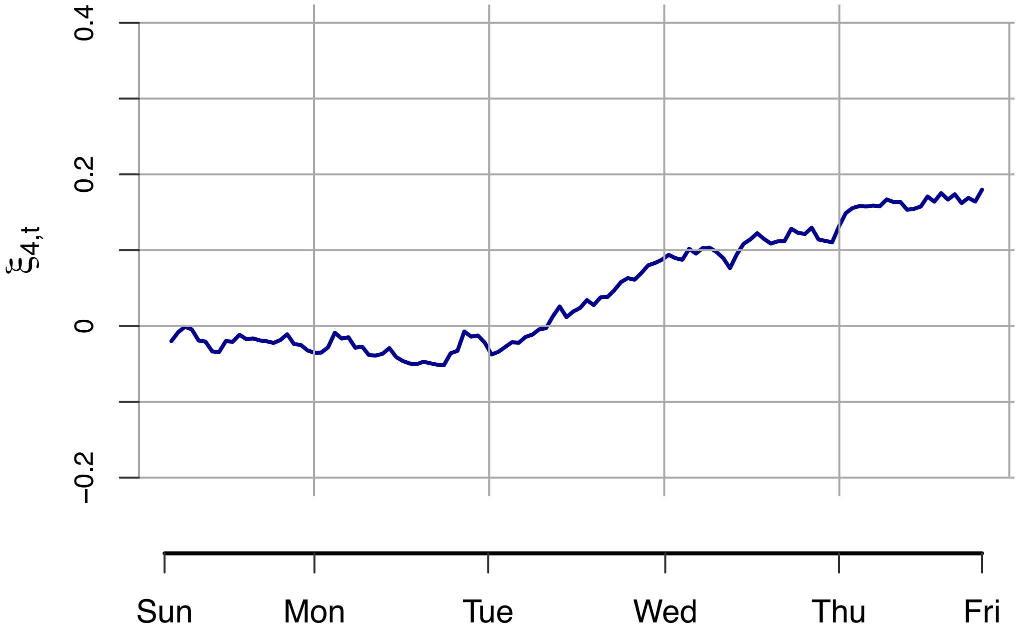

Figure 7.12. Simulated observed path of investor 4. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

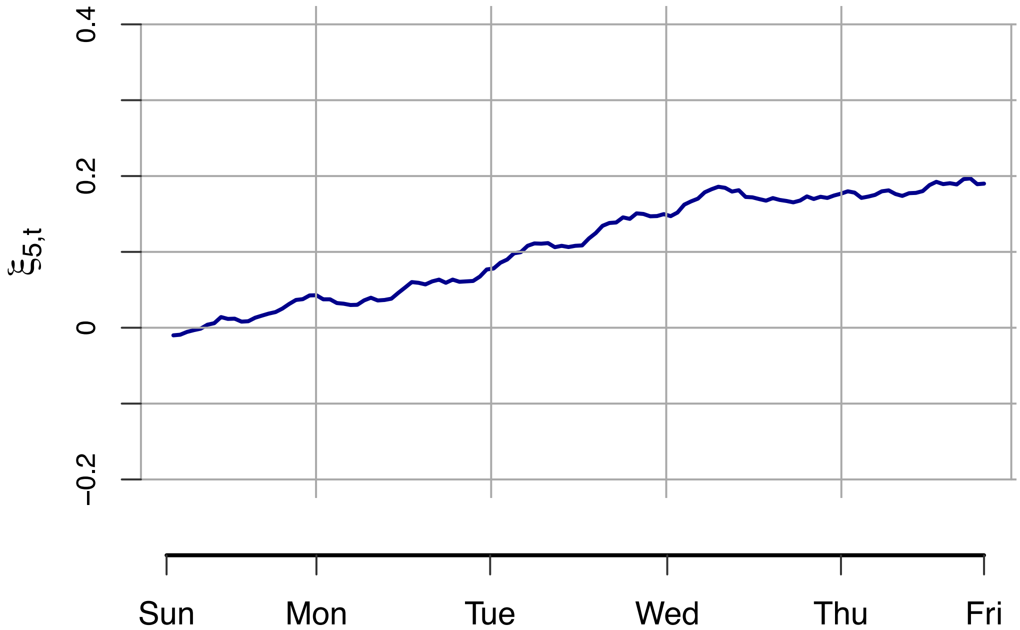

Figure 7.13. Simulated observed path of investor 5. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

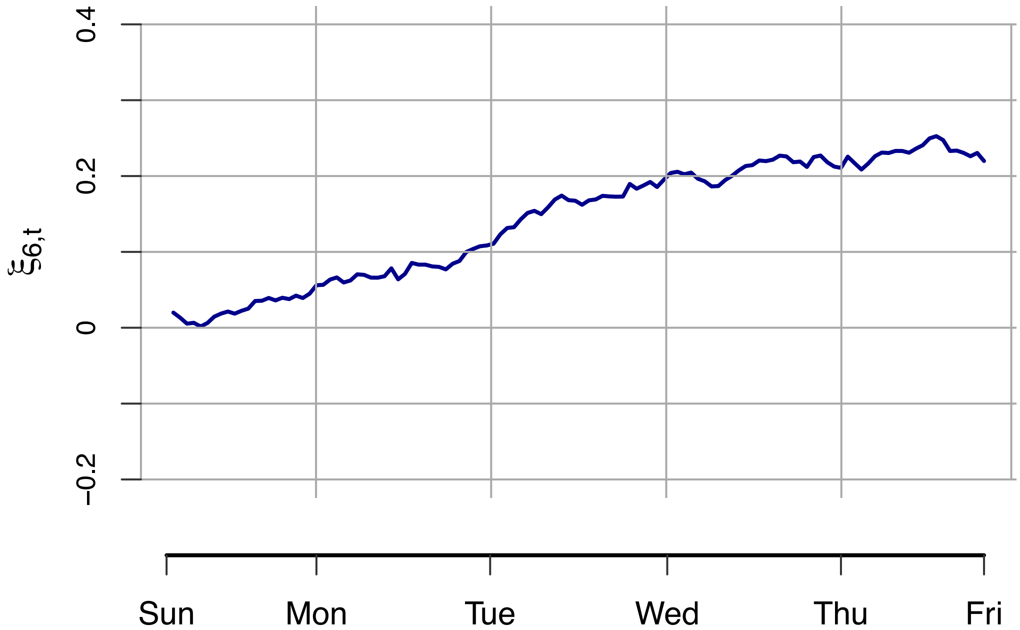

Figure 7.14. Simulated observed path of investor 6. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

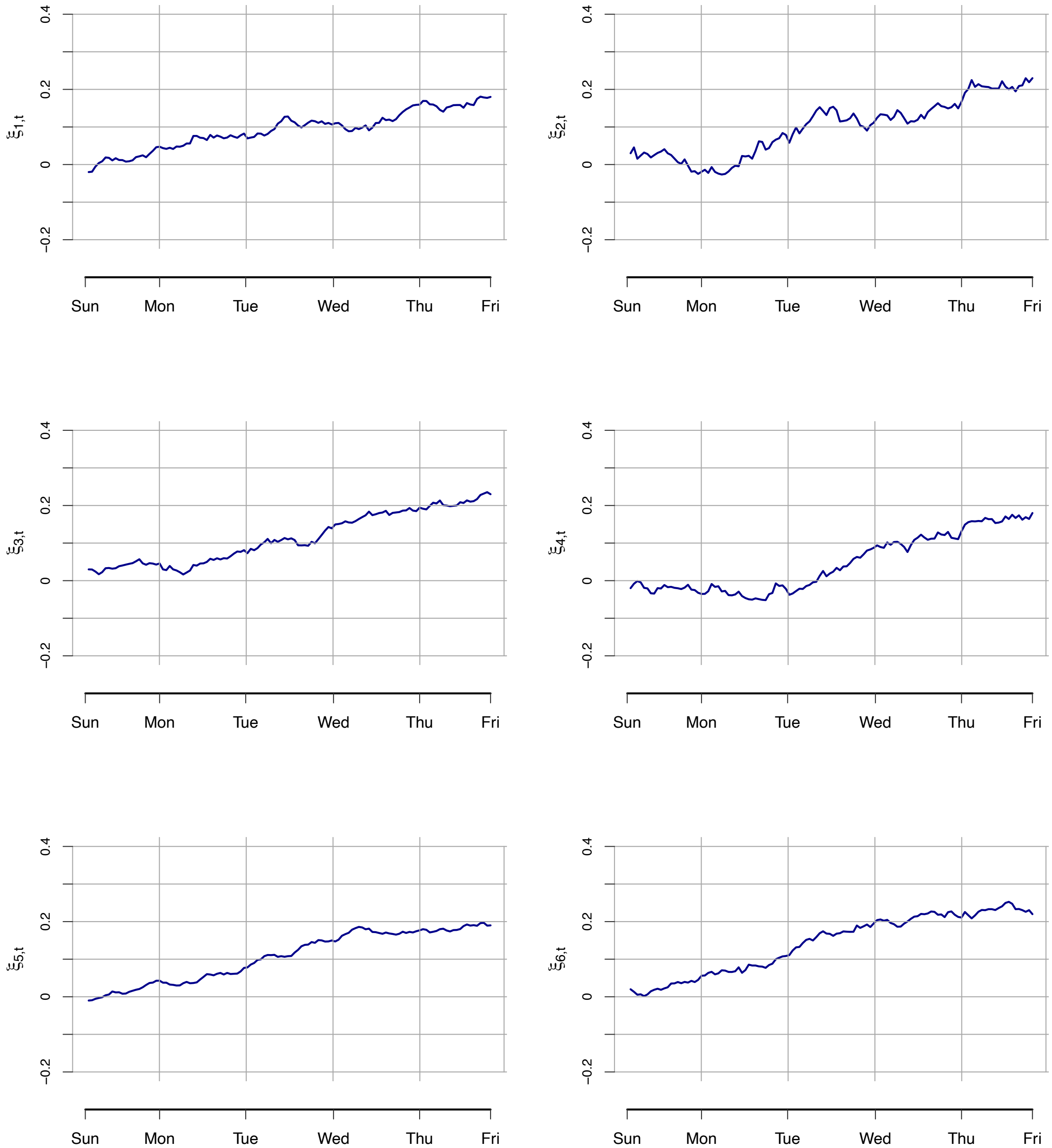

Figure 7.15. Simulated observed paths of all six investors. At each time t on the horizontal axis, the forecasted value of the generating random variable θ_7 is the slope of a straight line from the initial point of the forecast path to the point on the the forecast path at time t.

Next, with six investors we have a seven-dimensional unobservable process θ_t. The first six components $\theta_{1,t},\ldots,\theta_{6,t}$, of the vector θ_t are the six Brownian bridge processes that represent the forecast noise of the forecasts, and the seventh component of θ_t is the generating random variable $\theta_{7,t} = \theta_7$.

Each of the six Brownian bridge processes $\theta_{i,t}$, $1 \le i \le 6$ evolves as

$$\begin{aligned} d\theta_{i,t} &= -\frac{\theta_{i,t}}{T-t}dt + dW_{1,i,t} \\ &= \frac{1}{T-t}(0-\theta_{i,t})\,dt + dW_{1,i,t} \end{aligned} \tag{7.50}$$

The seventh component, the generating random variable $\theta_{7,t} = \theta_7$ evolves as

$$d\theta_{7,t} = 0dt + 0dW_{1,t} \tag{7.51}$$

Therefore, the vector process θ_t evolves as

$$d\theta_t = a_{1,t}\theta_t dt + b_1 dW_{1,t} \tag{7.52}$$

where

$$
\theta_t = \begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \\ \theta_{3,t} \\ \theta_{4,t} \\ \theta_{5,t} \\ \theta_{6,t} \\ \theta_{7,t} \end{pmatrix}, \; a_{1,t} = \begin{pmatrix} -\frac{1}{T-t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{T-t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{T-t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{T-t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{T-t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{T-t} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}
$$

and
$$
b_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{B_{2,1}}{S_{2,2}} & \frac{B_{2,2}}{S_{2,2}} & 0 & 0 & 0 & 0 \\ \frac{B_{3,1}}{S_{3,3}} & \frac{B_{3,2}}{S_{3,3}} & \frac{B_{3,3}}{S_{3,3}} & 0 & 0 & 0 \\ \frac{B_{4,1}}{S_{4,4}} & \frac{B_{4,2}}{S_{4,4}} & \frac{B_{4,3}}{S_{4,4}} & \frac{B_{4,4}}{S_{4,4}} & 0 & 0 \\ \frac{B_{5,1}}{S_{5,5}} & \frac{B_{5,2}}{S_{5,5}} & \frac{B_{5,3}}{S_{5,5}} & \frac{B_{5,4}}{S_{5,5}} & \frac{B_{5,5}}{S_{5,5}} & 0 \\ \frac{B_{6,1}}{S_{6,6}} & \frac{B_{6,2}}{S_{6,6}} & \frac{B_{6,3}}{S_{6,6}} & \frac{B_{6,4}}{S_{6,6}} & \frac{B_{6,5}}{S_{6,6}} & \frac{B_{6,6}}{S_{6,6}} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.
$$

It can be shown that the matrix b_1 is the lower triangular component of the Cholesky decomposition of the correlation matrix ρ with a row of zeros added at the bottom.

The filtering equation for conditional mean is

$$
dm_t \quad = \quad a_{1,t} m_t dt + \left(b_1 B_1' + \gamma_t A_{1,t}'\right)\left(B_1 B_1'\right)^{-1} (d\xi_t - A_{1,t} m_t dt) \tag{7.53}
$$

A heuristic comparison for jointly normally distributed random vectors θ and ξ

$$\mathbb{E}\left(\theta \middle| \xi\right) \quad = \quad \mathbb{E}(\theta) + \operatorname{cov}(\theta, \xi)\left[\operatorname{var}(\xi)\right]^{-1}\left[\xi - \mathbb{E}(\xi)\right] \tag{7.54}$$

The filtering equation for conditional variance is

$$d\gamma_t \quad = \quad \left[a_{1,t}\gamma_t + \gamma_t a'_{1,t} + b_1 b'_1 - \left(b_1 B'_1 + \gamma_t A'_{1,t}\right)\left(B_1 B'_1\right)^{-1}\left(b_1 B'_1 + \gamma_t A'_{1,t}\right)'\right]dt \tag{7.55}$$

A heuristic comparison for jointly normally distributed random vectors θ and ξ

$$\operatorname{cov}\left(\theta \middle| \xi\right) \quad = \quad \operatorname{var}(\theta) - \operatorname{cov}(\theta, \xi)\left[\operatorname{var}(\xi)\right]^{-1}\left[\operatorname{cov}(\theta, \xi)\right]' \tag{7.56}$$

Equation (7.53) and Equation (7.55) can be converted to approximating difference equations and solved numerically.

I used initial value $m_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ m_{7,0} \end{pmatrix}$, where $m_{7,0} = \dfrac{\xi_{1,0} + \xi_{2,0} + \xi_{3,0} + \xi_{4,0} + \xi_{5,0} + \xi_{6,0}}{6}$.

I also used the initial value $\gamma_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{0,7,7} \end{pmatrix}$

where $\gamma_{0,7,7} = \sqrt{\frac{1}{6}\sum_{i=1}^{6} S_{i,i}^2}$.

Figure 7.16 shows the resulting path of $m_{7,t}$, which quickly converges to small fluctuations around the true value $\theta_7 = 0.04$.

Figure 7.17 shows the conditional variance $\gamma_{7,t}$. Because the vector (θ_t, ξ_t) is normally distributed, the conditional variance γ_t is a function of time only and does not depend on the observed values of the vector ξ_t.

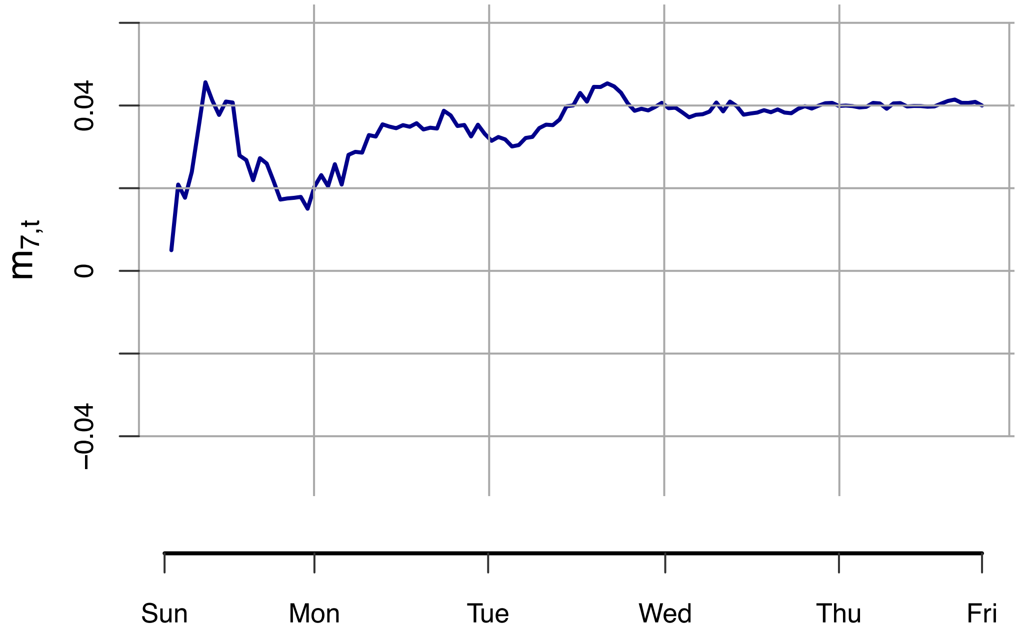

Figure 7.16. Path of the conditional mean $m_{7,t}$ for the unobservable log return for the week under consideration.

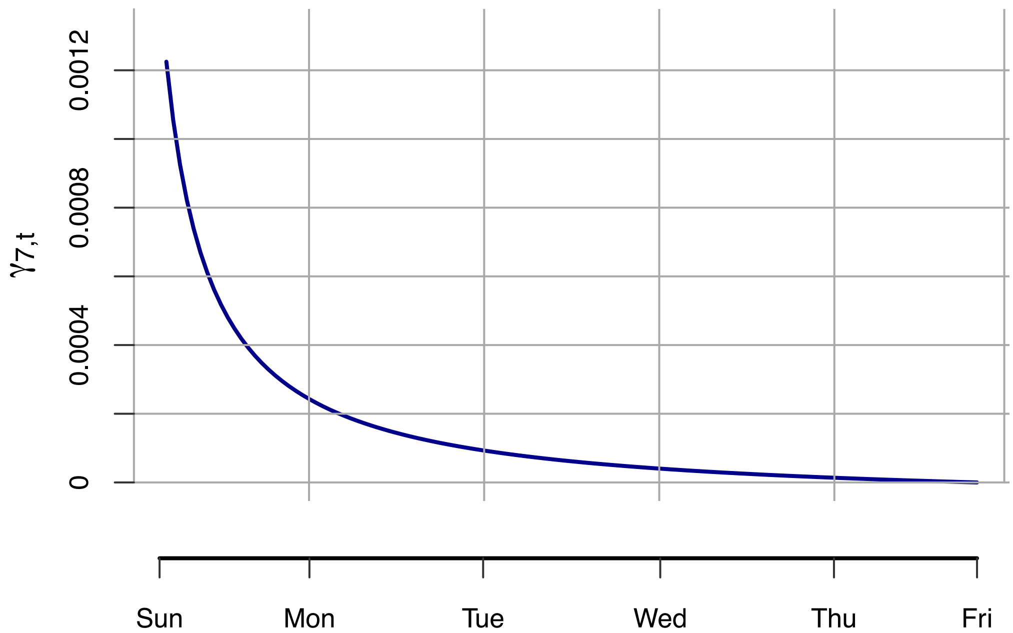

Figure 7.17. Conditional variance $\gamma_{7,7,t}$ for the unobservable log return for the week under consideration.

For example, at the end of Wednesday, conditional mean is $m_{7,t} = 0.0396$ and conditional variance is $\gamma_{7,7,t} = 4.2295 \times 10^{-5}$.

We can can calculate the conditional distribution of weekly log return $\tau = \psi(\theta)$

$$\begin{aligned} \mathbb{P}\left(\tau \leq y \middle| \mathcal{F}_t^{\xi}\right) &= \mathbb{P}\left[\psi(\theta) \leq y \middle| \mathcal{F}_t^{\xi}\right] \\ &= \mathbb{P}\left[\theta \leq \psi^{-1}(y) \middle| \mathcal{F}_t^{\xi}\right] \\ &= \Phi\left[\frac{\psi^{-1}(y) - m}{\sqrt{\gamma}}\right] \end{aligned}$$

where Φ is standard normal distribution, $\psi^{-1}(y) = \widetilde{\Phi}^{-1}\left[F(y)\right]$, and $\widetilde{\Phi}$ is the fitted normal distribution used by the investors.

We also have

$$\mathbb{P}\left(\tau > y \middle| \mathcal{F}_t^{\xi}\right) = \Phi\left[\frac{m - \psi^{-1}(y)}{\sqrt{\gamma}}\right]$$

If $y = 0.034$ we have $\psi^{-1}(0.034) = 0.04$ and we have the conditional probability

$$\begin{aligned}\mathbb{P}\left(\tau > 0.034 \middle| \mathcal{F}_t^{\xi}\right) &= \Phi\left(\frac{0.0396 - 0.04}{\sqrt{4.2295 \times 10^{-5}}}\right) \\ &= 0.4755\end{aligned}$$

For comparison, the unconditional (marginal) probability using fitted generalized hyperbolic distribution is

$$\mathbb{P}(\tau > 0.034) = 0.060$$

8 Conclusion

This chapter offers a summary of the contributions, and describes opportunities for future work.

The central statistical problem of survival analysis is to determine and characterize the conditional distribution of a survival time given a history of some observed health markers.

The dissertation presents two families of models in which the health markers evolve randomly over time in a manner that can be represented by Ito stochastic processes, and the survival time is modeled as a suitably chosen deterministic function of a random variable, called here a generating random variable, that is related to the random evolution of the health markers. The deterministic function is chosen to give the survival time the desired distribution function.

In the first family of models, the generating variable of the survival time is an Ito integral over the positive half-line, with the observable health marker at any given time represented by the same integral up to that time.

More precisely, in the first family of models, the health marker is a solution of a specified stochastic differential equation, chosen so that the solution process of the stochastic differential equation converges to a finite random variable when time goes to infinity, and the distribution of this limit random variable depends on the history of the solution process.

In that setup, the dissertation provides formulas for the conditional distribution function of the survival time, given the observed history of the health marker, and its relatives: the conditional survival function and the conditional hazard function. The relevance of this family of models is demonstrated via a simulation study and a comparison of the new model with the traditional proportional hazards model.

A limitation of the first family of models presented in this dissertation is that they involve one health marker, which can represent a single symptom or treatment or a weighted average index of symptoms and treatments. A potential way to remove this limitation is discussed below in paragraphs describing future research opportunities. In addition, the second family of models does not have this limitation.

The second family of models involves a linear filtering framework, in which the generating variable, or a vector of generating variables, linearly affects a number of observable health markers that evolve as Ito processes.

The generating variable, or the vector of generating variables, is unobservable, and its conditional mean and variance, given the health markers, are estimated by stochastic filtering methods. Because the conditional distribution of the generating variables, given the history of the health markers, is normal, conditional mean and variance determine the conditional distribution. In addition, the generating random variables can be limits of unobservable Ito processes, called here generating processes, that influence the evolution of the health markers.

To demonstrate the relevance of this model the dissertation describes the results of a simulation study with a four-dimensional vector of health markers.

The usefulness of the filtering framework is not limited to computing conditional distributions of survival times – instead of a positive survival time we can use a return to a financial asset which can be positive or negative.

To apply the model in that context, the dissertation fits a generalized hyperbolic distribution to weekly log returns to the S & P 500 stock market index.

There are a number of investors who observe individual forecast markers and use them to form a joint forecast of the weekly log return. Because the weekly log return, unobservable during the week, is observable at the end of the week, the modeling of forecasting noise is done through a multi-dimensional Brownian bridge process.

In the filtering models considered in this dissertation, the conditional variance is a deterministic function of time. This is a limitation, but may be overcome through an enlargement of the set of stochastic differential equations used to model the health markers, and/or an enlargement of the set of stochastic differential equations used to model the generating processes.

The modeling frameworks developed in this dissertation offer several avenues for further work and development.

In the context of the first family of models, further development may involve the characterization of the conditional distribution of survival time that arises from an alternative specification of the stochastic differential equation that describes the evolution of the health marker.

In addition, although the health markers that are used in this family of models can be a weighted average of several symptoms and/or treatments, a more general setup would allow several correlated health marker processes. Such extension can be accomplished by a linear transformation of the several generating variables associated with the several health markers, and using the fact that we know the distribution of a product of a finite number of independent uniform random variables on the unit interval.

In the second family of models, further development could include the introduction of observable measures of treatment, such as blood level of a drug that affects the evolution of one or more of the generating processes.

Another possible extension is a multi-dimensional model of generating processes and/or generating variables, applied to the problem of modeling competing risks.

Another extension of the filtering model of survival times is a model with non-linear dependence of the health markers on the generating processes or generating variables.

In the application to investment analysis, the fitting of the generalized hyperbolic distribution to log returns suggests that the variance of log returns has a generalized inverse Gaussian distribution. With that assumption, the return forecasting model can be modified into a variance forecasting model, keeping in mind that the variance of stock log returns is never observable.

References

[1] O. O. Aalen, O. Borgan, and H. K. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer, 2008.

[2] O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London Series A*, 353: 401–419, 1977.

[3] T. Besedes and T. J. Prusa. On the duration of trade. Working Paper 9936, National Bureau of Economic Research, September 2003.

[4] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69:89–95, 2001.

[5] A. N. Borodin and P. Salminen. *Handbook of Brownian Motion – Facts and Formulae*. Birkhauser Verlag, 2002.

[6] T. B. Brakenhoff, M. Mitroiu, R. H. Keogh, K. G. M. Moons, R. H. H. Groenwold, and M. van Smeden. Measurement error is often neglected in medical literature: a systematic review. *Journal of Clinical Epidemiology*, 98:89–97, 2018.

[7] T. B. Brakenhoff, M. van Smeden, F. L. J. Visseren, and R. H. H. Groenwold. Random measurement error: Why worry? an example of cardiovascular risk factors. *PLoS One*, 2018. DOI: 10.1371/journal.pone.0192298.

[8] H. Braun. A simple method for testing goodness-of-fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society*, 42:53–63, 1980.

[9] R. S. Bucy and P. D. Joseph. *Filtering for stochastic processes with applications to guidance*. John Wiley & Sons, 1968.

[10] G. Casella and R. L. Berger. *Statistical Inference*. Cengage Learning, 2001.

[11] D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC, 2015.

[12] D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

[13] D. R. Cox. Partial likelihood. *Biometrika*, 62:543–558, 1975.

[14] G. E. Dallal and L. Wilkinson. An analytic approximation to the distribution of Lilliefors' test for normality. *The American Statistician*, 40:294–296, 1986.

[15] M. L. Delignette-Muller and C. Dutang. Frequently asked questions. Vignette for R package fitdistrplus, 2022.

[16] B. Efron. The efficiency of Cox's likelihood function for censored data. *Journal of American Statistical Association*, 72:557–565, 1977.

[17] F. Emmert-Streib and M. Dehmer. Introduction to survival analysis in practice. *Machine Learning and Knowledge Extraction*, 1:1013–38, 2019.

[18] L. D. Fisher and D. Y. Lin. Time-dependent covariates in the Cox proportional hazards regression model. *Annual Review of Public Health*, 20:145–157, 1999.

[19] A. Friedman. *Stochastic Differential Equations and Applications*. Academic Press, 1975.

[20] D. Gamerman and M. West. An application of dynamic survival models in unemployment studies. *The Statistician*, 36:269–274, 1987.

[21] I. I. Gihman and A. V. Skorohod. *Stochastic Differential Equations*. Springer Verlag, 1972.

[22] J. G. Ibrahim, M. H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, 2001.

[23] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis Of Failure Time Data*. John Wiley & Sons, Inc., 2002.

[24] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.

[25] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss Models: From Data to Decisions*. Wiley, 2019.

[26] H. W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62:399–402, 1967.

[27] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes I & II*. Springer, 2001.

[28] E. Pitacco. Survival models in actuarial mathematics: From Halley to longevity risk. 7th International Congress on Insurance: Mathematics and Economics, 2003.

[29] M. Rausand, A. Barros, and A. Hoyland. *System Reliability Theory: Models, Statistical Methods, and Applications*. Wiley, 2020. Third Edition.

[30] W. J. Reed. The normal-Laplace distribution and its relatives. In N. Balakrishnan, J. M. Sarabia, and E. Castillo, editors, *Advances in Distribution Theory, Order Statistics, and Inference*, pages 61–74. Birkhuser, 2006.

[31] W. J. Reed and M. Jorgensen. The double Pareto lognormal distribution: A new parametric model for size distribution. *Communications in Statistics Theory and Methods*, 33:1733–1753, 2004.

[32] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer Verlag, 1994.

[33] S. L. Spruance, J. E. Reid, M. Grace, and M. Samore. Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*, 48:2787–92, 2004.

[34] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

[35] M. A. Stephens. Tests based on edf statistics. In R. B. D'Agostino and M. A. Stephens, editors, *Goodness-of-fit techniques*. Dekker, 1986.

[36] M. Tableman and J. S. Kim. *Survival Analysis Using S: Analysis of Time-to-Event Data*. Chapman & Hall / CRC, 2003.

[37] C. C. Tadeu, P. Zebrowski, and M. Wildemeersch. Assessing the time intervals between economic recessions. *PLoS ONE*, 15, 2020.

[38] H. C. Thode. *Testing for normality*. Dekker, 2002.

[39] von Hammerstein E. A. Tail behaviour and tail dependence of generalized hyperbolic distributions. In J. Kallsen and A. Papapantoleon, editors, *Springer Proceedings in Mathematics & Statistics*, pages 3–40. Springer, 2016.

[40] D. Vose. *Risk analysis, a quantitative guide*. Wiley, 2000.

[41] W. H. Wong. Theory of partial likelihood. *The Annals of Statistics*, 14:88–123, 1986.

[42] J. Wu, F. W. Crawford, D. A. Kim, and D. Stafford. Exposure, hazard, and survival analysis of diffusion on social networks. *Stat Med.*, 37:2561–85, 2018.