Parameter estimation in several classes of non-Markovian random processes defined by Stochastic Differential Equations

by

Robert Charles Reiner, Jr.

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Anna Amirdjanova, Co-Chair
Professor Vijay Nair, Co-Chair
Professor Tailen Hsing
Assistant Professor Erhan Bayraktar

In loving memory of my father, Robert Charles Reiner, Sr 9/7/5

ACKNOWLEDGEMENTS

I would like to thank all of the people who have so greatly effected my life over the last five years. Most importantly my wife and best friend Mandy Izzo. It isn't clear what or where I would be without her, but it is obvious to everyone who knows both of us that that place would not be healthy; emotionally, physically nor sanitarily. My mother Colleen Reiner has provided continual support that has kept me on a path that has so far worked out quite well and she too has had a great influence on the man I have become. My sister, Colette Verkuil continues to awe me with her achievements and makes me try to be me best.

While I have made many friends since arriving in Ann Arbor, my two compatriots, Jason Goldstick and Adam Rothman have been with me every step of the process and this program would have been a much quieter, more boring, though undoubtedly more sober place. I owe a large part of my sanity to both of them.

My two advisors have offered continual, valuable advice and two different but greatly influential outlooks. Anna Amirdjanova always reminds me of the beauty of mathematics while Vijay Nair points out the equal beauty in applicability. They have both shaped me in countless ways.

Finally, though I lost him the third day of my first semester here in Ann Arbor, my father, Robert Reiner Sr, will always be both my toughest critic as well as my biggest supporter. His words and opinions will live on in me for the rest of my days.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

CHAPTER I

Introduction

1.1 Stochastic differential equations

This thesis consists of two topics devoted to parameter estimation of solutions of stochastic evolution equations driven by Gaussian processes. The two problems, however, use vastly different inference methods: maximum likelihood approach derived by stochastic calculus techniques for Volterra processes in the first problem and method of moments and simulation-based tools in the second problem. We study certain stochastic differential equations of the form

$$dX_t = f(X_t, Y_t, t, \theta_1)dt + g(X_t, Y_t, t, \theta_2)dY_t,$$

where (Y_t) is a given Gaussian process with known covariance kernel, and f and g are some known "drift" and "volatility" functions which depend on unknown parameters of interest (θ_1, θ_2). In this general form, the model can be endowed with a very rich and flexible structure (both in terms of memory properties and the shape of finite-dimensional distributions), yet allows for a short and intuitive dynamical description, making it attractive for potential applications.

In this work, we first consider the case of a non-semimartingale Y driving the dynamics of X, where Y is a Gaussian random field with covariance structure of the form $\int_0^t \int_0^s K(t,u)K(s,v)dudv$ for a general Volterra kernel K. Next we study a monotone increasing integral functional of a standard Brownian motion, which can be formally regarded as a solution to the degenerate stochastic differential equation with $g \equiv 0$. In both cases the resulting process X_t is generally non-Markovian, which makes the problems interesting from mathematical viewpoint and useful in many applications where Markov assumption is impractical.

Volterra processes are one of the most recent additions to the field of continuous Gaussian processes and represent generalizations of the popular fractional Brownian motion (fBm), originally introduced by Kolmogorov under the name of a Wiener spiral. However, the properties and use of the fBm as a modelling tool were relatively unknown until Mandelbrot's seminal 1968 paper [37], in which the fBm received its modern name. Fractional Brownian motion's important difference from the classical Brownian motion is the generally non-Markovian nature of the fBm. Each fBm has an associated Hurst parameter, H (named after hydrologist Harold Hurst), where $H \in (0,1)$ and the correlations between increments of the fBm are defined to die out at the polynomial rate (H controls the degree of the polynomial decay). When $H < \frac{1}{2}$, fractional Brownian motion exhibits short-range dependence, when $H > \frac{1}{2}$, it exhibits long-range dependence, and $H = \frac{1}{2}$ corresponds to the standard Brownian motion case.

Further generalizations, bifractional [22, 53, 58] and multifractional [7] Brownian motions, have been made to the fractional Brownian motion to allow for multidimensional and time-varying Hurst index. However, a much more powerful generalization of all of the above processes, called Volterra process, has appeared in the literature relatively recently. Volterra process is defined as a stochastic integral of a time-dependent deterministic kernel with respect to a standard Brownian motion, where the form of the integral (Volterra) kernel can be kept very general.

On the other hand, researchers have long been interested in the study of certain integral functionals of Brownian motion in connection with various physical, financial and engineering applications. For example, studies of the integral of the absolute value of a standard Brownian motion go back to the work of Cameron and Martin [9] and Kac [24]. It is interesting that the famous Feynman-Kac formula, relating solutions of certain parabolic partial differential equations to Laplace transforms of the solutions to stochastic differential equations, was introduced by Kac in 1949 [25] while trying to describe the distribution of the integral of the absolute value of Brownian motion. Apart from analytical beauty, the formula found immense computational uses for analysts and practitioners alike. More recently, a similar integral functional of Brownian motion was studied by Lachal and distributions of the integral process and of the first exit time from a bounded interval were derived (see [31, 32] for details).

1.2 Stochastic Calculus

While Brownian motion and its generalizations have been around for quite some time, the mathematical framework for stochastic calculus has only been around for the past 60 years. From the initial work of Itô and Skorohod, there have been many advancements. While the calculus for Brownian motion is well understood, it was not until the turn of the millennia [60, 2] that fractional Brownian motion's stochastic calculus began to be placed on solid mathematical footing. Since that time, stochastic integration with respect to fractional Brownian motion has been developed by many authors (see for example [4, 20]).

Most of the work done on Volterra process has taken place over the last decade. From the introduction of stochastic calculus for general Gaussian processes [16, 42, 3, 14, 36, 51], the general theory has advanced parallel to that of fractional Brownian motion (which seems to be a driving force behind a lot of the development). More recent work on stochastic integration with respect to Gaussian processes [15, 39, 10] has done much to complete the basic framework. Volterra processes themselves were introduced in their current form by Decreusefond in 2002 [13]. Since then, equivalence of processes [5, 6], integration [23, 30], and simulation [23] have all been discussed.

A major tool that is used throughout this thesis is that of Russo-Vallois calculus [54]. This concept of integration and quadratic variation has seen recent applications to finance, and seems, at the moment, to be the only tractable way to advance. The Russo-Vallois calculus has been used in stochastic differential equations driven by fractional Brownian motion [55, 18, 19, 17], and can be used in the extension to stochastic differential equations that are driven by Volterra processes.

1.3 Stochastic differential equations driven by Volterra processes

Stochastic differential equations (SDEs) arise naturally in many physical and biological experiments. In fact, Brown's original observations of a particle suspended in liquid are best described using the famous Langevin equation (rather than the Brownian motion itself) due to friction, or viscosity, of the fluid. However, until as recently as 2002 non-semimartingale structure of the fractional Brownian motion prevented researchers from studying stochastic differential equations driven by fractional Brownian motion, until first results establishing existence and uniqueness of strong solutions to these equations appeared in [43].

Extension of the above theory of stochastic differential equations to the case of SDEs driven by Volterra processes is the next natural step towards development of non-martingale tools needed for successful study of complex dynamical processes seen in practice. In that direction, in Chapter II, using Russo-Vallois calculus, we prove the existence and uniqueness of solutions to the SDEs driven by a general class of Volterra processes.

1.4 Parameter estimation for stochastic differential equations

A natural question that arises when observing any processes is statistical inference. The estimation of a drift coefficient in stochastic differential equations is one of particular interest. Estimation in stochastic differential equations driven by Brownian motion is well surveyed in [48]. The estimation of stochastic differential equations driven by fractional Brownian motion has also been solved in many cases [27, 8, 28, 49]. Again, however, the general problem, using Volterra processes, has not been investigated (with the exception of a degenerate differential equation [23]).

In Section 2.4 we derive maximum likelihood estimators of a drift parameter, derive their properties and partially address the practical concern that solutions of most SDEs are observed only at discrete times rather than on a continuous scale. Intuition for suitable time-discretized versions of the estimator naturally builds on the road map designed by Neuenkirch and Nourdin in [41] but requires departure from the use of many useful identities valid under fractional differentiation and fractional integration of the integral kernel associated with the fBm. In Chapter III we study the SDE dynamics and parameter estimation in the multiparameter setting, when the driving process is a Volterra random field.

1.5 Analysis of a certain integral functional of Brownian motion

Chapter IV is devoted to the study of an integral functional of a standard Brownian motion (B_t) of the form

$$X_t = \int_0^t (B_s^2)^\theta ds,$$

where θ is an (unobserved) parameter. This choice of the process is motivated by physical properties of many degradation processes (in the absence of catastrophic failures) which have continuous and monotone increasing random trajectories. In many applications one is interested in estimating the time to failure of various devices, such as time to cross some threshold, $D > 0$. It is natural to study the "time to failure" random variable T_D defined by

$$T_D := \inf\{t > 0 : X_t = D\}.$$

We first estimate θ based on observing several paths of the process X, and then estimate the entire distribution of T_D through simulation.

1.6 Moment estimation methods

While maximum likelihood estimation is the gold standard of frequentist inference procedures, there are many circumstances where a simple approach of matching sample moments with analytically known variable moments is preferable. In some situations, the computation of the likelihood of the data can be an arduous task impossible without computers, while the matching of moments can be elementary. In other cases, like those we will be dealing with, the computation of the likelihood is intractable. Since we will be dealing with integral functionals of Brownian motion, the computation of the likelihood proves to difficult. However, since Brownian motion is a simple Markovian Gaussian process, deconstructing the integral to compute moments will prove not as challenging. In addition to method of moment methods, we will recall the generalized method of moment estimation procedure of Hansen [21]. This method is perfect for one of the data setups we consider in that we have two moment conditions but only a single parameter to be estimated. Since the parameter is over-identified, we utilize an optimal distance metric to determine the most efficient estimate of the parameter of all estimates made be the two moment conditions.

All of the moment conditions rely on the law of large numbers to ensure that the sample averages will converge to the average of the random variable. All of the method of moment estimators also rely on the observations being independent and identically distributed to compute these sample averages. In specialized cases we consider a situation where all of the processes are observed at the same time. For these cases we use the method of moments estimators. However, a more general case where the observations are not made all at the same time is also considered. We still assume independence of observations, but now the observations clearly do not have the same distribution. For this situation we introduce a new estimator based on the more general Kolmogorov law of large numbers. This estimator, called the asymptotic method of moments estimator, similarly relies on the fact that only for the true value of the parameter will the sample average converge to population average, but in this case, the sample observations are the difference between each observation of a path and the expected value of that observation. We find that this new estimator performs quite well even for small samples.

1.7 Outline of the thesis

In Chapter 2 we introduce Volterra processes as well as stochastic differential equations driven by them. After developing an estimator of a drift parameter, we establish several asymptotic properties of the estimator as well as indicate a discretized estimator whose convergence would depend on knowledge of the specific form of the Volterra processes covariance. Chapter 3 extends the results in Chapter 2 to the multi-parameter setting. After establishing existence and uniqueness of solutions to differential equations driven by a multi-parameter Volterra process, an estimator is introduced and a form of consistency is proven. In Chapter 4 we consider an integral functional of Brownian motion and estimate several quantities of interest associated with it. We introduce the concept of an asymptotic method of moments estimator to allow us to utilize the moment conditions derived for a general data observation setup. Finally, in Chapter 5 we conclude the thesis by describing both advantages and shortcomings of the work as well as future directions for both projects.

CHAPTER II

Parameter estimation in one-dimensional Stochastic Differential Equations

2.1 Introduction

Advancements in parameter estimation for stochastic differential equations driven by Gaussian processes has always been directly preceded by related advancements in stochastic calculus. While the stochastic calculus and parameter estimation in the case of Brownian motion noise has been well studied (see for example [45] and the references therein), the stochastic calculus for fractional Brownian motion was not well established until the late 1990's (see for example [60, 2]). Once this extension of stochastic calculus was made, parameter estimation of stochastic differential equations driven by fractional Brownian motion followed relatively quickly ([27, 29, 26, 44]). Interestingly though, existence and uniqueness results of solutions to these stochastic differential equations were developed after the initial parameter estimation work ([43]).

Just as fractional Brownian motion generalized Brownian motion greatly in that it allowed for non-Markovian dynamics, Volterra processes allow for considerably more flexibility than fractional Brownian motion. While both Brownian motion and fractional Brownian motion are examples of Volterra processes, the fairly general conditions on the kernels that define Volterra processes through an integral relationship to Brownian motion allow for much more complex Gaussian processes. Volterra processes themselves were introduced in their current form by Decreusefond in 2002 [13]. The latest extensions of stochastic calculus to general Gaussian processes [16, 42, 3, 14, 36, 51] and specifically the more recent work on stochastic integration with respect to Gaussian processes [15, 39, 10] has allowed for more analysis to be done with Volterra processes. Since then, equivalence of processes [5, 6, 46], integration [23, 30], and simulation [23] have all been investigated.

In this chapter we discuss a stochastic differential equation in one dimension driven by a general class of Gaussian processes, estimation of a drift parameter in the equation as well as properties of this estimator. We first introduce Volterra processes, the stochastic calculus we use to work with these processes, establish results on the existence and uniqueness to the stochastic differential equations that are driven by Volterra processes as well as give conditions, critical for our parameter estimation method, that martingales associated with a given class of Volterra processes exist. In Section 2.4 we define our maximum likelihood estimator of a drift parameter for stochastic differential equations of a given general form based on continuous observations of a path, and then we establish asymptotic results for this estimator in certain specific cases. Finally, we lay a groundwork towards estimating the parameter in a more general scenario where the path of the processes is observed only on a discretized mesh.

2.2 Preliminaries

In this chapter, we will adopt the following notation :

- f'_x will represent $\frac{df}{dx}$
- The function $\mathbb{1}$ will represent the function that is identically equal to 1.

The extension of the stochastic calculus we will use requires growth conditions of the random processes, namely Hölder continuity.

Definition II.1. A function $\phi : [t_1, t_2] \subset \mathbb{R} \to \mathbb{R}$ where $t_1 < t_2$, is Hölder continuous of index $\alpha \in \mathbb{R}_+$ (also called α-Hölder continuous) if

i) ϕ is continuous.

ii) $\|\phi(\cdot)\|_{[t_1,t_2],\alpha} < \infty$ where

$$\|\phi(\cdot)\|_{[t_1,t_2],\alpha} = \sup_{s_1 \neq s_2,\, t_1 \leq s_1, s_2 \leq t_2} \frac{|\phi(s_1) - \phi(s_2)|}{|s_1 - s_2|^\alpha}.$$

We let $\mathscr{C}_\alpha([t_1, t_2])$ denote the space of all α-Hölder continuous functions on the interval $[t_1, t_2]$.

Directly following Decreusefond [13], we will use the following framework for the 1 parameter Volterra processes. Let us consider $\mathbb{T} = [0, T] \subset \mathbb{R}_+$ or $\mathbb{T} = \mathbb{R}_+$. Fix $\gamma > 0$. Let $\|\cdot\|_2$ be the standard norm in $L^2(\mathbb{T})$. Assume $K^\gamma : \mathbb{T} \times \mathbb{T} \to [0, \infty)$ is a deterministic function such that the following three conditions hold:

(C1) $K^\gamma(0, t) = 0$ for all $t \in \mathbb{T}$ and $K^\gamma(t, t') = 0$ for $t < t'$.

(C2) There exist constants C and $\gamma > 0$ such that for all $t, t' \in \mathbb{T}$

$$\int_{\mathbb{T}} \left(K^\gamma(t, s) - K^\gamma(t', s)\right)^2 ds \leq C \|t - t'\|_2^{2\gamma}.$$

(C3) K^γ is injective as a transformation of functions in $L^2(\mathbb{T})$:

$$(K^\gamma f)(t) = \int_{\mathbb{T}} K^\gamma(t, s) f(s) ds, \quad f \in L^2(\mathbb{T}).$$

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathbf{P})$ be a filtered, complete probability space. We then define 1 parameter Volterra processes as follows.

Definition II.2 (Volterra Process)**.** A mean-zero, 1-parameter Gaussian random field, $V^\gamma = \{V_t^\gamma\}_{t\in\mathbb{T}}$ with covariance

$$R(t,t') = \mathbb{E}\left[V_t^\gamma V_{t'}^\gamma\right] = \int_{\mathbb{T}} K^\gamma(t,s)K^\gamma(t',s)ds. \tag{2.1}$$

where kernel K^γ satisfies conditions (C1)-(C3) is called a 1-parameter Volterra process.

We assume our probability space is large enough to allow us to express the process V^γ as

$$V_t^\gamma(\omega) = \int_{\mathbb{T}} K^\gamma(t,s)d\mathbb{B}_s(\omega). \tag{2.2}$$

where $\mathbb{B} = \{\mathbb{B}_t\}_{t\in\mathbb{T}}$ is a standard 1-dimensional Brownian process. Below, several examples of Volterra processes are provided.

Example 2.2.1 (Standard Brownian motion)**.** Let

$$K(t,t') = \mathbb{1}_{[0,t]}(t').$$

Then V^γ, with $\gamma = \frac{1}{2}$, is a standard Brownian motion with covariance

$$R(t,t') = t \wedge t'.$$

Example 2.2.2 (1-dimensional fractional Brownian motion)**.** Let

$$K(t,t') = \frac{(t-t')^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} F\left(\frac{1}{2}-H, H-\frac{1}{2}, H+\frac{1}{2}, 1-\frac{t}{t'}\right) \mathbb{1}_{[0,t]}(t'),$$

where (see [1]) $F(a,b,c,t)$ is the Gauss hypergeometric function. Then V^γ, with $\gamma = H$, is fractional Brownian motion with Hurst parameter $H \in (0,1)$ and covariance

$$R(t,t') = \frac{1}{2}\left(t'^{2H} + t^{2H} - |t-t'|^{2H}\right).$$

Example 2.2.3 (1-dimensional Ornstein-Uhlenbeck process)**.** Let

$$K(t,t') = \mathbf{e}^{\theta(t-t')}\mathbb{1}_{[0,t]}(t').$$

Then V^γ, with $\gamma = \frac{1}{2}$, is the Ornstein-Uhlenbeck process, i.e.,

$$dV_t^\gamma = \theta V_t^\gamma dt + d\mathbb{B}_t.$$

Example 2.2.4 (Multifractal Brownian motion)**.** Let $H \in C^r(\mathbb{R};(0,1))$ with $\sup_{t\in\mathbb{R}} H(t) < r$, for some $r \in (0,1)$ and

$$K(t,t') = \frac{1}{C_1 H(t)}\left((t-t')_+^{H(t)-\frac{1}{2}} - (-t')_+^{H(t)-\frac{1}{2}}\right), \quad t,t' \in \mathbb{R}.$$

Then V^γ, with $\gamma = \inf_{t\in\mathbb{R}} H(t)$, is the multifractal Brownian motion of Benassi et al [7].

Example 2.2.5 (Logarithmic Brownian motion)**.** Let $\beta > 0$ and define

$$\epsilon^2(r) = \frac{\beta}{r}\left[\log\left(\frac{1}{r}\right)\right]^{-\beta-1},$$

and

$$K(t,t') = \epsilon(t-t')\mathbb{1}_{[0,t]}(t') \quad t,t' \in [0,1).$$

Then V^γ is the logarithmic Brownian motion of Mocioalca and Viens [39]. This is a very interesting Volterra process since, for example when $\beta = 1$, it can exhibit longer range dependence than <u>any</u> fractional Brownian motion with any Hurst index.

By condition (C1), V^γ is adapted to the natural filtration of $\mathbb{B}$. Condition (C2) implies that $K(t,\cdot) \in L^2(\mathbb{T})$ for all $t \in \mathbb{T}$ and thus V^γ is well defined. Using Kolmogorov's criterion, we see that condition (C2) also guarantees that there exists a Hölder continuous modification of V^γ of index β for all $\beta < \gamma$. To emphasize this property, we denote a Volterra process which has a Hölder continuous modification of index less than or equal to γ by V^γ.

The primary notion of stochastic integration we use is that of Russo-Vallois [54]. Since, in general, Volterra processes are not semimartingales, the standard stochastic calculus does not apply. We first define *ucp* convergence.

Definition II.3. A family of processes $\left(H_t^{(\epsilon)}\right)$ converge to the process (H_t) in the uniform convergence in probability on each compact interval (*ucp*) sense if

$$\sup_{t\in[0,T]} \left|H_t^{(\epsilon)} - H_t\right| \xrightarrow{\mathbf{P}} 0 \quad \text{as } \epsilon \to 0 \qquad \forall\, T \geq 0,$$

denoted as:

$$\underset{\epsilon\longrightarrow 0}{\text{lim-ucp}}\ H_t^{(\epsilon)} = H_t.$$

Now, we define the forward integral as follows (the backward, symmetric integrals and brackets are similarly defined and can be found in [54]):

Definition II.4. Let $X = (X_t)_{t\in[0,T]}$ and $Y = (Y_t)_{t\in[0,T]}$ be two stochastic processes with continuous paths. The (Russo-Vallois) forward integral $\int_0^t Y_s d^- X_s$, is defined by

$$\underset{\epsilon\longrightarrow 0}{\text{lim-ucp}}\ \epsilon^{-1}\int_0^t Y_s \left(X_{(s+\epsilon)\wedge T} - X_s\right) ds$$

provided the limit exists. Similarly, the (Russo-Vallois) backward integral $\int_0^t Y_s d^+ X_s$ is defined by

$$\underset{\epsilon\longrightarrow 0}{\text{lim-ucp}}\ \epsilon^{-1}\int_0^t Y_s \left(X_s - X_{(s-\epsilon)\vee 0}\right) ds,$$

and the (Russo-Vallois) symmetric integral $\int_0^t Y_s d^\circ X_s$ is defined by

$$\underset{\epsilon\longrightarrow 0}{\text{lim-ucp}}\ (2\epsilon)^{-1}\int_0^t \left(Y_{(s+\epsilon)\wedge T} + Y_s)\right) \left(X_{(s+\epsilon)\wedge T} - X_s\right) ds$$

provided each limit exists.

Due to the Hölder continuity of Volterra precesses, we will be able to relate the integral of Russo-Vallois to that of Young ([59]) defined here.

Definition II.5 (Young's integral)**.** We say that the integral

$$\int_t^{t'} f(s)dg(s)$$

exists, with value I, in the sense of Young if the sum

$$\sum_{i=1}^{N} f(s_i)\left(g(t_i) - g(t_{i-1})\right)$$

where $t = t_0 \leq s_1 \leq t_1 \leq \ldots \leq t_{N-1} \leq s_N \leq t_N = t'$, differs from I by at most ϵ when all the lengths $t_i - t_{i-1}$ are less than δ where $\epsilon \to 0$ as $\delta \to 0$

Critical to the estimation method we will employ is the existence of a martingale associated with a given Volterra process. The following martingale representation result, analogous to the one found in [34] for fractional Brownian sheets, defines these martingales as well as establishes a sufficient condition for their existence.

Theorem II.6. *Let $X = (X_t)_{t\in\mathbb{T}}$ be a continuous mean-zero 1-parameter Gaussian random process with the covariance function R, i.e.*

$$R(t,t') = \mathbb{E}\left[X_t X_{t'}\right].$$

For arbitrary continuous curves $\mathcal{C} : \mathbb{T} \to \mathbb{R}$ and $\forall t \in \mathbb{T}$, suppose that there exists a family of kernels $k_{\mathcal{C}}^t : [0,t) \to \mathbb{R}$ such that

$$\iint_{[0,t)\times[0,t')} k_{\mathcal{C}}^t(s) k_{\mathcal{C}}^{t'}(s') R(ds, ds') = \int_{[0,t\wedge t')} \mathcal{C}(s) k_{\mathcal{C}}^{t\wedge t'}(s) ds, \quad \forall t,t' \in \mathbb{T} \tag{2.3}$$

where $a \wedge b := \min(a,b)$. Define the process $N^{\mathcal{C}} = (N_t^{\mathcal{C}})_{t\in\mathbb{T}}$ by:

$$N_t^{\mathcal{C}} = \int_{[0,t)} k_{\mathcal{C}}^t(s) dX_s. \tag{2.4}$$

Then $N^{\mathcal{C}} = (N_t^{\mathcal{C}})_{t\in\mathbb{T}}$ is a 1-parameter Gaussian martingale with variance

$$\left\langle N^{\mathcal{C}} \right\rangle_t = \int_{[0,t)} \mathcal{C}(s) k_{\mathcal{C}}^t(s) ds. \tag{2.5}$$

Moreover, for all $\mathcal{C}, \widetilde{\mathcal{C}} \in C(\mathbb{T}, \mathbb{R})$,

$$N_t^{\widetilde{\mathcal{C}}} = \int_{[0,t)} q^{\mathcal{C},\widetilde{\mathcal{C}}}(s) dN_s^{\mathcal{C}} \tag{2.6}$$

where

$$q_t^{\mathcal{C},\widetilde{\mathcal{C}}} = \frac{d\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t}{d\left\langle N^{\mathcal{C}}\right\rangle_t} \tag{2.7}$$

with

$$\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t = \iint_{[0,t)^2} k_{\mathcal{C}}^t(v) k_{\widetilde{\mathcal{C}}}^t(s') R(ds, ds'). \tag{2.8}$$

Proof.

It is clear that $(N_t^{\mathcal{C}}, t \in \mathbb{T})$ is a centered Gaussian process starting from 0 at $t = 0$. From Eq. (3.1) and Eq. (3.10), we have

$$\begin{aligned}
\mathbb{E}\left[N_t^{\mathcal{C}} N_{t'}^{\mathcal{C}}\right] &= \iint\limits_{[0,t)\times[0,t')} k_{\mathcal{C}}^t(s) k_{\mathcal{C}}^{t'}(s') R(ds, ds') \\
&= \int_{[0,t\wedge t')} \mathcal{C}(s) k_{\mathcal{C}}^{t\wedge t'}(s) ds \\
&= \left\langle N^{\mathcal{C}}\right\rangle_{t\wedge t'}.
\end{aligned}$$

Thus $N_z^{\mathcal{C}}$ has independent increments, and $(N_t^{\mathcal{C}}, t \in \mathbb{T})$ is a 1-parameter Gaussian martingale.

Now, let $\mathcal{C}, \widetilde{\mathcal{C}} \in C(\mathbb{T}, \mathbb{R})$ be arbitrary. Then, since both

$$N_t^{\mathcal{C}} = \int_{[0,t)} k_{\mathcal{C}}^t(s) dX_s$$

and

$$N_t^{\widetilde{\mathcal{C}}} = \int_{[0,t)} k_{\widetilde{\mathcal{C}}}^t(s) dX_s$$

are martingales, defining f, dependent on both $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ as

$$f(t) = \frac{d\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t}{d\left\langle N^{\mathcal{C}}\right\rangle_t} \quad \text{a.s.}$$

we have necessarily that

$$N_t^{\widetilde{\mathcal{C}}} = \int_{[0,t)} f(s) dN_s^{\mathcal{C}}$$

and to show the dependence of f on both $\mathcal{C}$ and $\widetilde{\mathcal{C}}$, we denote

$$f(t) = q^{\mathcal{C},\widetilde{\mathcal{C}}} = \frac{d\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t}{d\left\langle N^{\mathcal{C}}\right\rangle_t}.$$

□

2.3 Existence and uniqueness of solutions

In this section we will prove the existence and uniqueness of stochastic differential equations driven by Volterra processes in the case where the Hölder continuity index, γ, is in $(\frac{1}{2}, 1]$.

First, recall a useful result for ordinary differential equations.

Theorem II.7 ([55]). *Let b, $\sigma : [0,\infty) \times \mathbb{R} \to \mathbb{R}$, and $g \in \mathscr{C}_\gamma(\mathbb{R})$, where $\frac{1}{2} < \gamma \leq 1$. Suppose that b is globally Lipschitz in t and x, and $\sigma \in \mathscr{C}_1(\mathbb{R})$ with σ, σ_t', σ_t' globally Lipschitz in t and x. Then for every $T > 0$ and $\gamma > \beta > 1 - \gamma$, the ordinary differential equation*

$$dx(t) = b(t, x(t))dt + \sigma(t, x(t))dg(t) \quad t \in (0, T), \qquad (2.9)$$

$$x(0) = x_0,$$

has a unique solution in $\mathscr{C}_\beta([0,T])$, where the integration is in the framework of Young [59].

We also recall the following proposition which relates Russo-Vallois calculus to that of Young.

Proposition II.8 ([52])**.** *Let X, Y be two real processes indexed by $[0,T]$ whose paths are, respectively, a.s. in $\mathscr{C}_\alpha([0,T)]$ and $\mathscr{C}_\beta([0,T])$, with $\alpha, \beta > 0$ and $\alpha+\beta > 1$. Then the three integrals $\int_0^\cdot Y d^+X$, $\int_0^\cdot Y d^-X$, and $\int_0^\cdot Y d^\circ X$ exist and coincide with the Young integral $\int_0^\cdot Y dX$*

We can now state the main result of this section.

Theorem II.9. *Let b, $\sigma : [0,\infty) \times \mathbb{R} \to \mathbb{R}$, and V^γ a Volterra process where $\frac{1}{2} < \gamma < 1$. Suppose b is globally Lipschitz in t and x, and $\sigma \in \mathscr{C}_1(\mathbb{R})$ with σ, σ_t', and σ_x' globally Lipschitz in t and x. Then for every $T > 0$ and $\gamma > \beta > 1-\gamma$, the stochastic differential equation*

$$dX_t(\omega) = b(t, X_t(\omega))dt + \sigma(t, X_t(\omega))d^-V_t^\gamma(\omega), \qquad t \in [0,T] \tag{2.10}$$

$$X_0(\omega) = Z(\omega)$$

has a unique solution in $\mathscr{C}_\beta([0,T)]$ with probability 1.

Proof. First, we note that since $V^\gamma \in \mathscr{C}_\beta([0,T])$ for every index $\beta < \gamma$, and σ is Hölder continuous with index 1, $\beta + 1 > 1$, and therefore by Proposition II.8

$$\int_0^t \sigma(s, X_s)d^-V_s^\gamma = \int_0^t \sigma(s, X_s)dV_s^\gamma$$

almost surely, where the last integral is Young's integral. Then the result follows directly from Theorem II.7 applied path-wise. □

2.4 Maximum Likelihood estimation

This section concerns estimation of a drift parameter for a stochastic differential equation of the following form:

$$dX_t = A(t, X_t, \theta)dt + \sigma(t)dV_t^\gamma \tag{2.11}$$

where V^γ is a 1-parameter Volterra process of index γ, $\theta \in \Theta \subset \mathbb{R}$ and $\sigma(t)$ is a positive non-vanishing function on $[0, T]$.

2.4.1 Maximum likelihood estimator when drift is a polynomial in θ.

Theorem II.10. *Under the assumptions of Theorem II.6 and Theorem II.9, let V^γ be a 1-parameter Volterra process with covariance function R, where*

$$R(t, t') = \mathbb{E}\left[V_t^\gamma V_{t'}^\gamma\right] = \int_0^{t\wedge t'} K^\gamma(t, s) K^\gamma(t', s) ds$$

where $K^\gamma(t, s)$, $0 \le s < t \le T$ is a deterministic kernel. Define the process $X = (X_t)_{t\in[0,T]}$ by the equations

$$dX_t = A(t, X_t, \theta)dt + \sigma(t)dV_t^\gamma, \quad t \in (0, T)$$

$$X_0 = \xi \text{ a.s.}$$

where $A(t, x, \theta) = \sum_{i=0}^p a_i(t, x)\theta^i$ and σ (a positive non-vanishing function on $[0, T]$) are known functions but $\theta \in \Theta \subset \mathbb{R}$ is unknown. Assuming that the function $k_{\mathbb{1}}^t(s)$, defined by Eq. (3.1) with $\mathcal{C}(s) = 1$, $\forall s$, is smooth enough so that $\frac{k_{\mathbb{1}}^t(s)}{\sigma(s)} \in \mathscr{C}_\beta(\mathbb{R})$ where $\beta + \gamma > 1$, the maximum likelihood estimator, $\hat{\theta}_T$, of θ is given by:

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\operatorname{argmax}} \int_0^T \left(\sum_{i=0}^p J_i(t)\theta^i\right) dU_t - \frac{1}{2}\int_0^T \left(\sum_{i=0}^p J_i(t)\theta^i\right)^2 d\left\langle N^*\right\rangle_t \tag{2.12}$$

where

$$\left\langle N^*\right\rangle_t = \int_{[0,t)} k_{\mathbb{1}}^t(s) ds, \tag{2.13}$$

$$U_t = \int_{[0,t)} \frac{k_{\mathbb{1}}^t(s)}{\sigma(s)} dX_s, \tag{2.14}$$

$$J_i(t) = \frac{d}{d\left\langle N^*\right\rangle_t} \int_{[0,t)} k_{\mathbb{1}}^t(s) \frac{a_i(s, X_s)}{\sigma(s)} ds \quad i \in \{0, 1, \dots, p\}. \tag{2.15}$$

Proof. Let $Q_\theta(t)$ be as defined as:

$$Q_\theta(t) = \frac{d}{d\left\langle N^*\right\rangle_t} \int_0^t k_{\mathbb{1}}^t(s) \frac{A(s, X_s, \theta)}{\sigma(s)} ds \quad t \in [0, T].$$

Also, defining N_t^* and $\langle N^* \rangle_t$ as

$$N_t^* = \int_0^t k_{\mathbb{1}}^t(s) dV_s^\gamma, \qquad \langle N^* \rangle_t = \int_{[0,t)} k_{\mathbb{1}}^t(s) ds$$

by Theorem II.6, N_t^* is the fundamental martingale associated with V_t^γ. Defining U_t, as above, by

$$U_t = \int_0^t \frac{k_{\mathbb{1}}^t(s)}{\sigma(s)} dX_s$$

the process $U = \{U_t; 0 \le t \le T\}$ is an $(\mathscr{F}_t)$-semimartingale with the decomposition

$$U_t = \int_0^t Q_\theta(s) d \langle N^* \rangle_s + \int_0^t k_{\mathbb{1}}^t(s) dV_s^\gamma.$$

Let $\mathbf{P}_\theta^T$ be the measure induced by the process $\{X_t; 0 \le t \le T\}$ when θ is the true parameter. We then get that the Radon-Nikodym derivative of $\mathbf{P}_\theta^T$ with respect to $\mathbf{P}_0^T$ is given by:

$$\frac{d\mathbf{P}_\theta^T}{d\mathbf{P}_0^T} = \exp\left\{ \int_0^T Q_\theta(t) dU_t - \frac{1}{2} \int_0^T Q_\theta^2(t) d \langle N^* \rangle_t \right\}$$

Let $L_T(\theta)$ denote the Radon-Nikodym derivative $\frac{d\mathbf{P}_\theta^T}{d\mathbf{P}_0^T}$. Each element of the set of maximum likelihood estimators (MLE), $\hat{\theta}_T$, is defined by the relation:

$$L_T(\hat{\theta}_T) = \sup_{\theta \in \Theta} L_T(\theta).$$

Note that we have

$$\begin{aligned} Q_\theta(t) &= \frac{d}{d \langle N^* \rangle_t} \int_0^t k_{\mathbb{1}}^t(s) \frac{A(s, X_s, \theta)}{\sigma(s)} ds \\ &= \sum_{i=0}^N \left(\theta^i \frac{d}{d \langle N^* \rangle_t} \int_0^t k_{\mathbb{1}}^t(s) \frac{a_i(s, Y_s)}{\sigma(s)} ds \right) \\ &= \sum_{i=0}^N J_i(t) \theta^i. \end{aligned}$$

Then,

$$\ln\left(L_T\right)(\theta) = \int_0^T \left(\sum_{i=0}^N J_i(t)\theta^i \right) dU_t - \frac{1}{2} \int_0^T \left(\sum_{i=0}^N J_i(t)\theta^i \right)^2 d \langle N^* \rangle_t . \tag{2.16}$$

Thus the MLE satisfies

$$\hat{\theta}_T \in \underset{\theta\in\Theta}{\operatorname{argmax}} \left\{ \int_0^T \left(\sum_{i=0}^N J_i(t)\theta^i \right) dU_t - \frac{1}{2} \int_0^T \left(\sum_{i=0}^N J_i(t)\theta^i \right)^2 d\left\langle N^* \right\rangle_t \right\}.$$

□

2.4.2 Case of linear drift

A specific case of interest is when the drift function, $A(t, X_t, \theta)$, is linear in θ, i.e.

$$A(t, x, \theta) = a_0(t, x) + \theta a_1(t, x). \tag{2.17}$$

In this case, the MLE is unique and has the following analytic expression for the MLE directly from Eq. (2.16).

Corollary II.11. *Under the assumptions of Theorem II.10, and using the same notation, when the drift term $A(t, x, \theta)$ is given by Eq. (2.17) the MLE, $\hat{\theta}_T$, of θ is given by:*

$$\hat{\theta}_T = \frac{\int_0^T J_1(t)dU_t - \int_0^T J_0(t)J_1(t)d\left\langle N^* \right\rangle_t}{\int_0^T J_1^2(t)d\left\langle N^* \right\rangle_t}, \tag{2.18}$$

where $\left\langle N^* \right\rangle_t$, U_t and J_i are given by Eq. (2.13)-Eq. (2.15) respectively.

2.4.3 Properties of Maximum Likelihood Estimate

Here we prove several properties for the maximum likelihood estimator, $\hat{\theta}_T$ of θ from Eq. (2.11) in the linear case discussed in Corollary II.11. Specifically we prove that under certain assumptions our estimator is strongly consistent, a law of iterated logarithm holds, as well as a central limit theorem.

Theorem II.12. *The MLE, $\hat{\theta}_T$ given by Eq. (2.18), is strongly consistent provided*

$$\int_0^T J_1^2(t)d\left\langle N^* \right\rangle_t \to \infty \quad \text{a.s. } [\mathbf{P}_{\theta_0}] \text{ as } T \to \infty.$$

Proof. Let θ_0 be the true parameter. Then, since

$$dU_t = (J_0(t) + \theta_0 J_1(t))\, d\langle N^*\rangle_t + dN_t^*,$$

we have

$$\frac{d\mathbf{P}_\theta^T}{d\mathbf{P}_{\theta_0}^T} = \exp\left\{(\theta - \theta_0)\int_0^T J_1(t)dN_t^* - \frac{1}{2}(\theta-\theta_0)^2\int_0^T J_1^2(t)d\langle N^*\rangle_t\right\}.$$

Following this representation of the Radon-Nikodym derivative, we obtain that

$$\hat{\theta}_T - \theta_0 = \frac{\int_0^T J_1(t)dN_t^*}{\int_0^{T_1} J_1^2(t)d\langle N^*\rangle_t}.$$

Thus

$$G_T \equiv \int_0^T J_1(t)dN_t^*, \quad T \geq 0 \tag{2.19}$$

is a local martingale with the quadratic variation process

$$\langle G\rangle_T = \int_0^T J_1^2(t)d\langle N^*\rangle_t \to \infty \quad \text{a.s.} \tag{2.20}$$

and thus, by the strong law of large numbers for square-integrable martingales, (Corollary 1, p. 144 in [35]]),

$$\frac{G_T}{\langle G\rangle_T} \to 0 \quad \text{w.p. 1.}$$

Hence, $\hat{\theta}_T - \theta_0 = \frac{G_T}{\langle G\rangle_T} \to 0 \quad \text{w.p. 1.}$ □

For the next theorem, we recall a result on the law of iterated logarithm for local martingales.

Theorem II.13 ([33], Théorème 3, Translated from French)**.** *If M is a local martingale with locally integrable paths,*

$$\mathbb{E}\left[\sup_t |M_t - M_{t-}|\right] < \infty,$$

and

$$\langle M \rangle_t \to \infty \quad a.s.\ \mathbf{P} \text{ as t} \to \infty,$$

then

$$\limsup_{t\to\infty} \frac{M_t}{\sqrt{2\langle M \rangle_t \ln\left(\ln\left(\langle M \rangle_t\right)\right)}} = 1 \quad a.s.\ \mathbf{P}.$$

Now, we establish a law of iterated logarithm for the estimator Eq. (2.18).

Theorem II.14 (Law of iterated logarithm)**.** *Under the assumptions of Theorem II.12, we have that*

$$\limsup_{T\to\infty}(\hat{\theta}_T - \theta_0) \times \left(\frac{2\ln\left(\ln\left(\int_0^T J_1^2(t) d\langle N^* \rangle_t\right)\right)}{\int_0^T J_1^2(t) d\langle N^* \rangle_t} \right)^{-\frac{1}{2}} = 1 \quad \text{a.s. } [\mathbf{P}_{\theta_0}]$$

where $\langle N^* \rangle_t$ is given by Eq. (2.13), J_1 is given by Eq. (2.15), $\hat{\theta}_T$ is given by Eq. (2.18), and where $\theta_0 \in \Theta$ is the true parameter.

Proof. Using the notation from Theorem II.12, namely G_T the local martingale given by Eq. (2.19) with its corresponding quadratic variation process $\langle G \rangle_T$ given by Eq. (2.20), we have again that

$$\hat{\theta}_T - \theta_0 = \frac{G_T}{\langle G \rangle_T}. \tag{2.21}$$

From Theorem II.13, we have that

$$\limsup_{T\to\infty} \frac{G_T}{\sqrt{2\langle G \rangle_T \log\log \langle G \rangle_T}} = 1 \quad \text{a.s. } [\mathbf{P}_{\theta_0}]. \tag{2.22}$$

Since, by Eq. (2.21)

$$\begin{aligned}\frac{G_T}{\sqrt{2\langle G \rangle_T \log\log \langle G \rangle_T}} &= \frac{G_T\left(\hat{\theta}_T - \theta_0\right)}{\sqrt{2\langle G \rangle_T \log\log \langle G \rangle_T}} \times \frac{\langle G \rangle_T}{G_T} \\ &= \frac{\sqrt{\langle G \rangle_T}\left(\hat{\theta}_T - \theta_0\right)}{\sqrt{2\log\log \langle G \rangle_T}},\end{aligned}$$

then, Eq. (2.22) implies

$$\limsup_{T\to\infty} \frac{\sqrt{\langle G\rangle_T}\left(\hat{\theta}_T - \theta_0\right)}{\sqrt{2\log\log\langle G\rangle_T}} = 1 \quad \text{a.s. } [\mathbf{P}_{\theta_0}].$$

□

Theorem II.15. *Assume that functions $a_1(t,x)$ and $\sigma(t)$ are such that*

$$G_t = \int_0^t J_1(s)dN_s^*$$

is a local continuous martingale and that there exists a normalizing function I_t, $t \geq 0$ s.t.

$$I_T^2\langle G_T\rangle = I_T^2\int_0^T J_2^2(t)d\langle N^*\rangle_t \to \eta^2, \qquad \textit{in probability as } T\to\infty,$$

where $I_T \to 0$ as $T\to\infty$ and η is a random variable such that $\mathbf{P}(\eta > 0) = 1$. Then

$$(I_T G_T, I_T^2\langle G_T\rangle) \to (\eta Z, \eta^2) \textit{ in distribution as } T\to\infty,$$

where the random variable Z has the standard Normal distribution and $Z \perp\!\!\!\perp \eta$.

Proof. Follows from the central limit theorem for martingales (see, for example, [47]).

□

Theorem II.16. *Under the assumptions of Theorem II.15,*

$$I_T^{-1}(\hat{\theta}_T - \theta_0) \to \frac{Z}{\eta} \qquad \textit{in distribution as } T\to\infty$$

where Z is a standard Normal random variable and $Z \perp\!\!\!\perp \eta$

Proof. We note that

$$I_T^{-1}(\hat{\theta}_T - \theta_0) = \frac{I_T G_T}{I_T^2\langle R_T\rangle},$$

and the desired result follows immediately from Theorem II.15. □

2.4.4 Discretization of MLE

The above maximum likelihood estimator, $\hat{\theta}_T$ is based on continuously observing the process X_t. In any practical application this is an unreasonable requirement. Typically one could at best assume to observe the process at a mesh of time points. As such we would like to develop a consistent estimator based only on the discrete set of observations. Assuming that we have $n+1$ observations equally spread over the interval $[0,T]$ (including the observation $X_0 = 0$), and introducing the notation $t_k = \frac{kT}{n}$, for $k = 0, \dots, n$, we assume that $a_0(t, X_t) \equiv 0$, and have the following partially discretized estimator:

$$\bar{\theta}_n = \frac{\sum_{m=0}^{n-1} J_1(t_m)\left(U_{t_{m+1}} - U_{t_m}\right)}{\sum_{m=0}^{n-1} |J_1(t_m)|^2 \left(\langle N^*\rangle_{t_{m+1}} - \langle N^*\rangle_{t_m}\right)}. \tag{2.23}$$

Now, consider the two semi-martingales

$$A_n = \int_0^{t_n} J_1(s)dU_s = \int_0^T J_1(s)dU_s,$$

$$B_n = \sum_{m=0}^{n-1} J_1(t_m)\left(U_{t_{m+1}} - U_{t_m}\right).$$

Since

$$\langle B\rangle_n = \sum_{m=0}^{n-1} |J_1(t_m)|^2 \left(\langle N^*\rangle_{t_{m+1}} - \langle N^*\rangle_{t_m}\right),$$

we have

$$\hat{\theta}_n = \frac{A_n}{\langle A\rangle_n} \qquad \text{and} \qquad \bar{\theta}_n = \frac{B_n}{\langle B\rangle_n}.$$

The following proposition, a generalization of Proposition 5 from [57], gives conditions for the partially discretized estimator, $\bar{\theta}$ to converge to the maximum likelihood estimator, $\hat{\theta}$.

Proposition II.17. *If there exist constants α, $\gamma > 0$ such that*

(C1) $\dfrac{n^\alpha \langle A - B\rangle_n}{\langle B\rangle_n}$ *is almost surely bounded for n large*

(C2) $\frac{A_n - B_n}{\langle B \rangle_n}$ *converges to 0 almost surely as* $n \to \infty$.

then

$$\lim_{n\to\infty} |\hat{\theta}_n - \bar{\theta}_n| = 0 \quad a.s.$$

Proof. We have

$$\hat{\theta}_n - \bar{\theta}_n = \frac{A_n}{\langle A \rangle_n} - \frac{B_n}{\langle B \rangle_n} = \frac{A_n - B_n}{\langle B \rangle_n} + \frac{A_n}{\langle A \rangle_n}\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n}. \tag{2.24}$$

Using condition (C2) it is clear that we only have to show

$$\frac{A_n}{\langle A \rangle_n}\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \xrightarrow[n\to\infty]{a.s.} 0.$$

Now, consider

$$\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n}.$$

We have, almost surely for large n

$$\begin{aligned}
\left|\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n}\right| &= \frac{|\langle B - A, B + A \rangle_n|}{\langle B \rangle_n} \\
&\le \frac{|\langle B + A \rangle_n|^{\frac{1}{2}} |\langle B - A \rangle_n|^{\frac{1}{2}}}{\langle B \rangle_n} && \text{by Hölder's inequality} \\
&\le \frac{\left(|\langle A \rangle_n|^{\frac{1}{2}} + |\langle B \rangle_n|^{\frac{1}{2}}\right) |\langle B - A \rangle_n|^{\frac{1}{2}}}{\langle B \rangle_n} && \text{by Minkowski's inequality} \\
&\le \frac{\sqrt{2}\,(|\langle A \rangle_n| + |\langle B \rangle_n|)^{\frac{1}{2}} |\langle B - A \rangle_n|^{\frac{1}{2}}}{\langle B \rangle_n} \\
&\le \frac{\sqrt{2}\,(|\langle A \rangle_n + \langle B \rangle_n|)^{\frac{1}{2}} |K^{\frac{1}{2}} n^{-\frac{\alpha}{2}}| \langle B \rangle_n|^{\frac{1}{2}}}{\langle B \rangle_n} && \text{for some constant } K > 0 \text{ by (C1)} \\
&= \sqrt{2Kn^{-\alpha}}\; \frac{|\langle A \rangle_n + \langle B \rangle_n|^{\frac{1}{2}}}{|\langle B \rangle_n|^{\frac{1}{2}}}.
\end{aligned}$$

Let $Kn^{-\alpha} = \epsilon$. Then,

$$\begin{aligned}
& & \left| \frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \right| \leq \sqrt{2\epsilon} \left| \frac{\langle A \rangle_n}{\langle B \rangle_n} + 1 \right|^{\frac{1}{2}} \\
&\Leftrightarrow & \left(\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \right)^2 \leq 2\epsilon \left(\frac{\langle A \rangle_n}{\langle B \rangle_n} + 1 \right) \\
&\Leftrightarrow & \left(\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \right)^2 - 2\epsilon \frac{\langle A \rangle_n}{\langle B \rangle_n} + 2\epsilon + \epsilon^2 \leq 4\epsilon + \epsilon^2 \\
&\Leftrightarrow & \left(\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \right)^2 + 2 \left(\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \right) \epsilon + \epsilon^2 \leq 4\epsilon + \epsilon^2 \\
&\Leftrightarrow & \left(\frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} + \epsilon \right)^2 \leq 4\epsilon + \epsilon^2 \\
&\Leftrightarrow & \left| \frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} + \epsilon \right| \leq \sqrt{4\epsilon + \epsilon^2} \quad \forall \epsilon \\
&\Leftrightarrow & \frac{\langle B \rangle_n - \langle A \rangle_n}{\langle B \rangle_n} \xrightarrow[n\to\infty]{a.s} 0.
\end{aligned}$$

Since $\frac{A_n}{\langle A \rangle_n} = \hat{\theta}_n$ and $\hat{\theta}$ is strongly consistent and thus a.s. bounded for large n, we have our desired result. □

The next logical step, a fully discretized estimator of the form:

$$\widetilde{\theta}_n = \frac{\sum_{m=0}^{n-1} \widetilde{J}_1(t_m) \left(\widetilde{U}_{t_{m+1}} - \widetilde{U}_{t_m} \right)}{\sum_{m=0}^{n-1} \widetilde{J}_1(t_m) \left(\langle N^* \rangle_{t_{m+1}} - \langle N^* \rangle_{t_m} \right)}, \tag{2.25}$$

where

$$\begin{aligned}
\widetilde{J}_1(t_m) &= \frac{d}{d \langle N^* \rangle_{t_m}} \sum_{l=0}^{m-1} k_{\mathbb{1}}^z(t_l) a_1 \left(X_{t_l} \right) \left(t_{l+1} - t_l \right), \\
\widetilde{U}_{t_m} &= \sum_{l=0}^{m-1} k_{\mathbb{1}}^z(t_l) \left(X_{t_{l+1}} - X_{t_l} \right),
\end{aligned}$$

can not, as yet, be generally shown to converge to $\hat{\theta}$ since this convergence depends directly on the specific form of the kernels defining the Volterra process and its associated fundamental martingale. The estimator given in Eq. (2.25) has been previously shown to converge in the case of fractional Brownian motion by Tudor and Viens[57].

CHAPTER III

Parameter estimation in multi-dimensional Stochastic Differential Equations

3.1 Introduction

The development of estimation in stochastic differential equation with multi-parameter Gaussian noise has received considerably less attention than problems concerning 1-parameter noise. Part of the reason for this is the lack of full development of martingale theory in multiple parameters. Due to the loss of total ordering, even the definition of martingale requires extra care. However, utilizing strong martingales associated with the Volterra process, a maximum likelihood estimator can be defined and using sectorial limits, several asymptotic properties of this estimator can be established. Another difficulty that is overcome is proving the existence and uniqueness of a strong solution to the stochastic differential equation in the multi-parameter Volterra noise case.

In this chapter, after carefully defining strong martingales in the hyper-plane, existence of a solution to the stochastic differential equation is shown, a maximum likelihood estimator is defined and strong consistency of the estimator is established.

3.2 Preliminaries

First, we introduce the following notation (again where $\triangleq$ means 'is denoted as'):

- f'_t will represent $\frac{df}{dt}$, or if f is a function of several variables, i.e. $f(s,t,u)$, then $f'_{(1)} \triangleq f'_s$, $f'_{(2)} \triangleq f'_t$ etc.
- For $t^1, t^2 \in \mathbb{R}^d$, $t^i = (t^i_1, t^i_2, \ldots, t^i_d)$, $i = 1, 2$, $s = (s_1, \ldots, s_d) \in \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$
  - When it exists,

$$\int_{t^1}^{t^2} f(s)ds \triangleq \int_{t^1_1}^{t^2_1} \cdots \int_{t^1_d}^{t^2_d} f(s_1, \ldots, s_d)ds_d \ldots ds_1.$$

  - If $t^1_i < t^2_i$ for all $i = 1, \ldots, d$, then

$$[t^1, t^2] \triangleq [t^1_1, t^2_1] \times \cdots \times [t^1_d, t^2_d].$$

Due to the fact that we will be taking limits in a plane, we need to define exactly how this limit is to be interpreted. Because of the underdevelopment of general multi-parameter martingale and random processes limit theorems, we will use simpler sectorial limits.

Definition III.1 (Sectorial Limits)**.** For $d \in \mathbb{N}$, let $\prod_d$ be the collection of all permutations of $\{1 \ldots d\}$. For any $f : \mathbb{R}^d_+ \to \mathbb{R}$ and any $\pi \in \prod_d$, define, for any $s \in \mathbb{R}^d$ and fixed $t \in (\mathbb{R} \times \infty)^d$

$$\pi - \lim_{s \to t} f(s) \triangleq \lim_{s_{\pi(1)} \to t_{\pi(1)}} \cdots \lim_{s_{\pi(d)} \to t_{\pi(d)}} f(s),$$

if it exists. We say f has **sectorial limits** at t if $\pi - \lim_{s \to t} f(s)$ exists for all $\pi \in \prod_d$. If all the limits are the same, we denote the common limit (the **sectorial limit**) as

$$\lim_{s \rightsquigarrow t} f(s).$$

An additional difficulty in dealing with processes in a d-parameter space is that we need to consider how we can define increments of the process. To that end, we first define partial ordering.

Definition III.2 (Partial Ordering)**.** Let $a = (a_1, \ldots, a_d), b = (b_1, \ldots, b_d) \in \mathbb{R}^d_+$ be arbitrary. We say

i. $a \preccurlyeq b$ if and only if $a_i \leq b_i$ for all $i = 1, \ldots, d$.

ii. $a \prec b$ if and only if $a_i < b_i$ for all $i = 1, \ldots, d$.

iii. $a \curlywedge b \triangleq (\min(a_1, b_1), \ldots, \min(a_d, b_d))$.

iv. $a \curlyvee b \triangleq (\max(a_1, b_1), \ldots, \max(a_d, b_d))$.

We can now define a dth dimensional increment of a random process in the hyperplane.

Definition III.3 (Increment)**.** Let $X = \{X_t, t \in \mathbb{R}^d_+\}$ be a d-dimensional process, and let $t^1 = (t^1_1, t^1_2, \ldots, t^1_d),\ t^2 = (t^2_1, t^2_2, \ldots, t^2_d) \in \mathbb{R}^d_+$ be such that $t^1 \prec t^2$. Then, we define the increment $X((t^1, t^2])$ as:

$$X((t^1, t^2]) = \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{d-\sum_{j=1}^{d} i_j} X_{t_1^{i_1}, t_2^{i_2}, \ldots, t_d^{i_d}}.$$

In particular, when $d = 1$, $X((t^1, t^2]) = X_{t_1^2} - X_{t_1^1}$ and when $d = 2$, $X((t^1, t^2]) = X_{t_1^2, t_2^2} - X_{t_1^1, t_2^2} - X_{t_1^2, t_2^1} + X_{t_1^1, t_2^1}$.

For the complete probability space $(\Omega, \mathscr{F}, \mathbf{P})$ in $\mathbb{R}^d_+$, we will require our filtration to satisfy the following standard conditions:

Definition III.4. We say that filtration $\{\mathscr{F}_t, t \in \mathbb{R}^d_+\}$ satisfies the conditions (F1)-(F4) if:

(F1) For all $t^1, t^2 \in \mathbb{R}^d_+$ where $t^1 \preccurlyeq t^2$, $\mathscr{F}_t \subset \mathscr{F}_{t^2}$,

(F2) $\mathscr{F}_0$ contains all the $\mathbf{P}$-null sets of $\mathscr{F}$,

(F3) For all $t^1 \in \mathbb{R}^d_+$, $\mathscr{F}_{t^1} = \bigcap_{t^1 \prec t^2} \mathscr{F}_{t^2}$, (F4) For each $t^1 \in \mathbb{R}_+^d$, the collection $\{\mathscr{F}_{t^1}^i\}_{i=1}^d$ are conditionally independent given $\mathscr{F}_{t^1}$, where $\mathscr{F}_{t^1}^i$ is defined by:

$$\mathscr{F}_{t^1}^i = \bigvee_{t^2 \in \mathcal{J}_i} \mathscr{F}_{t^2} = \sigma \left\{ \bigcup_{t^2 \in \mathcal{J}_i} \mathscr{F}_{t^2} \right\} \qquad \text{where } \mathcal{J}_i = \mathbb{R}_+^{i-1} \times \{t_i^1\} \times \mathbb{R}_+^{d-i}.$$

Definition III.5 (Quadratic Variation)**.** If X_t is a continuous process, then X's quadratic variation, denoted $\langle X \rangle_t$ is defined by:

$$\langle X \rangle_t (\omega) = \lim_{\triangle t_k \to 0} \sum_{t_k \preccurlyeq t} \left(X_{t_{k+1}}(\omega) - X_{t_k}(\omega) \right),$$

where the $\{t_k\}$ form a partition over $[0, t]$.

d-parameter martingales are defined similarly to 1 parameter martingales.

Definition III.6 (d-parameter martingale)**.** Let $(\mathscr{F}_t)_{t \in \mathbb{R}_+^d}$ be a filtration satisfying (F1)-(F4). The process $X = \{X_t, t \in \mathbb{R}_+^d\}$ is called a d-parameter martingale with respect to $(\mathscr{F}_t)$ if:

i) For each $t \in \mathbb{R}_+^d$, X_t is adapted to $\mathscr{F}_t$ and integrable.

ii) For each $t^1 \preccurlyeq t^2$, $\mathbb{E}\left[X_{t^2} | \mathscr{F}_{t^1}\right] = X_{t^1}$ a.s..

To use a multi-parameter version of Theorem II.6, we will need to deal with more restrictive processes than martingales called strong martingales.

Definition III.7 (d-dimensional strong martingale)**.** Let $X = \{X_t, t \in \mathbb{R}_+^d\}$ be a process such that X_t is integrable for all $t \in \mathbb{R}_+^d$ and let filtration $(\mathscr{F}_t)_{t \in \mathbb{R}_+^d}$ satisfy (F1)-(F4). Then X is called a d-dimensional strong martingale with respect to $(\mathscr{F}_t)$ if:

i) X is adapted to $(\mathscr{F}_t)$,

ii) X vanishes on all axes (i.e. $X_t = 0$ a.s. $\forall\, t = (0, \ldots, 0, t_i, 0, \ldots, 0)$, $t_i \in \mathbb{R}_+^d$, $i = 1, \ldots, d$),

iii) $\mathbb{E}\left[X((t^1,t^2])\,\middle|\,\bigvee_{i=1}^{d}\mathscr{F}_{t^1}^{i}\right]=0$ a.s. $\forall\, t^1 \prec t^2$.

The following proposition gives a sufficient condition for a Gaussian process to have independent increments.

Proposition III.8. *If $X=\{X_t, t\in\mathbb{R}_+^d\}$ is a d-parameter Gaussian random field with covariance*

$$\mathbb{E}\left[X_{t^1}X_{t^2}\right]=\langle X\rangle_{t^1\wedge t^2} \qquad \forall\, t^1,t^2\in\mathbb{R}_+^d.$$

Then X has independent increments.

Proof. Let $t^1,t^2,t^3,t^4\in\mathbb{R}_+^d$ where $t^i=(t_1^i,t_2^i,\ldots,t_d^i)$ for $i=0,1,2,4$ be such that $t^1\prec t^2$, $t^3\prec t^4$ and there exists a $j\in\{1,2,\ldots,d\}$ such that $t_j^2<t_j^3$. Then we have

$$\begin{aligned}
&\mathbb{E}\left[X((t^1,t^2])X((t^3,t^4])\right]=\\
&\quad=\mathbb{E}\left[\left(\sum_{i_1=0}^{2}\cdots\sum_{i_d=0}^{2}(-1)^{d-\sum_{j=1}^{d}i_j}X_{t_1^{i_1},\ldots,t_d^{i_d}}\right)\left(\sum_{k_1=2}^{4}\cdots\sum_{k_d=2}^{4}(-1)^{d-\sum_{l=1}^{d}k_l}X_{t_1^{k_1},\ldots,t_d^{k_d}}\right)\right]\\
&\quad=\sum_{i_1=0}^{2}\cdots\sum_{i_d=0}^{2}\sum_{k_1=2}^{4}\cdots\sum_{k_d=2}^{4}(-1)^{2d-\left(\sum_{j=1}^{d}i_j+\sum_{l=1}^{d}k_l\right)}\mathbb{E}\left[X_{t_1^{i_1},\ldots,t_d^{i_d}}X_{t_1^{k_1},\ldots,t_d^{k_d}}\right]\\
&\quad=\sum_{i_1=0}^{2}\cdots\sum_{i_d=0}^{2}\sum_{k_1=2}^{4}\cdots\sum_{k_d=2}^{4}(-1)^{\sum_{j=1}^{d}i_j+\sum_{l=1}^{d}k_l}\,\langle X\rangle_{t_1^{i_1}\wedge t_2^{k_1},\ldots,t_j^{i_j},\ldots,t_d^{i_d}\wedge t_d^{k_d}}\\
&\quad=0.
\end{aligned}$$

Now, since X is Gaussian, X has independent increments. □

As an example, let $d=2$. Then $X((t^1,t^2])=X_{t_1^2,t_2^2}-X_{t_1^1,t_2^2}-X_{t_1^2,t_2^1}+X_{t_1^1,t_2^1}$. Let $t^1,t^2,t^3,t^4\in\mathbb{R}_+^3$ where $t^i=(t_1^i,t_2^i)$ for $i=0,1,2,4$ be such that $t^1\prec t^2$, $t^3\prec t^4$ and, with out loss of generality, $t_1^2 < t_1^3$. Then,

$$
\begin{aligned}
&\mathbb{E}\left[X((t^1,t^2])X((t^3,t^4])\right] = \\
&= \mathbb{E}\left[\left(X_{t_1^2,t_2^2} - X_{t_1^1,t_2^2} - X_{t_1^2,t_2^1} + X_{t_1^1,t_2^1}\right)\left(X_{t_1^4,t_2^4} - X_{t_1^3,t_2^4} - X_{t_1^4,t_2^3} + X_{t_1^3,t_2^3}\right)\right] \\
&= \mathbb{E}\left[X_{t_1^2,t_2^2}X_{t_1^3,t_2^4}\right] - \mathbb{E}\left[X_{t_1^2,t_2^2}X_{t_1^3,t_2^4}\right] - \mathbb{E}\left[X_{t_1^2,t_2^2}X_{t_1^4,t_2^3}\right] + \mathbb{E}\left[X_{t_1^2,t_2^2}X_{t_1^3,t_2^3}\right] \\
&\qquad - \mathbb{E}\left[X_{t_1^1,t_2^2}X_{t_1^3,t_2^4}\right] + \mathbb{E}\left[X_{t_1^1,t_2^2}X_{t_1^3,t_2^4}\right] + \mathbb{E}\left[X_{t_1^1,t_2^2}X_{t_1^4,t_2^3}\right] - \mathbb{E}\left[X_{t_1^1,t_2^2}X_{t_1^3,t_2^3}\right] \\
&\qquad - \mathbb{E}\left[X_{t_1^2,t_2^1}X_{t_1^3,t_2^4}\right] + \mathbb{E}\left[X_{t_1^2,t_2^1}X_{t_1^3,t_2^4}\right] + \mathbb{E}\left[X_{t_1^2,t_2^1}X_{t_1^4,t_2^3}\right] - \mathbb{E}\left[X_{t_1^2,t_2^1}X_{t_1^3,t_2^3}\right] \\
&\qquad + \mathbb{E}\left[X_{t_1^1,t_2^1}X_{t_1^3,t_2^4}\right] - \mathbb{E}\left[X_{t_1^1,t_2^1}X_{t_1^3,t_2^4}\right] - \mathbb{E}\left[X_{t_1^1,t_2^1}X_{t_1^4,t_2^3}\right] + \mathbb{E}\left[X_{t_1^1,t_2^1}X_{t_1^3,t_2^3}\right] \\
&= \langle X\rangle_{t_1^2,t_2^2\wedge t_2^4} - \langle X\rangle_{t_1^2,t_2^2\wedge t_2^4} - \langle X\rangle_{t_1^2,t_2^2\wedge t_2^3} + \langle X\rangle_{t_1^2,t_2^2\wedge t_2^3} \\
&\quad - \langle X\rangle_{t_1^1,t_2^2\wedge t_2^4} + \langle X\rangle_{t_1^1,t_2^2\wedge t_2^4} + \langle X\rangle_{t_1^1,t_2^2\wedge t_2^3} - \langle X\rangle_{t_1^1,t_2^2\wedge t_2^3} \\
&\quad - \langle X\rangle_{t_1^2,t_2^1\wedge t_2^4} + \langle X\rangle_{t_1^2,t_2^1\wedge t_2^4} + \langle X\rangle_{t_1^2,t_2^1\wedge t_2^3} - \langle X\rangle_{t_1^2,t_2^1\wedge t_2^3} \\
&\quad + \langle X\rangle_{t_1^1,t_2^1\wedge t_2^4} - \langle X\rangle_{t_1^1,t_2^1\wedge t_2^3} - \langle X\rangle_{t_1^1,t_2^1\wedge t_2^3} + \langle X\rangle_{t_1^1,t_2^1\wedge t_2^3} \\
&= 0.
\end{aligned}
$$

Using Proposition III.8, and following the exact same proof as for Theorem II.6, we have the multiple parameter version of the associated martingale representation theorem.

Theorem III.9. *Let $X = (X_t)_{t\in\mathbb{T}}$ be a continuous mean-zero d-parameter Gaussian random process with the covariance function R, i.e.*

$$R(t,t') = \mathbb{E}\left[X_t X_{t'}\right].$$

For arbitrary continuous curves $\mathcal{C} : \mathbb{T} \to \mathbb{R}$ and $\forall t \in \mathbb{T}$, suppose that there exists a family of kernels $k_{\mathcal{C}}^t : [0,t) \to \mathbb{R}$ such that

$$\iint\limits_{[0,t)\times[0,t')} k_{\mathcal{C}}^t(s)k_{\mathcal{C}}^{t'}(s')R(ds,ds') = \int_{[0,t\wedge t')} \mathcal{C}(s)k_{\mathcal{C}}^{t\wedge t'}(s)ds, \quad \forall t,t' \in \mathbb{T}, \tag{3.1}$$

where $a \curlywedge b := \min(a,b)$. *Define the process* $N^{\mathcal{C}} = (N_t^{\mathcal{C}})_{t\in\mathbb{T}}$ *by:*

$$N_t^{\mathcal{C}} = \int_{[0,t)} k_{\mathcal{C}}^t(s)dX_s. \tag{3.2}$$

Then $N^{\mathcal{C}} = (N_t^{\mathcal{C}})_{t\in\mathbb{T}}$ *is a* d*-parameter strong Gaussian martingale with variance*

$$\langle N^{\mathcal{C}}\rangle_t = \int_{[0,t)} \mathcal{C}(s)k_{\mathcal{C}}^t(s)ds. \tag{3.3}$$

Moreover, for all $\mathcal{C}, \widetilde{\mathcal{C}} \in C(\mathbb{T},\mathbb{R})$,

$$N_t^{\widetilde{\mathcal{C}}} = \int_{[0,t)} q^{\mathcal{C},\widetilde{\mathcal{C}}}(s)dN_s^{\mathcal{C}}, \tag{3.4}$$

where

$$q_t^{\mathcal{C},\widetilde{\mathcal{C}}} = \frac{d\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t}{d\left\langle N^{\mathcal{C}}\right\rangle_t}, \tag{3.5}$$

with

$$\left\langle N^{\mathcal{C}}, N^{\widetilde{\mathcal{C}}}\right\rangle_t = \iint_{[0,t)^2} k_{\mathcal{C}}^t(v)k_{\widetilde{\mathcal{C}}}^t(v')R(dv,dv'). \tag{3.6}$$

The definition of Volterra processes in d parameters will again rely on the growth condition of Hölder continuity. We recall the following definition of Hölder continuity in d dimensions.

Definition III.10. A function $\phi : [t^1,t^2] \subset \mathbb{R}^d \to \mathbb{R}$ where $t^1 \prec t^2$, is Hölder continuous of index $\alpha = (\alpha_1,\ldots,\alpha_d) \in \mathbb{R}_+^d$ (also called α-Hölder continuous) if

i) ϕ is continuous.

ii) For all $1 \le i \le d$, $\left\|\phi(t_1^1,\ldots,t_{i-1}^1,\cdot,t_{i+1}^1,\ldots,t_d^1\right\|_{[t_i^1,t_i^2],\alpha_i} < \infty$ where

$$\left\|\phi(t_1^1,\ldots,t_{i-1}^1,\cdot,t_{i+1}^1,\ldots,t_d^1)\right\|_{[t_i^1,t_i^2],\alpha_i} = \sup_{u\neq v,\, t_i^1\le u,v\le t_i^2} \frac{|\phi(t_1^1,\ldots,t_{i-1}^1,u,t_{i+1}^1,\ldots,t_d^1) - \phi(t_1^1,\ldots,t_{i-1}^1,v,t_{i+1}^1,\ldots,t_d^1)|}{|u-v|^{\alpha_i}}.$$

iii) $\|\phi\|_{[t^1,t^2],\alpha} < \infty$, where

$$\|\phi\|_{[t^1,t^2],\alpha} = \sup_{u \prec v} \frac{|\phi((u,v])|}{\prod_{i=1}^d |u_i - v_i|^{\alpha_i}} \quad \text{where } u = (u_1, \ldots, u_d) \text{ and } v = (v_1, \ldots, v_d).$$

We let $\mathscr{C}_\alpha([t^1,t^2])$ denote the space of all α-Hölder continuous functions on the interval $[t^1,t^2]$. Next, we denote the space $\mathscr{C}_\alpha([t^1,t^2])$ equipped with the norm

$$\|x\|_{[t^1,t^2],\alpha,\infty} = \|x\|_\infty + \sup_{t_1^1 \le u_1 \le t_1^2} \|x(u_1,\cdot)\|_{[t_2^1,t_2^2],\alpha_2} + \sup_{t_1^2 \le u_2 \le t_2^2} \|x(\cdot,u_2)\|_{[t_1^1,t_1^2],\alpha_1} + \|x\|_{[t^1,t^2],\alpha}$$

by $\mathscr{C}_{\alpha,\infty}([t^1,t^2])$.

For a fixed $C > 0$, we let

$$\mathscr{C}_{\alpha,C}([t^1,t^2]) = \{\phi \in \mathscr{C}_\alpha([t^1,t^2]) \, : \, \|\phi\|_{[t^1,t^2],\alpha} \le C\},$$

and for a fixed $a \in \mathbb{R}$, we let

$$\mathscr{C}_{\alpha,C}([t^1,t^2],a) = \{\phi \in \mathscr{C}_{\alpha,C}([t^1,t^2]) \, : \, \phi(t) = a\}.$$

Finally, for $\phi_i \in \mathscr{C}_{\alpha_i}([t_i^1,t_i^2])$, $i = 1,2$, we let

$$\mathscr{C}_{\alpha,\infty,C,\phi_1,\phi_2}([t^1,t^2]) = \Big\{ x \in \mathscr{C}_{\alpha,\infty}([t^1,t^2]) : x(t_1^1,\cdot) = \phi_1, x(\cdot,t_2^1) = \phi_2, \\ \|x\|_{[t^1,t^2],\alpha} \le C, \\ \sup_{t_1^1 \le u_1 \le t_1^2} \|x(u_1,\cdot)\|_{[t_2^1,t_2^2],\alpha_2} \le C, \\ \sup_{t_2^1 \le u_2 \le t_2^2} \|x(\cdot,u_2)\|_{[t_1^1,t_1^2],\alpha_1} \le C \Big\}.$$

For the multi-parameter version of Volterra processes, let $\gamma = (\gamma_1, \ldots, \gamma_d) \in \mathbb{R}_+^d$. We define the multi-parameter kernel, $K^\gamma : [0,T]^d \times [0,T]^d \to [0,\infty)$ as follows:

$$K^\gamma(t^1,t^2) = \prod_{i=1}^d K^{\gamma_i}(t_i^1,t_i^2),$$

where $t^1, t^2 \in [0,T]^d$, and for each $i \in \{1,\dots,d\}$, K^{γ_i} satisfies conditions (C1)-(C3) (see p. 11).

Finally, we define a d parameter Volterra sheet.

Definition III.11 (d-parameter Volterra sheet)**.** A centered, d-parameter Gaussian random field, $V^\gamma = \{V_t^\gamma\}_{t\in[0,T]^d}$ with integral representation

$$V_t^\gamma = \int_{[0,T]} K^\gamma(t,s)d\mathbb{B}_s, \tag{3.7}$$

where $\mathbb{B} = \{\mathbb{B}_t\}_{t\in[0,T]}$ is a standard d-dimensional Brownian sheet is called a d-parameter Volterra process.

We note that V^γ has a.s. γ-Hölder continuous paths. Using the Kolmogorov-Chentsov Theorem, we need only show that, for all $t^1 \prec t^2$ where $t^1, t^2 \in [0,T]^d$,

$$\mathbb{E}\left[\left(V^\gamma\left((t^1,t^2]\right)\right)^p\right] \le C[(t_1^2 - t_1^1)\cdots(t_d^2 - t_d^1)]^{1+\alpha}$$

for some constants $C, \alpha > 0$ and $p \ge 2$.

We have, with $s = (s_1, \dots, s_d)$,

$$\begin{aligned}
\mathbb{E}\left[\left(V^\gamma\left((t^1,t^2]\right)\right)^2\right] &= \mathbb{E}\left[\left(\int_{[0,T]^d} \prod_{i=1}^d \left(K^{\gamma_i}(t_i^2, s_i) - K^{\gamma_i}(t_i^1, s_i)\right) d\mathbb{B}_s\right)^2\right] \\
&= \int_{[0,T]^d} \prod_{i=1}^d \left(K^{\gamma_i}(t_i^2, s_i) - K^{\gamma_i}(t_i^1, s_i)\right)^2 ds \\
&= \prod_{i=1}^d \left(\int_{[0,T]} \left(K^{\gamma_i}(t_i^2, s_i) - K^{\gamma_i}(t_i^1, s_i)\right)^2 ds_i\right) \\
&\le \prod_{i=1}^d C_i (t_i^2 - t_i^1)^{2\gamma_i}.
\end{aligned}$$

Since $V^\gamma\left((t^1,t^2]\right)$ has a mean-zero Normal distribution, we have the following relation of moments:

$$\mathbb{E}\left[\left(V^\gamma\left((t^1,t^2]\right)\right)^p\right] = C_p\, \mathbb{E}\left[\left(V^\gamma\left((t^1,t^2]\right)\right)^2\right]^{\frac{p}{2}}.$$

Thus, the desired result holds. We note that this also shows that in the 1-parameter case, V^γ has a.s. γ-Hölder continuous paths.

3.3 Existence and uniqueness of solutions

In this section we prove the existence and uniqueness of a strong solution to the stochastic differential equation whose drift parameter we will be estimating. From this point on, for simplicity, we will only be concerned with the two parameter case ($d = 2$). Extending all of the following results to any dimension $d \in \mathbb{N}$ is direct.

As in the 1 parameter case, we first need a result on the non-random differential equation where we have Hölder continuity growth conditions.

Proposition III.12. *Let* $\beta_1, \beta_2 \in \left(\frac{1}{2}, 1\right]$ *and* α_1, α_2 *be such that* $\beta_i > \alpha_i > 1-\beta_i$*. Let* $g \in \mathscr{C}_{\mathbb{R}^2,\beta}$ *and* $b, \sigma : \mathbb{R}^3 \to \mathbb{R}$ *such that* b *is bounded and Lipshitz in each dimension, and* σ *is bounded and has bounded third derivatives. Then for every* $K > 0$ *and* $t^1, t^2 \in \mathbb{R}^2$, $t^1 \prec t^2$*, there exists an* $\epsilon_0 > 0$*, independent of* t^1, t^2*, such that for every* $\phi_i \in \mathscr{C}_{\alpha_i,K}([t_i^1, t_i^2])$*,* $i = 1, 2$*, the operator*

$$F : \mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([t^1, t^1 + \epsilon_0]) \to \mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([t^1, t^1 + \epsilon_0]),$$

defined by

$$(Fx)_{u,v} = \phi_1(u)+\phi_2(v)+\int_{t_1^1}^{u}\int_{t_2^1}^{v} b(s_1, s_2, x_{s_1,s_2})ds_1 ds_2+\int_{t_1^1}^{u}\int_{t_2^1}^{v} \sigma(s_1, s_2, x_{s_1,s_2})dg(s_1, s_2),$$

is a contraction.

Proof. First,

$$\left\| \int_{t_1^1}^{\cdot} \int_{t_2^1}^{\cdot} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right\|_{[t^1,t^2],\alpha,\infty} =$$

$$= \left\| \int_{t_1^1}^{\cdot} \int_{t_2^1}^{\cdot} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right\|_{\infty} + \sup_{t_1^1 \le u_1 \le t_1^2} \left\| \int_{t_1^1}^{\cdot} \int_{t_2^1}^{\cdot} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right\|_{[t_2^1,t_2^2],\alpha_2}$$

$$+ \sup_{t_2^1 \le u_2 \le t_2^2} \left\| \int_{t_1^1}^{\cdot} \int_{t_2^1}^{\cdot} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right\|_{[t_1^1,t_1^2],\alpha_1}$$

$$+ \sup_{u \prec v} \frac{\left| \int_{u_1}^{v_1} \int_{u_2}^{v_2} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right|}{|v_1 - u_1|^{\alpha_1} |v_2 - u_2|^{\alpha_2}}$$

$$\le \|b\|_{\infty} (t_2^2 - t_2^1)(t_1^2 - t_1^1) + \sup_{t_1^1 \le u_1 \le t_1^2} \sup_{t_2^1 \le u_2, v_2 \le t_2^2} \frac{\left| \int_{t_1^1}^{u_1} \int_{u_2}^{v_2} b(s_1, s_2, x_{s_1,s_2}) ds_2 ds_1 \right|}{|v_2 - u_2|^{\alpha_2}}$$

$$+ \sup_{t_2^1 \le u_2 \le t_2^2} \sup_{t_1^1 \le u_1, v_1 \le t_1^2} \frac{\left| \int_{t_2^1}^{u_2} \int_{u_1}^{v_1} b(s_1, s_2, x_{s_1,s_2}) ds_1 ds_2 \right|}{|v_1 - u_1|^{\alpha_1}}$$

$$+ \|b\|_{\infty} \sup_{u \prec v} \frac{|v_1 - u_1||v_2 - u_2|}{|v_1 - u_1|^{\alpha_1} |v_2 - u_2|^{\alpha_2}}$$

$$\le \|b\|_{\infty} (t_1^2 - t_1^1)(t_2^2 - t_2^1) + \|b\|_{\infty} \sup_{t_1^1 \le u_1 \le t_1^2} \sup_{t_2^1 \le u_2, v_2 \le t_2^2} \frac{(u_1 - t_1^1)|v_2 - u_2|}{|v_2 - u_2|^{\alpha_2}}$$

$$+ \|b\|_{\infty} \sup_{t_2^1 \le u_2 \le t_2^2} \sup_{t_1^1 \le u_1, v_1 \le t_1^2} \frac{|v_1 - u_1|(u_2 - t_2^1)}{|v_1 - u_1|^{\alpha_1}}$$

$$+ \|b\|_{\infty} (t_2^2 - t_2^1)(t_1^2 - t_1^1)$$

$$(3.8) \le \|b\|_{\infty} (t_1^2 - t_1^1)^{1-\alpha_1} (t_2^2 - t_2^1)^{1-\alpha_2} \left((t_1^2 - t_1^1)^{\alpha_1} (t_2^2 - t_2^1)^{\alpha_2} + (t_1^2 - t_1^1)^{\alpha_1} + (t_2^2 - t_2^1)^{\alpha_2} + 1 \right).$$

Using Eq. (3.18) p. 773 from Reference [56], we have

(3.9)

$$\left\| \int_{t_1^1}^{\cdot} \int_{t_2^1}^{\cdot} \sigma(s_1, s_2, x_{s_1,s_2}) dg(s_1, s_2) \right\|_{[t^1,t^2],\alpha,\infty}$$

$$\le \|\sigma(\cdot, \cdot, x)\|_{[t^1,t^2],\alpha,\infty} \|g\|_{[t^1,t^2],\beta} (t_1^2 - t_1^1)^{\beta_1 - \alpha_1} (t_2^2 - t_2^1)^{\beta_2 - \alpha_2}$$

$$\times \left((t_1^2 - t_1^1)^{\alpha_1} (t_2^2 - t_2^1)^{\alpha_2} + (t_1^2 - t_1^1)^{\alpha_1} + (t_2^2 - t_2^1)^{\alpha_2} + 1 \right).$$

Now,

$$\begin{aligned}\|\sigma(\cdot,\cdot,x)\|_{[t^1,t^2],\alpha,\infty} =& \|\sigma\|_\infty + \sup_{t_1^1\le u_1\le t_1^2} \|\sigma(u_1,\cdot,x_{u_1,\cdot})\|_{[t_2^1,t_2^2],\alpha_2} \\ &+ \sup_{t_2^1\le u_2\le t_2^2} \|\sigma(\cdot,u_2,x_{\cdot,u_2})\|_{[t_1^1,t_1^2],\alpha_1} + \sup_{u\prec v} \frac{|\sigma(\cdot,\cdot,x)((u,v])|}{|v_1-u_1|^{\alpha_1}|v_2-u_2|^{\alpha_2}} \\ =& \|\sigma\|_\infty + \sup_{t_1^1\le u_1\le t_1^2} \sup_{t_2^1\le u_2,v_2\le t_2^2} \frac{|\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(u_1,v_2,x_{u_1,v_2})|}{|u_2-v_2|^{\alpha_2}} \\ &+ \sup_{t_2^1\le u_2\le t_2^2} \sup_{t_1^1\le u_1,v_1\le t_1^2} \frac{|\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(v_1,u_2,x_{v_1,u_2})|}{|u_1-v_1|^{\alpha_1}} \\ &+ \sup_{u\prec v} \frac{|\sigma(\cdot,\cdot,x)((u,v])|}{|v_1-u_1|^{\alpha_1}|v_2-u_2|^{\alpha_2}}.\end{aligned}$$

we have

$$\begin{aligned}&\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(u_1,v_2,x_{u_1,v_2}) = \\ &\quad = (u_2-v_2)\int_0^1 \sigma'_{(2)}\left(u_1,\lambda u_2+(1-\lambda)v_2,\lambda x_{u_1,u_2}+(1-\lambda)x_{u_1,v_2}\right)d\lambda \\ &\qquad + (x_{u_1,u_2}-x_{u_1,v_2})\int_0^1 \sigma'_{(3)}\left(u_1,\lambda u_2+(1-\lambda)v_2,\lambda x_{u_1,u_2}+(1-\lambda)x_{u_1,v_1}\right)d\lambda,\end{aligned}$$

$$\begin{aligned}|\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(u_1,v_2,x_{u_1,v_2})| \le& |u_2-v_2| \left\|\sigma'_{(2)}\right\|_\infty + |x_{u_1,u_2}-x_{u_1,v_2}| \left\|\sigma'_{(3)}\right\|_\infty \\ \le& |u_2-v_2| \left\|\sigma'_{(2)}\right\|_\infty \\ &+ |u_2-v_2|^{\alpha_2} \|x_{u_1,\cdot}\|_{[t_2^1,t_2^2],\alpha_2} \left\|\sigma'_{(3)}\right\|_\infty,\end{aligned}$$

so

$$\frac{|\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(u_1,v_2,x_{u_1,v_2})|}{|u_2-v_2|^{\alpha_2}} \le |u_2-v_2|^{1-\alpha_2} \left\|\sigma'_{(2)}\right\|_\infty + \|x_{u_1,\cdot}\|_{[t_2^1,t_2^2],\alpha_2} \left\|\sigma'_{(3)}\right\|_\infty,$$

and

$$\begin{aligned}&\sup_{t_1^1\le u_1\le t_1^2} \sup_{t_2^1\le u_2,v_2\le t_2^2} \frac{|\sigma(u_1,u_2,x_{u_1,u_2})-\sigma(u_1,v_2,x_{u_1,v_2})|}{|u_2-v_2|^{\alpha_2}} \le \\ &\qquad \le (t_2^2-t_2^1)^{1-\alpha_2} \left\|\sigma'_{(2)}\right\|_\infty + \sup_{t_1^1\le u_1\le t_1^2} \|x_{u_1,\cdot}\|_{[t_2^1,t_2^2],\alpha_2} \left\|\sigma'_{(3)}\right\|_\infty,\end{aligned}$$

similarly

$$\sup_{t_2^1 \le u_2 \le t_2^2} \sup_{t_1^1 \le u_1, v_1 \le t_1^2} \frac{|\sigma(u_1, u_2, x_{u_1,u_2}) - \sigma(v_1, u_2, x_{v_1,u_2})|}{|u_1 - v_1|^{\alpha_1}} \le$$

$$(t_1^2 - t_1^1)^{1-\alpha_1} \left\|\sigma'_{(2)}\right\|_\infty + \sup_{t_2^1 \le u_2 \le t_2^2} \|x_{\cdot,u_2}\|_{[t_1^1, t_1^2], \alpha_1} \left\|\sigma'_{(3)}\right\|_\infty .$$

Next

$$\begin{aligned}
&\sigma(\cdot,\cdot,x)((u,v]) = \\
&\quad = \underbrace{(v_2 - u_2) \int_0^1 \sigma'_{(2)}(v_1, \lambda v_2 + (1-\lambda)u_2, \lambda x_{v_1,v_2} + (1-\lambda)x_{v_1,u_2}) d\lambda}_{=a} \\
&\qquad + \underbrace{(x_{v_1,v_2} - x_{v_1,u_2}) \int_0^1 \sigma'_{(3)}(v_1, \lambda v_2 + (1-\lambda)u_2, \lambda x_{v_1,v_2} + (1-\lambda)x_{v_1,u_2}) d\lambda}_{=b} \\
&\qquad - \underbrace{(v_2 - u_2) \int_0^1 \sigma'_{(2)}(u_1, \lambda v_2 + (1-\lambda)u_2, \lambda x_{u_1,v_2} + (1-\lambda)x_{u_1,u_2}) d\lambda}_{=c} \\
&\qquad - \underbrace{(x_{u_1,v_2} - x_{u_1,u_2}) \int_0^1 \sigma'_{(3)}(u_1, \lambda v_2 + (1-\lambda)u_2, \lambda x_{u_1,v_2} + (1-\lambda)x_{u_1,u_2}) d\lambda}_{=d},
\end{aligned}$$

so

$$|\sigma(\cdot,\cdot,x)((u,v])| \le |a - c| + |b - d|,$$

and

$$
\begin{aligned}
|a-c| \leq & |(v_2-u_2)(v_1-u_1)\left\|\sigma''_{(2,1)}\right\|_\infty + \\
& + |(v_2-u_2)\left\|\sigma''_{(2,3)}\right\|_\infty \int_0^1 \lambda(x_{v_1,v_2}-x_{u_1,v_2})+(1-\lambda)(x_{v_1,u_2}x_{u_1,u_2})d\lambda \\
\leq & |v_2-u_2||v_1-u_1|^{\alpha_1}\times \\
& \times\left(\left\|\sigma''_{(2,1)}\right\|_\infty |v_1-u_1|^{1-\alpha_1} + \left\|\sigma''_{(2,3)}\right\|_\infty \sup_{t_1^1\leq u_1\leq t_1^2}\|x_{\cdot,u_1}\|_{[t_1^1,t_1^2],\alpha_1}\right) \\
|b-d| \leq & \Big|(x_{v_1,v_2}-x_{v_1,u_2}-x_{u_1,v_2}+x_{u_1,u_2})\times \\
& \times\int_0^1 \sigma'_{(3)}(v_1,\lambda v_2+(1-\lambda)u_2,\lambda x_{v_1,v_2}+(1-\lambda)x_{v_1,u_2})d\lambda \\
& + (x_{u_1,v_2}-x_{u_1,u_2})\int_0^1 \sigma'_{(3)}(v_1,\lambda v_2+(1-\lambda)u_2,\lambda x_{v_1,v_2}+(1-\lambda)x_{v_1,u_2}) \\
& - \sigma'_{(3)}(u_1,\lambda v_2+(1-\lambda)u_2,\lambda x_{u_1,v_2}+(1-\lambda)x_{u_1,u_2})d\lambda\Big|
\end{aligned}
$$

$$
\begin{aligned}
\leq & |v_2-u_2|^{\alpha_2}|v_1-u_1|^{\alpha_1}\|x\|_{[z^1,z^2],\alpha}\left\|\sigma'_{(3)}\right\|_\infty + |v_2-u_2|^{\alpha_2}\|x_{u_1,\cdot}\|_{[z_2^1,z_2^2],\alpha_2}\times \\
& \times\Big((v_1-u_1)\int_0^1\int_0^1 \sigma''_{(3,1)}(\mu v_1+(1-\mu)u_1, \\
& \qquad \lambda v_2+(1-\lambda)u_2, \\
& \qquad \mu(\lambda x_{v_1,v_2}+(1-\lambda)x_{v_1,u_2}) \\
& \qquad\qquad + (1-\mu)(\lambda x_{u_1,v_2}+(1-\lambda)x_{u_1,u_2}))d\mu d\lambda \\
& + \int_0^1 ((\lambda x_{v_1,v_2}+(1-\lambda)x_{v_1,u_2})-(\lambda x_{u_1,v_2}+(1-\lambda)x_{u_1,u_2}))\times \\
& \times\int_0^1 \sigma''_{(3,3)}(\mu v_1+(1-\mu)u_1, \\
& \qquad \lambda v_2+(1-\lambda)u_2, \\
& \qquad \mu(\lambda x_{v_1,v_2}+(1-\lambda)x_{v_1,u_2}) \\
& \qquad\qquad + (1-\mu)(\lambda x_{u_1,v_2}+(1-\lambda)x_{u_1,u_2}))d\mu\Big)
\end{aligned}
$$

$$\leq |v_2-u_2|^{\alpha_2}|v_1-u_1|^{\alpha_1}\left\|x\right\|_{[t^1,t^2],\alpha}\left\|\sigma'_{(3)}\right\|_\infty$$

$$+|v_2-u_2|^{\alpha_2}|v_1-u_1|\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\left\|\sigma''_{(3,1)}\right\|_\infty$$

$$+|v_2-u_2|^{\alpha_2}\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\left\|\sigma''_{(3,3)}\right\|_\infty\times$$

$$\times\int_0^1\lambda(x_{v_1,v_2}-x_{u_1,v_2})+(1-\lambda)(x_{v_1,u_2}-x_{u_1,u_2})d\lambda$$

$$\leq |v_2-u_2|^{\alpha_2}|v_1-u_1|^{\alpha_1}\Bigg(\left\|x\right\|_{[t^1,t^2],\alpha}\left\|\sigma'_{(3)}\right\|_\infty$$

$$+|v_1-u_1|^{1-\alpha_1}\sup_{t_1^1\leq u_1\leq t_1^2}\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\left\|\sigma''_{(3,1)}\right\|_\infty$$

$$+\sup_{t_1^1\leq u_1\leq t_1^2}\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\sup_{t_2^1\leq u_2\leq t_2^2}\left\|x_{\cdot,u_2}\right\|_{[t_1^1,t_1^2],\alpha_1}\left\|\sigma''_{(3,3)}\right\|_\infty\Bigg).$$

So,

$$\sup_{u\prec v}\frac{|\sigma(\cdot,\cdot,x)((u,v])|}{|v_1-u_1|^{\alpha_1}|v_2-u_2|^{\alpha_2}}\leq (t_1^2-t_1^1)^{1-\alpha_1}(t_2^2-t_2^1)^{1-\alpha_2}\left\|\sigma''_{(2,1)}\right\|_\infty$$

$$+(t_2^2-t_2^1)^{1-\alpha_2}\left\|\sigma''_{(2,3)}\right\|_\infty\sup_{t_1^1\leq u_1\leq t_1^2}\left\|x_{\cdot,u_1}\right\|_{[t_1^1,t_1^2],\alpha_1}$$

$$+\left\|x\right\|_{[t^1,t^2],\alpha}\left\|\sigma'_{(3)}\right\|_\infty \tag{3.10}$$

$$+|v_1-u_1|^{1-\alpha_1}\sup_{t_1^1\leq u_1\leq t_1^2}\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\left\|\sigma''_{(3,1)}\right\|_\infty$$

$$+\sup_{t_1^1\leq u_1\leq t_1^2}\left\|x_{u_1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2}\sup_{t_2^1\leq u_2\leq t_2^2}\left\|x_{\cdot,u_2}\right\|_{[t_1^1,t_1^2],\alpha_1}\left\|\sigma''_{(3,3)}\right\|_\infty.$$

Thus, from Eq. (3.8), Eq. (3.9) and Eq. (3.10), $Fx\in\mathscr{C}_{\alpha,\infty}([t^1,t^2])$ if $x\in\mathscr{C}_{\alpha,\infty}([t^1,t^2])$ and there exists an $\epsilon_1>0$ small enough such that $Fx\in\mathscr{C}_{\alpha,\infty,2K,\phi_1,\phi_2}([t^1,t^1+\epsilon_1])$ if $x\in\mathscr{C}_{\alpha,\infty,2K,\phi_1,\phi_2}([t^1,t^1+\epsilon_1])$.

Now, to bound $\left\|\sigma(\cdot,\cdot,x)-\sigma(\cdot,\cdot,x)\right\|_{[t^1,t^2],\alpha}$ for $x,y\in\mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([z^1,z^1+\epsilon_1])$, we consider

$$\underbrace{\sigma\left(t_1^2, t_2^2, x_{t_1^2,t_2^2}\right) - \sigma\left(t_1^2, t_2^2, y_{t_1^2,t_2^2}\right)}_{=A} - \underbrace{\left(\sigma\left(t_1^1, t_2^1, x_{t_1^2,t_2^1}\right) - \sigma\left(t_1^1, t_2^1, y_{t_1^2,t_2^1}\right)\right)}_{=B}$$
$$- \underbrace{\left(\sigma\left(t_1^1, t_2^2, x_{t_1^1,t_2^2}\right) - \sigma\left(t_1^1, t_2^2, y_{t_1^1,t_2^2}\right)\right)}_{=C} + \underbrace{\left(\sigma\left(t_1^1, t_2^1, x_{t_1^1,t_2^1}\right) - \sigma\left(t_1^1, t_2^1, y_{t_1^1,t_2^1}\right)\right)}_{=D}$$

. Noting that

$$A = \left(x_{t_1^2,t_2^2} - y_{t_1^2,t_2^2}\right) \int_0^1 \sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) d\lambda,$$
$$B = \left(x_{t_1^2,t_2^1} - y_{t_1^2,t_2^1}\right) \int_0^1 \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right) d\lambda,$$
$$C = \left(x_{t_1^1,t_2^2} - y_{t_1^1,t_2^2}\right) \int_0^1 \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) d\lambda,$$
$$D = \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \sigma'_{(3)}\left(t_1^1, t_2^1, \lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right) d\lambda,$$

we have,

$$
\begin{aligned}
&A - B - C + D = \\
&\quad = \underbrace{\begin{aligned}\Big(x_{t_1^2,t_2^2} - y_{t_1^2,t_2^2} - \left(x_{t_1^2,t_2^1} - y_{t_1^2,t_2^1}\right) - \left(x_{t_1^1,t_2^2} - y_{t_1^1,t_2^2}\right) \\
+ \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right)\Big) \int_0^1 \sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) d\lambda\end{aligned}}_{=A'} \\
&\quad \underbrace{\begin{aligned}&+ \left(x_{t_1^2,t_2^1} - y_{t_1^2,t_2^1}\right) \int_0^1 \Big(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
&\qquad - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Big) d\lambda \\
&+ \left(x_{t_1^1,t_2^2} - y_{t_1^1,t_2^2}\right) \int_0^1 \Big(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
&\qquad - \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right)\Big) d\lambda \\
&- \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \Big(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
&\qquad - \sigma'_{(3)}\left(t_1^1, t_2^1, \lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Big) d\lambda\,.\end{aligned}}_{=D'}
\end{aligned}
$$

We also see that

$$
\begin{aligned}
D' = \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \Bigg(& \sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
& - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right) \\
& - \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \\
& + \sigma'_{(3)}\left(t_1^1, t_2^1, \lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right) \Bigg) d\lambda \\
- \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \Bigg(& \sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
& - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right) \Bigg) d\lambda \\
- \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \Bigg(& \sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
& - \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \Bigg) d\lambda.
\end{aligned}
$$

We can rewrite $A - B - C + D$ as follows

$$
\begin{aligned}
A - B - C + D = A' & \\
& \underbrace{\begin{aligned} &+ \left(\left(x_{t_1^2,t_2^1} - y_{t_1^2,t_2^1}\right) - \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right)\right) \times \\ &\quad \times \int_0^1 \left(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right)\right. \\ &\qquad\qquad \left. - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\right) d\lambda \end{aligned}}_{=B'} \\
& \underbrace{\begin{aligned} &+ \left(\left(x_{t_1^1,t_2^2} - y_{t_1^1,t_2^2}\right) - \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right)\right) \times \\ &\quad \int_0^1 \left(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right)\right. \\ &\qquad\qquad \left. - \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right)\right) d\lambda \end{aligned}}_{=C'} \\
& + \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \int_0^1 \left(\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right)\right. \\
& \qquad - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right) \\
& \qquad - \sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \\
& \qquad \left. + \sigma'_{(3)}\left(t_1^1, t_2^1, \lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\right) d\lambda.
\end{aligned}
$$

Since

$$\sigma'_{(3)}\left(t_1^2, t_2^2, \lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) - \sigma'_{(3)}\left(t_1^2, t_2^1, \lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)$$
$$= (t_2^2 - t_2^1)\int_0^1 \sigma''_{(3,2)}\Bigg(t_1^2, \mu t_2^2 + (1-\mu)t_2^1, \mu\left(\lambda x_{t_1^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) d\mu$$
$$+ \left(\lambda\left(x_{t_1^2,t_2^2} - x_{t_1^2,t_2^1}\right) + (1-\lambda)\left(y_{t_1^2,t_2^2} - y_{t_1^2,t_2^1}\right)\right) \times$$
$$\times \int_0^1 \sigma''_{(3,3)}\Bigg(t_1^2, \mu t_2^2 + (1-\mu)t_2^1, \mu\left(\lambda x_{t_1^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) d\mu,$$

and

$$\sigma'_{(3)}\left(t_1^1, t_2^2, \lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) - \sigma'_{(3)}\left(t_1^1, t_2^1, \lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)$$
$$= (t_2^2 - t_2^1)\int_0^1 \sigma''_{(3,2)}\Bigg(t_1^1, \mu t_2^2 + (1-\mu)t_2^1, \mu\left(\lambda x_{t_1^1} + (1-\lambda) y_{t_1^1,t_2^2}\right) + (1-\mu)\left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Bigg) d\mu$$
$$+ \left(\lambda\left(x_{t_1^1,t_2^2} - x_{t_1^1,t_2^1}\right) + (1-\lambda)\left(y_{t_1^1,t_2^2} - y_{t_1^1,t_2^1}\right)\right) \times$$
$$\times \int_0^1 \sigma''_{(3,3)}\Bigg(t_1^1, \mu t_2^2 + (1-\mu)t_2^1, \mu\left(\lambda x_{t_1^1} + (1-\lambda) y_{t_1^1,t_2^2}\right) + (1-\mu)\left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Bigg) d\mu,$$

we can now rewrite $A - B - C + D$ as follows

$$
\begin{aligned}
&A - B - C + D = \\
&\quad = A' + B' + C' \\
&\quad\quad + \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right)\left(t_2^2 - t_2^1\right) \times \\
&\quad\quad\quad \underbrace{\begin{aligned}\times \int_0^1 \Bigg[\int_0^1 \Bigg(\sigma''_{(3,2)}\Bigg(& t_1^2, \mu t_2^2 + (1-\mu) t_2^1, \\
& \mu\left(\lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
& \quad + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) \\
- \sigma''_{(3,2)}\Bigg(& t_1^1, \mu t_2^2 + (1-\mu) t_2^1, \\
& \mu\left(\lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \\
& \quad +(1-\mu)\left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Bigg)\Bigg) d\mu\Bigg] d\lambda\end{aligned}}_{=D''} \\
&\quad\quad + \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \times \\
&\quad\quad\quad \underbrace{\begin{aligned}\times \int_0^1 \Bigg[& \left(\lambda\left(x_{t_1^2,t_2^2} - x_{t_1^2,t_2^1}\right) + (1-\lambda)\left(y_{t_1^2,t_2^2} - y_{t_1^2,t_2^1}\right)\right) \times \\
& \times \int_0^1 \sigma''_{(3,3)}\Bigg(t_1^2, \mu t_2^2 + (1-\mu) t_2^1, \\
& \qquad \mu\left(\lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
& \qquad\quad + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) d\mu \\
& - \left(\lambda\left(x_{t_1^1,t_2^2} - x_{t_1^1,t_2^1}\right) + (1-\lambda)\left(y_{t_1^1,t_2^2} - y_{t_1^1,t_2^1}\right)\right) \times \\
& \times \int_0^1 \sigma''_{(3,3)}\Bigg(t_1^1, \mu t_2^2 + (1-\mu) t_2^1, \\
& \qquad \mu\left(\lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \\
& \qquad\quad +(1-\mu)\left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Bigg) d\mu\Bigg] d\lambda\,.\end{aligned}}_{=E''}
\end{aligned}
$$

We can rewrite E'' as:

$$
\begin{aligned}
E'' = &\left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \times \\
&\underbrace{\begin{aligned}\times \int_0^1 \Bigg[&\left(\lambda x\left([t_1^1,t_1^2]\times[t_2^1,t_2^2]\right) + (1-\lambda) y\left([t_1^1,t_1^2]\times[t_2^1,t_2^2]\right)\right) \times \\
&\times \int_0^1 \sigma''_{(3,3)}\Bigg(t_1^2, \mu t_2^2 + (1-\mu)t_2^1, \\
&\qquad \mu\left(\lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
&\qquad\qquad + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) d\mu \Bigg] d\lambda\end{aligned}}_{=E'''} \\
&+ \left(x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}\right) \times \\
&\underbrace{\begin{aligned}\times \int_0^1 \Bigg[&\left(\lambda\left(x_{t_1^1,t_2^2} - x_{t_1^1,t_2^1}\right) + (1-\lambda)\left(y_{t_1^1,t_2^2} - y_{t_1^1,t_2^1}\right)\right) \times \\
&\times \int_0^1 \Bigg[\sigma''_{(3,3)}\Bigg(t_1^2, \mu t_2^2 + (1-\mu)t_2^1, \\
&\qquad \mu\left(\lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}\right) \\
&\qquad\qquad + (1-\mu)\left(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}\right)\Bigg) \\
&\quad - \sigma''_{(3,3)}\Bigg(t_1^1, \mu t_2^2 + (1-\mu)t_2^1, \\
&\qquad \mu\left(\lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) \\
&\qquad\qquad + (1-\mu)\left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right)\Bigg)\Bigg] d\mu \Bigg] d\lambda\,.\end{aligned}}_{=F'''}
\end{aligned}
$$

Now, we see

$$
\begin{aligned}
|A'| &\le \|x-y\|_{[z^1,z^2],\alpha} |t_1^2 - t_1^1|^{\alpha_1} |t_2^2 - t_2^1|^{\alpha_2} \left\|\sigma'_{(3)}\right\|_\infty, \\
|B'| &\le \left\|x_{\cdot,t_2^1} - y_{\cdot,t_2^1}\right\|_{[t_1^1,t_1^2],\alpha_1} |t_1^2 - t_1^1|^{\alpha_1} \Bigg(|t_2^2 - t_2^1| \left\|\sigma''_{(3,2)}\right\|_\infty + \\
&\qquad + \left\|\sigma''_{(3,2)}\right\|_\infty \int_0^1 \lambda(x_{t_1^2,t_2^2} - x_{t_1^2,t_2^1}) + (1-\lambda)(y_{t_1^2,t_2^2} - y_{t_1^2,t_2^1}) \Bigg),
\end{aligned}
$$

similarly

$$|C'| \leq \left\|x_{t_1^1,\cdot} - y_{t_1^1,\cdot}\right\|_{[t_2^1,t_2^2],\alpha_2} |t_2^2 - t_2^1|^{\alpha_2} \Bigg(|t_1^2 - t_1^1| \left\|\sigma''_{(3,2)}\right\|_\infty + \\ + \left\|\sigma''_{(3,2)}\right\|_\infty \int_0^1 \lambda(x_{t_1^2,t_2^2} - x_{t_1^1,t_2^2}) + (1-\lambda)(y_{t_1^2,t_2^2} - y_{t_1^1,t_2^2}) \Bigg),$$

$$|D''| \leq |x_{t_1^1,t_2^1} - yt_1^1, t_2^1|(t_2^2 - t_2^1)(t_1^2 - t_1^1) \left\|\sigma'''_{(3,2,1)}\right\|_\infty \\ + |x_{t_1^1,t_2^1} - yt_1^1, t_2^1| \left\|\sigma'''_{(3,2,3)}\right\|_\infty (t_2^2 - t_2^1) \times \\ \times \Bigg| \int_0^1 \int_0^1 \mu(\lambda x_{t_1^2,t_2^2} + (1-\lambda) y_{t_1^2,t_2^2}) + (1-\mu)(\lambda x_{t_1^2,t_2^1} + (1-\lambda) y_{t_1^2,t_2^1}) \\ - \mu(\lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}) - (1-\mu)(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}) d\mu d\lambda \Bigg|,$$

$$|E'''| \leq |x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}| \left\|\sigma''_{(3,3)}\right\|_\infty \left| \int_0^1 (\lambda x((t^1,t^2]) + (1-\lambda) y((t^1,t^2])) d\lambda \right|,$$

$$|F'''| \leq |x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}| \left\|\sigma'''_{(3,3,1)}\right\|_\infty (t_1^2 - t_1^1) \times \\ \times \int_0^1 \lambda(x_{t_1^1,t_2^2} - x_{t_1^1,t_2^1}) + (1-\lambda)(y_{t_1^1,t_2^2} - y_{t_1^1,t_2^1}) d\lambda \\ + |x_{t_1^1,t_2^1} - y_{t_1^1,t_2^1}| \left\|\sigma'''_{(3,3,3)}\right\|_\infty \int_0^1 \Bigg[\lambda(x_{t_1^1,t_2^2} - x_{t_1^1,t_2^1}) + (1-\lambda)(y_{t_1^1,t_2^2} - y_{t_1^1,t_2^1}) \times \\ \times \int_0^1 \mu \left(\lambda x_{t_1^1,t_2^2} + (1-\lambda) y_{t_1^1,t_2^2}\right) + (1-\mu) \left(\lambda x_{t_1^1,t_2^1} + (1-\lambda) y_{t_1^1,t_2^1}\right) \Bigg] d\mu d\lambda.$$

Thus, if $x, y \in \mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([t^1, t^1 + \epsilon_1])$ then there is a constant, C, based on K, the bounds σ and the bounds on σ's derivatives, such that

$$\|\sigma(\cdot,\cdot,x) - \sigma(\cdot,\cdot,x)\|_{[t^1,t^2],\alpha} \leq C \times \|x - y\|_{[t^1,t^2],\alpha}. \tag{3.11}$$

Thus, from Eq. (3.8), Eq. (3.9) and Eq. (3.11), there exists an ϵ_2, independent of t^1 and t^2, such that

$$\|Fx - Fy\|_{[t^1,t^1+\epsilon_2],\alpha,\infty} \leq \delta \|x - y\|_{[t^1,t^1+\epsilon_2],\alpha,\infty} \tag{3.12}$$

for some $\delta \in (0,1)$. Thus, if we let $\epsilon_0 = \min(\epsilon_1, \epsilon_2)$, we have that

$$F : \mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([t^1, t^1 + \epsilon_0]) \to \mathscr{C}_{\alpha,\infty,K,\phi_1,\phi_2}([t^1, t^1 + \epsilon_0])$$

is a contraction. □

Theorem III.13. *Under the same assumptions as Proposition III.12, the ordinary differential equation*

$$dx(t) = b(t, x(t))dt + \sigma(t, x(t))dg(t), \qquad x(0) = x_0 \in \mathbb{R}, \tag{3.13}$$

has a unique solution in $\mathscr{C}_\beta([0, \mathbb{T}])$ *where* $\beta = (\beta_1, \beta_2)$.

Proof. We now iteratively apply the result of Proposition III.12 in each direction in turn to have the global existence of a unique solution on

$$\begin{aligned}
[0,\epsilon]\times[0,\epsilon] &\xrightarrow{\text{by Proposition III.12}} [0,2\epsilon]\times[0,\epsilon] \\
&\vdots \\
&\xrightarrow{\text{by Proposition III.12}} [0,T_1]\times[0,\epsilon] \\
&\xrightarrow{\text{by Proposition III.12}} [0,T_1]\times[0,2\epsilon] \\
&\vdots \\
&\xrightarrow{\text{by Proposition III.12}} [0,T_1]\times[0,T_2].
\end{aligned}$$

Uniqueness follows directly from Eq. (3.12). □

We now state the desired 2-parameter result.

Theorem III.14. *Let* $(V_t^\gamma)_{t\in\mathbb{T}}$ *be a 2-dimensional Volterra process with* γ*-Hölder paths where* $\gamma = (\gamma_1, \gamma_2) \in \left(\frac{1}{2}, 1\right]^2$ *and let* α_i, β_i *be such that* $\frac{1}{2} < \beta_i < \gamma_i$*, and* $\beta_i > \alpha_i > 1 - \beta_i$ $i = 1, 2, \ldots, d$*. Let* b *and* σ *satisfy the same hypothesis as in the above proposition. Then the ordinary differential equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dV_t, \qquad X_0 = x_0 \in \mathbb{R}, \tag{3.14}$$

has a unique solution in $\mathscr{C}_\beta([0, \mathbb{T}])$ *with probability 1, where* $\beta = (\beta_1, \ldots, \beta_d)$.

Proof. Since $\alpha_i + \beta_i > 1$ for all $i \in \{1, \dots, d\}$, the integral $\int_0^t f(t)dV_t$ is almost surely well defined for $f \in \mathscr{C}_\alpha([0, \mathbb{T}])$. Then the result follows directly from Theorem III.13 applied point-wise. □

3.4 Maximum Likelihood estimation

As in Chapter II, this section will concern estimating parameters for stochastic differential equations of the following form:

$$dX_t = A(t, X_t, \theta)dt + \sigma(t)dV_t^\gamma,$$

where V^γ is a 2-parameter Volterra process, $\theta \in \Theta \subset \mathbb{R}$ and $\sigma(t)$ is a positive, non-vanishing function on $[0, T] = [0, T_1] \times [0, T_2]$.

First, we define the fundamental 1-parameter martingales associated with K^{γ_1} and K^{γ_2}. If V^{γ_i}, $i = 1, 2$, is the 1-parameter Volterra process defined by K^{γ_i} $i = 1, 2$, then $N_{t_i}^{*,i}$, defined by

$$N_{t_i}^{*,i} = \int_0^{t_i} k_{\mathbb{1}}^{s_i} dV_{s_i}^{\gamma_i} \quad i = 1, 2,$$

where $k_{\mathbb{1}}^{s_i}$ is the kernel defined in Eq. (3.1) with $\mathcal{C}(s) = 1$, for all s, for V^{γ_i} $i = 1, 2$, is the fundamental 1-parameter martingale associated with V^{γ_i} $i = 1, 2$.

Next, let

$$k_{\mathbb{1}}^t(s) = k_{\mathbb{1}}^{t_1}(s_1)k_{\mathbb{1}}^{t_2}(s_2).$$

Then, using Theorem III.9, with $s = (s_1, s_2)$,

$$N_t^* = \int_0^t k_{\mathbb{1}}^t(s)dV_s^\gamma$$

is the fundamental strong 2-parameter Gaussian martingale associated with V^γ.

3.4.1 Maximum likelihood estimator when drift is a polynomial in θ.

We have the following result on the maximum likelihood estimate of θ.

Theorem III.15. *Under the assumptions of Proposition III.12 and Theorem III.14, let V^γ be a 2-parameter Volterra process. Define the process $X = (X_t)_{t\in[0,T]}$ by the equations*

$$dX_t = A(t, X_t, \theta)dt + \sigma(t)dV_t^\gamma, \quad t \in (0, T)$$

$$X_0 = \xi \text{ a.s.}$$

where $A(t, X_t, \theta) = \sum_{i=0}^p a_i(t_1, t_2, X_{t_1,t_2})\theta^i$ and σ (a positive non-vanishing function on $[0, T]$) are known functions and $\theta \in \Theta \subset \mathbb{R}$. Assuming σ is bounded and has bounded third derivatives, the maximum likelihood estimator, $\hat{\theta}_T$, of θ is given by:

$$\operatorname*{argmax}_{\theta\in\Theta} \int_0^T Q_\theta(t)dU_t - \frac{1}{2}\int_0^T Q_\theta^2(t_1, t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2},$$

where

$$Q_\theta(t_1, t_2) = \frac{d}{d\left\langle N^{*,1}\right\rangle_{t_1}} \frac{d}{d\left\langle N^{*,2}\right\rangle_{t_2}} \int_0^{t_1}\int_0^{t_2} k_{\mathbb{1}}^{t_1,t_2}(s_1, s_2)\frac{A(s_1, s_2, X_{s_1,s_2}, \theta)}{\sigma(s_1, s_2)}ds_2ds_1,$$

and

$$U_{t_1,t_2} = \int_0^{t_1}\int_0^{t_2} Q_\theta(s_1, s_2)d\left\langle N^{*,2}\right\rangle_{s_2} d\left\langle N^{*,1}\right\rangle_{s_1} + N_{t_1,t_2}^*.$$

Proof. We let $\mathbf{P}_\theta^T$ be the measure induced by the process $\{X_t; 0 \preccurlyeq t \preccurlyeq T\}$ when θ is the true parameter. We then have that the Radon-Nikodym derivative of $\mathbf{P}_\theta^T$ with respect to $\mathbf{P}_0^T$ is given by:

$$\frac{d\mathbf{P}_\theta^T}{d\mathbf{P}_0^T} = \exp\left\{\int_0^T Q_\theta(t)dU_t - \frac{1}{2}\int_0^T Q_\theta^2(t_1, t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}\right\}.$$

Thus, the log-likelihood function is given by

$$l_T(\theta) = \int_0^T Q_\theta(t)dU_t - \frac{1}{2}\int_0^T Q_\theta^2(t_1, t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2},$$

and the MLE is given by

$$\hat{\theta}_T = \operatorname*{argmax}_{\theta\in\Theta} l_T(\theta).$$

$\square$

3.4.2 Case of linear drift

A specific case of interest is when the drift function, $A(t, X_t, \theta)$ is linear in θ, i.e.

$$A(t, X_t, \theta) = a_0(t, X_t) + \theta\, a_1(t, X_t).$$

In this case, we have an analytic expression for the MLE as given in the following corollary.

Corollary III.16. *Under the assumptions of the above Theorem, when the drift term is linear in* θ*, the MLE,* $\hat{\theta}_T$*, of* θ *is given by:*

$$\hat{\theta}_T = \frac{\int_0^{T_2}\int_0^{T_1} J_1(t_1,t_2)dU_{t_1,t_2} - \int_0^{T_1}\int_0^{T_2} J_0(t_1,t_2)J_1(t_1,t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}}{\int_0^{T_2}\int_0^{T_1} J_1^2(t_1,t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}},$$

where

$$J_i(t_1,t_2) = \frac{d}{d\left\langle N^{*,1}\right\rangle_{t_1}}\frac{d}{d\left\langle N^{*,2}\right\rangle_{t_2}}\int_0^{t_1}\int_0^{t_2} k_{\mathbb{1}}^{t_1,t_2}(s_1,s_2)\frac{a_i(s_1,s_2,X_{s_1,s_2})}{\sigma(s_1,s_2)}ds_2ds_1.$$

Proof. We see directly that in the linear case,

$$Q_\theta(t) = J_0(t) + J_1(t)\cdot\theta$$

Thus, the likelihood equation is

$$\int_0^T J_1(t)dU_t - \int_0^{T_2}\int_0^{T_1}(J_0(t_1,t_2) + \theta\, J_1(t_1,t_2))J_1(t_1,t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2} = 0,$$

and therefore, the MLE $\hat{\theta}_T$ is given by

$$\hat{\theta}_T = \frac{\int_0^{T_2}\int_0^{T_1} J_1(t_1,t_2)dU_{t_1,t_2} - \int_0^{T_1}\int_0^{T_2} J_0(t_1,t_2)J_1(t_1,t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}}{\int_0^{T_2}\int_0^{T_1} J_1^2(t_1,t_2)d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}}.$$

□

3.4.3 Properties of Maximum Likelihood Estimate

While there are considerably less multi-parameter martingale results available, we have, using sectorial limits, the following strong law of large numbers.

Lemma III.17. *Assume X is a strong, 2-parameter martingale with $\langle X\rangle_{\cdot,\infty} = \langle X\rangle_{\infty,\cdot} = \infty$ a.s.. Then*

$$\lim_{t\rightsquigarrow\infty} \frac{X_t}{\langle X\rangle_t} = 0 \text{ a.s..}$$

Proof. Fix t_1. Then by Corollary 1, p. 144 in [35]],

$$\begin{aligned} & \lim_{t_2\to\infty} \frac{X_{t_1,t_2}}{\langle X\rangle_{t_1,t_2}} = 0 \text{ a.s.} \\ \Rightarrow \qquad & \lim_{t_1\to\infty}\lim_{t_2\to\infty} \frac{X_{t_1,t_2}}{\langle X\rangle_{t_1,t_2}} = 0 \text{ a.s..} \end{aligned}$$

Similarly, with $\pi_2 : (1,2) \mapsto (2,1)$,

$$\pi_2 - \lim_{t\to\infty} \frac{X_t}{\langle X\rangle_t} = 0 \text{ a.s..}$$

Thus the desired result is shown. □

Now, we show that the estimator is sectorially strongly consistent (i.e. if θ_0 is the true parameter, $\lim_{T\rightsquigarrow\infty} \hat{\theta}_T - \theta_0 = 0$ a.s.).

Theorem III.18. *The MLE, $\hat{\theta}_T$, is sectorially strongly consistent provided*

$$\int_0^T J_1^2(t) d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2} \to \infty \quad \text{a.s. } [\mathbf{P}_{\theta_0}] \text{ as T} \to \infty.$$

Proof. Let θ_0 be the true parameter. Then, since

$$dU_t = (J_0(t) + \theta_0 J_1(t))\, d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2} + dN_t^*,$$

we have

$$\frac{d\mathbf{P}_\theta^{T_1}}{d\mathbf{P}_{\theta_0}^{T_1}} = \exp\left\{(\theta-\theta_0)\int_0^T J_1(t) dN_t^* - \frac{1}{2}(\theta-\theta_0)^2 \int_0^T J_1^2(t) d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}\right\}.$$

Following this representation of the Radon-Nikodym derivative, we obtain that

$$\hat{\theta}_{T_1} - \theta_0 = \frac{\int_0^T J_1(t) dN_t^*}{\int_0^T J_1^2(t) d\left\langle N^{*,1}\right\rangle_{t_1} d\left\langle N^{*,2}\right\rangle_{t_2}}$$

Thus

$$R_T \equiv \int_0^T J_1(t) dN_t^* \tag{3.15}$$

is a strong 2-parameter martingale with the quadratic variation process

$$\langle R \rangle_T = \int_0^T J_1^2(t) d \left\langle N^{*,1} \right\rangle_{t_1} d \left\langle N^{*,2} \right\rangle_{t_2}, \tag{3.16}$$

and the result follows directly from Corollary III.17. □

CHAPTER IV

Parameter estimation in Integrals of functions of Brownian Motion

4.1 Introduction

In this chapter, we consider statistical inference based on observing a process, $X_t = X(t;\theta)$ defined by the following stochastic differential equation

$$dX_t = f(\mathbb{B}_t;\theta)dt \quad X_0 = 0.$$

This stochastic differential equation is not of the form considered in earlier chapters where there is a non-random drift component and a random volatility component. The problem in this chapter is motivated by the need to develop degradation models where the degradation process has non-decreasing sample paths.

Degradation data occur in the analysis of survival and reliability data where one observes how the performance of a subject or device changes over time. The increased availability of sensor technology has made it possible to collect and analyze data on how devices "age" over time. This is becoming more common in the monitoring and maintenance of expensive systems, sometimes called predictive or condition-based maintenance [12]. In time, one can anticipate such techniques being used with patients' health care as we move even more to electronic medical records.

Most of the literature in longitudinal data analysis, growth curves, and even degradation models assume that the data at each point in time are Gaussian. For example, Brownian motion with linear drift has been used to model degradation data in reliability applications. Part of the reason for the popularity of this model is that the time-to-failure, defined as the first-passage time of the process over a certain threshold, has been developed and is known to follow an inverse-Gaussian distribution [11]. Nair and Wang [40] and others have considered time-transformed versions of this process that accommodates more complex degradation shapes.

But one disadvantage with these models is that they do not have non-decreasing sample paths (degradation levels), which is common in many applications. The methods developed in this section is a first effort at addressing this problem. We consider processes that are integrals of positive powers of the Brownian motion with zero mean. Specifically, the process is $X_t = \int_0^t (\mathbb{B}_s^2)^\theta \, ds$. We observe X_{jt_k} for different devices or subjects at time points $t_{k_1}, \ldots, t_{k_j}$. The goal is to make inference about θ and the distribution of the underlying "time-to-failure" $T_D = \inf\{t : X(t) = D\}$ for some fixed D, including prediction of the conditional distribution to failure given past observations of X_t.

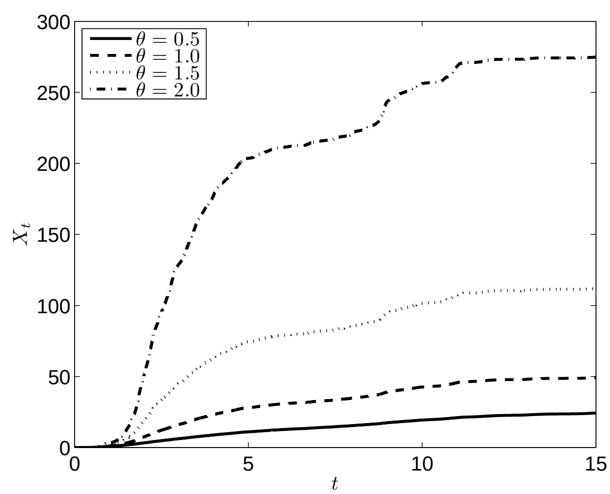


Figure 4.1: Sample path X_t for $\theta = 0.5, 1, 1.5$ and 2

4.2 Preliminaries and Problem Statement

4.2.1 Notation and Definitions

For simplicity of presentation, we will adopt the following notation (again $\triangleq$ means 'is denoted as'):

1. $V_{\mu,\sigma}$ will be a Normal random variable with mean μ and standard deviation σ.

2. For $\theta > 0$, $Y_t \triangleq (\mathbb{B}_t^2)^\theta$.

3. For $\theta > 0$,

$$X_t \triangleq \int_0^t Y_t ds = \int_0^t \left(\mathbb{B}_s^2\right)^\theta ds. \tag{4.1}$$

4. For $D > 0$, define T_D as

$$T_D = \inf_{t \geq 0} \{t : X_t = D\}. \tag{4.2}$$

5. U is a Chi-square random variable with 1 degree of freedom. The density function of U is

$$f_U(u) = \frac{\mathbf{e}^{-\frac{u}{2}}}{\sqrt{2\pi u}}.$$

6. U_λ is a non-central Chi-square random variable with 1 degree of freedom and non-centrality parameter $\lambda > 0$. The density function of U_λ is

$$f_{U_\lambda}(u) = \frac{1}{2}\mathbf{e}^{-\frac{u+\lambda}{2}} \left(\frac{u}{\lambda}\right)^{-\frac{1}{4}} I_{-\frac{1}{2}}(\sqrt{\lambda u}).$$

where $I_a(y)$ is the modified Bessel function of the first kind (see for example [1]) given by

$$I_a(y) = \left(\frac{y}{2}\right)^a \sum_{j=0}^{\infty} \frac{\left(\frac{y^2}{4}\right)^j}{j!\Gamma[a+j+1]}.$$

Note that

$$\left(\frac{V_{\mu,\sigma}}{\sigma}\right)^2 \sim U_\lambda \qquad \text{with } \lambda = \left(\frac{\mu}{\sigma}\right)^2.$$

We will make use of the following two propositions that analytically define fractional moments of the two above random variables.

Proposition IV.1. *If U is a Chi-Square random variable with 1 degree of freedom for any $\theta > 0$, we have*

$$\mathbb{E}\left[U^{\theta}\right] = \frac{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta}}{\sqrt{\pi}}. \tag{4.3}$$

Proof. Recall the density of a Chi-square random variable with 1 degree of freedom is

$$f_U(u) = \frac{\mathbf{e}^{-\frac{u}{2}}}{\sqrt{2\pi u}}. \tag{4.4}$$

By direct calculation we have

$$\begin{aligned}
\mathbb{E}\left[U^{\theta}\right] &= \int_0^{\infty} u^{\theta} f_U(u) du \\
&= \int_0^{\infty} u^{\theta} \frac{\mathbf{e}^{-\frac{u}{2}}}{\sqrt{2\pi u}} du \\
&= \int_0^{\infty} \frac{u^{\left(\theta+\frac{1}{2}\right)-1}\, \mathbf{e}^{-\frac{u}{2}}}{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta+\frac{1}{2}}} \left(\frac{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta+\frac{1}{2}}}{\sqrt{2\pi}}\right) du \\
&= \left(\frac{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta+\frac{1}{2}}}{\sqrt{2\pi}}\right) \underbrace{\int_0^{\infty} \frac{u^{\left(\theta+\frac{1}{2}\right)-1}\, \mathbf{e}^{-\frac{u}{2}}}{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta+\frac{1}{2}}} du}_{=1} \\
&= \frac{\Gamma\left[\theta + \frac{1}{2}\right] 2^{\theta}}{\sqrt{\pi}}.
\end{aligned}$$

□

We also have the following more general result.

Proposition IV.2. *If U_λ is a non-central Chi-Square random variable with 1 degree of freedom and non-centrality parameter $\lambda > 0$, for any $\theta > 0$ we have*

$$\mathbb{E}\left[U_{\lambda}^{\theta}\right] = \frac{2^{\theta}\, \mathbf{e}^{-\frac{\lambda}{2}} \Gamma\left[\theta + \frac{1}{2}\right] {}_1F_1\left(\frac{1}{2} + \theta, \frac{1}{2}, \frac{\lambda}{2}\right)}{\sqrt{\pi}}, \tag{4.5}$$

where ${}_1F_1(a;b;z)$ is the Confluent Hypergeometric Function of the First Kind (see for example [1]).

Proof. Recall the density of a non-central Chi-Square random variable with 1 degree of freedom and non-centrality parameter λ is

$$f_{U_\lambda}(u) = \frac{1}{2}\,\mathbf{e}^{-\frac{u+\lambda}{2}} \left(\frac{u}{\lambda}\right)^{-\frac{1}{4}} I_{-\frac{1}{2}}(\sqrt{\lambda u}). \tag{4.6}$$

Thus we have,

$$\begin{aligned}
\mathbb{E}\left[U_\lambda^\theta\right] &= \int_0^\infty u^\theta f_U(u)du \\
&= \int_0^\infty u^\theta \frac{1}{2}\,\mathbf{e}^{-\frac{u+\lambda}{2}} \left(\frac{u}{\lambda}\right)^{-\frac{1}{4}} \left(\frac{\sqrt{\lambda u}}{2}\right)^{-\frac{1}{2}} \sum_{j=0}^\infty \frac{\left(\frac{\lambda u}{4}\right)^j}{j!\Gamma[j+\frac{1}{2}]} du \\
&= \sum_{j=0}^\infty \frac{1}{2^{\frac{1}{2}+2j}}\,\mathbf{e}^{-\frac{\lambda}{2}} \frac{\lambda^j}{j!\Gamma\left[j+\frac{1}{2}\right]} \underbrace{\int_0^\infty u^{\theta+j-\frac{1}{2}}\,\mathbf{e}^{-\frac{1}{2}} da}_{=2^{\frac{1}{2}+j+\theta}\Gamma\left[\frac{1}{2}+j+\theta\right]} \\
&= 2^\theta \sum_{j=0}^\infty \frac{\left(\frac{\lambda}{2}\right)^j \mathbf{e}^{-\frac{\lambda}{2}}}{j!} \frac{\Gamma\left[j+\theta+\frac{1}{2}\right]}{\Gamma\left[j+\frac{1}{2}\right]} \\
&= 2^\theta * \mathbf{e}^{-\frac{\lambda}{2}} \frac{\Gamma\left[\theta+\frac{1}{2}\right]}{\Gamma\left[\frac{1}{2}\right]} \underbrace{\sum_{i=0}^\infty \frac{\left(\frac{1}{2}+\theta\right)_j}{\left(\frac{1}{2}\right)_j} \frac{\left(\frac{\lambda}{2}\right)^j}{j!}}_{={}_1F_1\left(\frac{1}{2}+\theta,\frac{1}{2},\frac{\lambda}{2}\right)},
\end{aligned}$$

where

$$(a)_n = a(a+1)(a+2)\cdot(a+n-1),$$

and ${}_1F_1(a;b;z)$ is defined as in [1]. □

Remark IV.3. Eq. (4.5) reduces to Eq. (4.3) when $\lambda = 0$, i.e. for $\theta > 0$,

$$\mathbb{E}\left[U_{\lambda=0}^\theta\right] = \mathbb{E}\left[U^\theta\right].$$

The processes that will be dealt with in this chapter are integrals of Brownian Motion. We recall its definition.

Definition IV.4 (Brownian Motion)**.** Standard Brownian motion on the positive real line, denoted $\{\mathbb{B}_t, 0 \leq t \leq \infty\}$, is a continuous random process defined by the following four properties:

1. $\mathbb{B}_0 = 0$.
2. $\mathbb{B}_t$ is almost surely continuous.
3. $\mathbb{B}_t$ has independent increments, i.e. for $0 \leq s \leq t \leq u \leq v$, $\mathbb{B}_v - \mathbb{B}_u$ and $\mathbb{B}_t - \mathbb{B}_s$ are independent random variables.
4. For $t \geq s \geq 0$, $\mathbb{B}_t - \mathbb{B}_s \sim N\left(0, \sigma^2 = t - s\right)$.

4.2.2 Problem Statement

We are concerned with estimating a parameter in the following simple differential equation:

$$dX_t = \left(\mathbb{B}_t^2\right)^\theta dt \quad \text{with } X_0 = 0 \tag{4.7}$$

As stated above, $\{X_t, t \geq 0\}$ is the following continuous, monotonically non-decreasing random process

$$X_t \triangleq \int_0^t \left(\mathbb{B}_s^2\right)^\theta ds$$

for a fixed $\theta > 0$. We will also be concerned with estimating the first passage time of the process, i.e. the time at which the process first equals a given fixed level, $D > 0$, which we will denote as T_D. Since the process starts at zero and is almost surely increasing, this time is unique for each path of the process. Unfortunately we are not able to directly estimate T_D based on observations, so we will instead follow the approach of first estimating θ based on observations of X_t, and then, through simulation, estimate T_D.

Thus, our first estimation problem concerns θ. We will consider two different observation setups.

- Non-identically distributed observations

  We will first consider a general situation where we observe n independent paths of X_t once each. The times that each path is observed may not be the same across the n different paths.

- Identically distributed observations

  Second we will consider a simpler sub-case where we again observe n independent paths of X_t once each, but now all n processes are observed at the same time.

The primary difficulty that must be overcome in the more general case is that the observations are not identically distributed.

Once we obtain an estimate of θ, we can easily simulate many paths of X_t, compute T_D for each one of them and by using the sample average, obtain an estimate of T_D. In fact, through this method, we will actually be obtaining an estimate of the distribution of T_D, which we will be able to use to provide prediction intervals based on the simulations.

4.3 Estimation of θ

4.3.1 Moments of X_t

Due to the complexity of the process X_t, the approach of maximum likelihood estimation of θ, which requires knowledge of the probability distribution of X_t, is intractable. However, through direct computation, the first two moments of X_t can be analytically obtained, which leads us to several moment-based estimators. To this end, we have the following proposition:

Proposition IV.5. *For a fixed $\theta > 0$ and $t > 0$, the first two moments of X_T are given by:*

(4.8)
$$\mathbb{E}\left[X_t\right] = \frac{\Gamma\left[\theta + \frac{1}{2}\right]}{2(\theta+1)\sqrt{\pi}}(2t)^{\theta+1}.$$

(4.9)
$$\mathbb{E}\left[(X_t)^2\right] = \frac{4^\theta \Gamma\left[\theta+\frac{1}{2}\right]^2 \Gamma[\theta+1]\Gamma\left[2\theta+\frac{3}{2}\right]}{\pi\Gamma\left[3\theta+\frac{5}{2}\right](\theta+1)} {}_3F_2\left(\left\{\theta+\frac{1}{2}, \theta+\frac{1}{2}, \theta+1\right\}, \left\{\frac{1}{2}, 3\theta+\frac{5}{2}\right\}, 1\right) t^{2(\theta+1)}.$$

Proof. We first prove Eq. (4.8). We have

$$\begin{aligned}
\mathbb{E}\left[X_t\right] &= \mathbb{E}\left[\int_0^t \left(\mathbb{B}_s^2\right)^\theta ds\right] \\
&= \int_0^t \mathbb{E}\left[\left(\mathbb{B}_s^2\right)^\theta\right] ds.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}\left[\left(\mathbb{B}_s^2\right)^\theta\right] &= \mathbb{E}\left[\left(V_{0,\sqrt{s}}^2\right)^\theta\right] \\
&= \mathbb{E}\left[\left(\left(\sqrt{s}\frac{V_{0,\sqrt{s}}}{\sqrt{s}}\right)^2\right)^\theta\right] \\
&= s^\theta \mathbb{E}\left[\left(V_{0,1}^2\right)^\theta\right] \\
&= s^\theta \mathbb{E}\left[U^\theta\right] \\
&= \frac{\Gamma\left[\theta+\frac{1}{2}\right]}{\sqrt{\pi}}(2s)^\theta,
\end{aligned}$$

where the last line follows from Eq. (4.3). Thus,

$$\begin{aligned}
\mathbb{E}\left[X_t\right] &= \int_0^t \frac{\Gamma\left[\theta+\frac{1}{2}\right]}{\sqrt{\pi}}(2s)^\theta ds \\
&= \frac{\Gamma\left[\theta+\frac{1}{2}\right]}{2(\theta+1)\sqrt{\pi}}(2t)^{\theta+1}.
\end{aligned}$$

Now we prove Eq. (4.9). First we notice that

$$\begin{aligned}
\mathbb{E}\left[(X_t)^2\right] &= \mathbb{E}\left[\left(\int_0^t Y_s ds\right)^2\right] \\
&= \mathbb{E}\left[\int_0^t \int_0^t Y_s Y_{s'} ds' ds\right] \\
&= \mathbb{E}\left[2\int_0^t \int_0^t \mathbb{1}_{[s'<s]} Y_s Y_{s'} ds' ds\right] \quad \text{by symmetry} \\
&= 2\mathbb{E}\left[\int_0^t \int_0^s Y_s Y_{s'} ds' ds\right] \\
&= 2\int_0^t \int_0^s \mathbb{E}\left[Y_s Y_{s'}\right] ds' ds. \qquad (4.10)
\end{aligned}$$

For $s' < s$, we have

$$\begin{aligned}
\mathbb{E}\left[Y_s Y_{s'}\right] &= \int_{-\infty}^{\infty} \mathbb{E}\left[Y_s Y_{s'} \middle| \mathbb{B}_{s'} = a\right] \mathbf{P}\left(\mathbb{B}_{s'} = a\right) da \\
&= \int_{-\infty}^{\infty} (a^2)^\theta \mathbb{E}\left[Y_s \middle| \mathbb{B}_{s'} = a\right] \mathbf{P}\left(V_{0,\sqrt{s'}} = a\right) da.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}\left[Y_s \middle| \mathbb{B}_{s'} = a\right] &= \mathbb{E}\left[\left(V^2_{a,\sqrt{s-s'}}\right)^\theta\right] \\
&= \mathbb{E}\left[\left((s-s')U_\lambda^2\right)^\theta\right] \quad \text{with } \lambda = \frac{a^2}{s-s'} \\
&= (s-s')^\theta \mathbb{E}\left[\left(U_\lambda^2\right)^\theta\right] \\
&= \frac{(s-s')^\theta 2^\theta \, \mathbf{e}^{-\frac{\lambda}{2}} \Gamma\left[\theta + \frac{1}{2}\right] {}_1F_1\left(\theta + \frac{1}{2}, \frac{1}{2}, \frac{\lambda}{2}\right)}{\sqrt{\pi}} \\
&= \left(2(s-s')\right)^\theta \mathbf{e}^{-\frac{\lambda}{2}} \sum_{n=0}^{\infty} \frac{\Gamma\left[\theta + \frac{1}{2} + n\right]}{\Gamma\left[\frac{1}{2} + n\right]} \frac{\left(\frac{\lambda}{2}\right)^n}{n!}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
\mathbb{E}\left[Y_s Y_{s'}\right] &= \int_{-\infty}^{\infty} (a^2)^{\theta} \left((2(s-s'))^{\theta}\, \mathbf{e}^{-\frac{\lambda}{2}} \sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]}{\Gamma\left[\frac{1}{2}+n\right]} \frac{\left(\frac{\lambda}{2}\right)^n}{n!} \right) \left(\frac{1}{\sqrt{2\pi s'}}\, \mathbf{e}^{-\frac{a^2}{2s}} \right) da \\
&= \frac{(2(s-s'))^{\theta}}{\sqrt{2\pi s'}} \sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]}{\Gamma\left[\frac{1}{2}+n\right] n!} \underbrace{\int_{-\infty}^{\infty} (a^2)^{\theta}\, \mathbf{e}^{-\frac{a^2}{2(s-s')}-\frac{a^2}{2s'}} \left(\frac{a^2}{2(s-s')} \right)^n da}_{=2^{\frac{1}{2}+\theta}\left(\frac{s'}{s}\right)^{\theta+\frac{1}{2}+n}(s-s')^{\theta+\frac{1}{2}}\Gamma\left[\theta+\frac{1}{2}+n\right]} \\
&= \frac{4^{\theta}(s-s')^{2\theta+\frac{1}{2}}}{\sqrt{\pi s'}} \left(\frac{s'}{s}\right)^{\theta+\frac{1}{2}} \underbrace{\sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]^2}{\Gamma\left[\frac{1}{2}+n\right]} \frac{\left(\frac{s'}{s}\right)^n}{n!}}_{=\frac{\Gamma\left[\theta+\frac{1}{2}\right]^2}{\sqrt{\pi}} {}_2F_1\left(\theta+\frac{1}{2},\theta+\frac{1}{2},\frac{1}{2},\frac{s'}{s}\right)} \qquad (4.11) \\
&= \frac{4^{\theta}(s-s')^{2\theta+\frac{1}{2}}\Gamma\left[\theta+\frac{1}{2}\right]^2}{\pi\sqrt{s'}} \left(\frac{s'}{s}\right)^{\theta+\frac{1}{2}} {}_2F_1\left(\theta+\frac{1}{2},\theta+\frac{1}{2},\frac{1}{2},\frac{s'}{s}\right),
\end{aligned}
$$

where ${}_2F_1(a;b;z)$ is the Hypergeometric Function (see for example [1]). Now, from Eq. (4.11), we have

$$
\begin{aligned}
\int_0^s \mathbb{E}\left[Y_s Y_{s'}\right] ds' &= \frac{4^{\theta}}{\sqrt{\pi} s^{\theta+\frac{1}{2}}} \int_0^s (s')^{\theta}(s-s')^{2\theta+\frac{1}{2}} \sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]^2}{\Gamma\left[\frac{1}{2}+n\right]} \frac{(s')^n}{s^n n!} ds' \\
&= \frac{4^{\theta}}{\sqrt{\pi} s^{\theta+\frac{1}{2}}} \sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]^2}{\Gamma\left[\frac{1}{2}+n\right] s^n n!} \underbrace{\int_0^s (s')^{\theta+n}(s-s')^{2\theta+\frac{1}{2}} ds'}_{=\frac{s^{3\theta+\frac{3}{2}+n}\Gamma[\theta+1+n]\Gamma\left[2\theta+\frac{3}{2}\right]}{\Gamma\left[3\theta+\frac{5}{2}+n\right]}} \\
&= \frac{4^{\theta} s^{2\theta+1}\Gamma\left[2\theta+\frac{3}{2}\right]}{\sqrt{\pi}} \quad \underbrace{\sum_{n=0}^{\infty} \frac{\Gamma\left[\theta+\frac{1}{2}+n\right]^2 \Gamma\left[\theta+1+n\right]}{\Gamma\left[3\theta+\frac{5}{2}+n\right]\Gamma\left[\frac{1}{2}+n\right]} \frac{1}{n!}}_{\frac{\Gamma\left[\theta+\frac{1}{2}\right]^2\Gamma[\theta+1]}{\sqrt{\pi}\Gamma\left[3\theta+\frac{5}{2}\right]} {}_3F_2\left(\left\{\theta+\frac{1}{2},\theta+\frac{1}{2},\theta+1\right\},\left\{\frac{1}{2},3\theta+\frac{5}{2}\right\},1\right)} \\
&= \frac{4^{\theta} s^{2\theta+1}\Gamma\left[\theta+\frac{1}{2}\right]^2 \Gamma\left[\theta+1\right]\Gamma\left[2\theta+\frac{3}{2}\right]}{\pi\Gamma\left[3\theta+\frac{5}{2}\right]} \times \qquad (4.12) \\
&\qquad \times {}_3F_2\left(\left\{ \theta+\frac{1}{2},\theta+\frac{1}{2},\theta+1 \right\}, \left\{ \frac{1}{2}, 3\theta+\frac{5}{2} \right\}, 1 \right),
\end{aligned}
$$

where ${}_3F_2(\{a_1,a_2,a_3\};\{b_1,b_2\};z)$ is a Generalized Hypergeometric Function (see for example [1]). Finally, we have

$$\int_0^t \int_0^s \mathbb{E}\left[Y_s Y_{s'}\right] ds' ds =$$

$$= \frac{4^\theta \Gamma\left[\theta+\frac{1}{2}\right]^2 \Gamma\left[\theta+1\right] \Gamma\left[2\theta+\frac{3}{2}\right]}{\pi\Gamma\left[3\theta+\frac{5}{2}\right]} \times$$

$$\times\, {}_3F_2\left(\left\{\theta+\frac{1}{2}, \theta+\frac{1}{2}, \theta+1\right\}, \left\{\frac{1}{2}, 3\theta+\frac{5}{2}\right\}, 1\right) \int_0^t s^{2\theta+1} ds$$

$$= \frac{4^\theta \Gamma\left[\theta+\frac{1}{2}\right]^2 \Gamma\left[\theta+1\right] \Gamma\left[2\theta+\frac{3}{2}\right]}{\pi\Gamma\left[3\theta+\frac{5}{2}\right]} \times$$

$$\times\, {}_3F_2\left(\left\{\theta+\frac{1}{2}, \theta+\frac{1}{2}, \theta+1\right\}, \left\{\frac{1}{2}, 3\theta+\frac{5}{2}\right\}, 1\right) \frac{t^{2(\theta+1)}}{2(\theta+1)}.$$

Thus

$$\mathbb{E}\left[(X_t)^2\right] = 2\int_0^t \int_0^s \mathbb{E}\left[Y_s Y_{s'}\right] ds' ds$$

(4.13)

$$= \frac{4^\theta \Gamma\left[\theta+\frac{1}{2}\right]^2 \Gamma\left[\theta+1\right] \Gamma\left[2\theta+\frac{3}{2}\right]}{\pi\Gamma\left[3\theta+\frac{5}{2}\right](\theta+1)} {}_3F_2\left(\left\{\theta+\frac{1}{2}, \theta+\frac{1}{2}, \theta+1\right\}, \left\{\frac{1}{2}, 3\theta+\frac{5}{2}\right\}, 1\right) t^{2(\theta+1)}.$$

□

4.3.2 Non-identically distributed observations

Asymptotic Method of Moments

The first estimator we develop is based on observing n independent paths of X_t, where each path is observed once at different times, $\{X_{i,t_i}\}_{i=1}^n$, $t_i > 0$ for all i. Though the expected value of each observation is different (i.e. the observations are independent but not identically distributed), through the use of the Kolmogorov Law of Large numbers and Kronecker's lemma, we develop an estimator that is almost surely consistent. We recall the following Corollary:

Corollary IV.6 (Corollary 7.4.1 [50]). *Let $\{X_n, n \geq 1\}$ be an independent sequence of random variables satisfying $\mathbb{E}\left[X_n^2\right] < \infty$. Suppose we have a monotone sequence*

$b_n \uparrow \infty$. *If*

$$\sum_k Var\left(\frac{X_k}{b_k}\right) < \infty$$

then

$$\frac{S_n - \mathbb{E}\left[S_n\right]}{b_n} \xrightarrow{a.s.} 0,$$

where

$$S_n = \sum_{k=1}^{n} X_k.$$

In particular, we have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_{i,t_i} - \mathbb{E}\left[X_{t_i}\right]}{n} = 0 \ a.s. \tag{4.14}$$

Thus, using an approach similar to the method of moments, we have the following estimator of θ. We consider only situations where $t > 5$ and $\theta \geq 0.1$ to ensure a well defined estimator.

Definition IV.7. Assuming that we have n paths of X_t, each observed once at a different time t_i ($t_i > 5$), we define $\hat{\theta}$ as

$$\hat{\theta} = \arg_{\theta>0} \left\{ \frac{\sum_{i=1}^{n} X_{i,t_i} - \mathbb{E}\left[X_{t_i}\right]}{n} = 0 \right\}. \tag{4.15}$$

Properties of $\hat{\theta}$

We first establish establish properties for the asymptotic method of moments estimator $\hat{\theta}$. The first property, consistency, is direct based on the definition of the estimator.

Proposition IV.8. *Assuming $\hat{\theta}$ is defined as in Eq. (4.15), we have*

$$\hat{\theta} \xrightarrow{a.s.} \theta \quad \text{as } n \to \infty.$$

Due to the fact that we assume that the observations are not made at a fixed, common time, we are dealing with independent but not identically distributed random variables. The standard central limit theorem does not apply, and we must appeal to a more general result; namely Lyapunov's condition. We recall that if, for a collection of independent, mean-zero random variables $\{X_i\}_{i=1}^n$, we can show there exists a $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i|^{2+\delta}\right] = 0$$

where $s_n^2 = \sum_{i=1}^n \mathbb{E}\left[X_i^2\right]$, then

$$\frac{\sum_{i=1}^n X_i}{s_n} \xrightarrow{\mathfrak{D}} N(0,1).$$

The process does not easily allow for calculation of non-integer moments, and due to the convexity of the absolute value, we must use the fourth moment (i.e. $\delta = 2$) if we would like to appeal to this condition. While the computations are currently untenable, there certainly appears to be evidence of convergence based on numerical calculations for serval various values of the parameter. This leads us to believe that the variance of our estimator decreases as a linear function of n, and based on the computed variances, we expect this estimator to preform well even with few observations.

4.3.3 Identically distributed observations

Method of Moment estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$

Here, we will assume a slightly simpler estimation problem. We still suppose that we observe n independent paths of X_t where each path is observed once, but now we assume that each path is observed at the same time. From the law of large numbers, we have

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{i=1}^{n} X_{t,i} = \mathbb{E}\left[X_t\right],$$

$$\text{and} \qquad \lim_{n\to\infty} \frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_{t,i}\right)^2 = \mathbb{E}\left[(X_t)^2\right].$$

The simplest estimators based on n independent observations of X_t for a fixed and common $t > 0$ are method of moments estimators based on the first two moments. Directly from the above proposition, we can define these estimators. Again, we will require $t > 5$ and $\theta \geq 0.1$ to ensure the estimators are well defined.

Definition IV.9. Assume, for a fixed $t > 5$ and unknown $\theta \geq 0.1$, that $\{X_{t,i}\}_{i=1}^{n}$ is an independent collection of n observations of the process defined in Eq. (4.1). Define

- $\hat{\theta}_1$ as the value of θ that solves

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_{t,i}^{\theta} = \mathbb{E}\left[X_t\right],$$

- $\hat{\theta}_2$ as the value of θ that solves

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_{t,i}^{\theta}\right)^2 = \mathbb{E}\left[(X_t)^2\right].$$

We note that $\hat{\theta}$ defined by Eq. (4.15) is identical to $\hat{\theta}_1$ when the observations are all made at the same time.

Generalized Method of Moment Estimator, $\hat{\theta}_G$

Since we are interested in estimating the single parameter θ, and we have two moment conditions, θ is overidentified, and we can use the generalized method of moment (GMM) approach. Though around since the early 1950's, GMM began to be used more heavily in the early 1980's following Peter Hansen's work on the asymptotic properties of the GMM estimator [21]. The idea of the GMM is that since there are more moment equations than unknowns, no one value of the parameter (or more in more general cases, parameters) can satisfy all equations. In this situation, we have a vector valued function of an observation and the parameter θ, $f(X, \theta)$ such that $\mathbb{E}[f(X, \theta)] = 0$. We seek the value of θ that comes closest to satisfying the equations. Defining the concept of closeness however is where the critical step is made.

Definition IV.10 (**W**-norm)**.** For a positive semi-definite $k \times k$ matrix $\mathbf{W}$, define the **W**-norm $\|\cdot\|_{\mathbf{W}}$ as

$$\|a\|_{\mathbf{W}} = a'\mathbf{W}a$$

Defined in [38], we will use the following 2-step estimation procedure, where we again only observe the process for $t > 5$ and $\theta \geq 0.1$ to ensure the estimators are well defined:

Definition IV.11 (Two-stage GMM estimation)**.** First define a preliminary estimate $\hat{\theta}_p$ by choosing $\mathbf{W} = \mathbf{I}_k$:

$$\hat{\theta}_p = \underset{\theta \in \Theta}{argmin} \, \|\overline{f}_n(\theta)\|_{\mathbf{W}}$$

where

$$\overline{f}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f(X_i, \theta)$$

Second, this estimate of θ is used to approximate the ideal $\mathbf{W}$ with $\widehat{\mathbf{W}}^*$:

$$\widehat{\mathbf{W}}^* = \left(\frac{1}{n}\sum_{i=1}^{n} f(x_i, \theta) f(x_i, \theta)'\right)^{-1}$$

and third, the final estimate is obtained, $\hat{\theta}_G$ by minimizing the distance using $\widehat{\mathbf{W}}^*$:

$$\hat{\theta}_G = \underset{\theta \in \Theta}{argmin} \, \|\overline{f}_n(\theta)\|_{\widehat{\mathbf{W}}^*}$$

For our particular problem, the natural choice for f is:

$$f(X_t, \theta) = \begin{pmatrix} X_t - \mathbb{E}[X_t] \\ (X_t)^2 - \mathbb{E}\left[(X_t)^2\right] \end{pmatrix} \tag{4.16}$$

We use $\widehat{\mathbf{W}}^*$ since it approximates $\mathbf{W}^*$ defined as

$$\mathbf{W}^* = \mathbb{E}\left[f(X,\theta)f(X,\theta)'\right]^{-1}$$

As will be shown in the next section, were we able to use $\mathbf{W}^*$ as our $\mathbf{W}$, the estimators asymptotic variance would be minimized. Since $\mathbf{W}^*$ depends on the parameter we are trying to estimate however, we must use a consistent estimator of it instead.

Properties of $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_G$

The method of moment estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, can be rewritten as GMM estimators who are the results of a single moment condition. While these two estimators are not the result of an over-identified parameter, the results derived for the GMM estimator, $\hat{\theta}_G$ hold for these estimators as well (though in some cases, the proof is considerably more direct when they are considered as traditional method of moment estimators). For our GMM estimator, while our choice of $\mathbf{W}$ is not the most efficient one, we still have several desirable asymptotic properties of $\hat{\theta}_G$.

Theorem IV.12 (Consistency)**.** *[Theorem 1.1 p.13, [38]] For a fixed $t > 0$,*

$$\hat{\theta}_G - \theta \xrightarrow{\mathbf{P}} 0 \qquad as\ n \to \infty$$

The consistency of the estimate does not actually depend on the choice of the weighting matrix $\mathbf{W}$. In fact, $\hat{\theta}_p$ from Definition IV.11 is asymptotically consistent.

The choice of $\mathbf{W}$ does however effect the asymptotic variance of the estimate. We first recall the following central limit theorem from Reference [38].

Theorem IV.13 (Asymptotic Normality). *[Theorem 1.2 p.19, [38]] Let* $\overline{F}_n(\theta) = \frac{d\bar{f}_n}{d\theta}(\theta)$*. Assuming there exists a vector* F *such that*

$$\overline{F}_n(\theta) \xrightarrow{\mathbf{P}} F \quad as\ n \to \infty$$

and defining $\mathbf{V} = (\mathbf{W}^*)^{-1}$*, we have, for a fixed* $t > 0$ *and any choice of weighting matrix* $\mathbf{W}$*,*

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \sigma^2)$$

where σ^2 *is defined as:*

$$(F'\mathbf{W}F)^{-1}F'\mathbf{W}\mathbf{V}\mathbf{W}F(F'\mathbf{W}F)^{-1}$$

As noted in [38], it is clear that choosing $\mathbf{W} = \mathbf{W}^* = \mathbf{V}^{-1}$ would minimize the asymptotic variance. However, as noted in the definition of the GMM estimator $\hat{\theta}$, we do not know θ. Since $\hat{\theta}_p$ is a consistent estimator of θ, $\hat{\mathbf{W}}^*$ is a consistent estimator of $\mathbf{W}^*$, and thus $\hat{\theta}_G$ is asymptotically efficient. It must be noted that since any initial choice of $\mathbf{W}$ in step 1 of Definition IV.11 leads to an efficient (but different) estimator $\hat{\theta}_G$, $\hat{\theta}_G$ is really an element of the class of all asymptotically efficient estimators. This class consists of a unique estimator for every initial choice of positive semi-definite 2×2 weighting matrix $\mathbf{W}$.

4.4 Simulation Studies on θ

4.4.1 Non-identically distributed observations

Due to the fact that this estimation problem does not have a current standard method, there is no estimator to compare to our estimator $\hat{\theta}$. For these simulation studies several parameter values were considered. In each case, 100 simulations were conducted, with each simulation considering estimating θ based on $n = 20, 4060$ or 80 observations. The values of θ considered were $\theta = 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75$ and 2. For each value of θ, due to the complexity of the process, the variance of the estimator is intractable, so parametric bootstrap standard errors are provided. We

	$\hat{\theta}$ (s.e.($\hat{\theta}$))			
θ	$n = 20$	$n = 40$	$n = 60$	$n = 80$
0.25	0.2328 (0.1010)	0.2280 (0.0865)	0.2456 (0.0714)	0.2356 (0.0617)
0.50	0.4782 (0.1256)	0.4958 (0.0896)	0.4868 (0.0690)	0.4858 (0.0619)
0.75	0.7350 (0.1333)	0.7514 (0.0890)	0.7490 (0.0794)	0.7365 (0.0636)
1.00	0.9520 (0.1484)	0.9905 (0.1039)	0.9980 (0.0875)	0.9897 (0.0838)
1.25	1.2100 (0.1570)	1.2168 (0.1238)	1.2283 (0.0990)	1.2336 (0.0829)
1.50	1.4483 (0.2000)	1.4704 (0.1450)	1.4753 (0.1134)	1.4853 (0.0939)
1.75	1.6912 (0.2094)	1.7115 (0.1343)	1.7267 (0.1477)	1.7352 (0.1033)
2.00	1.9134 (0.2576)	1.9444 (0.1901)	1.9666 (0.1575)	1.9710 (0.1381)

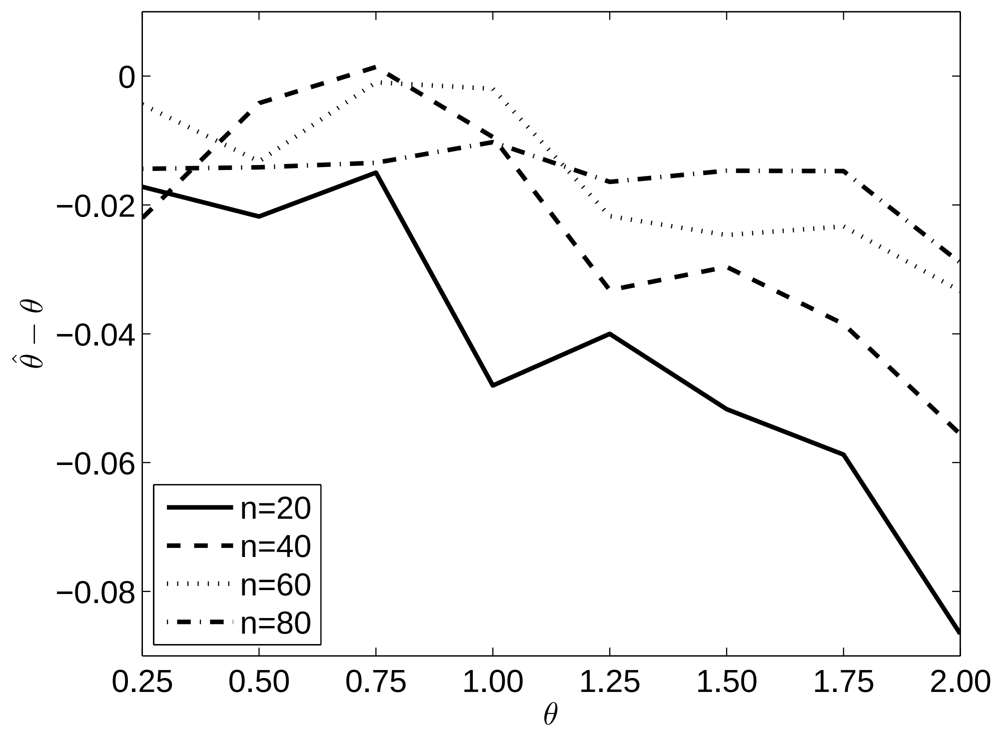
Table 4.1: Estimated values and standard errors for dependent observations

can see from Table 4.1 and Figure 4.2 that the bias as well as the standard error increases as θ increases for a fixed value of n. Additionally, and not surprisingly, from Table 4.1 and Figure 4.3 we see that, for a fixed θ, the bias and the standard error of $\hat{\theta}$ decreases as n increases. This is not surprising since we know that the estimator is both consistent and asymptotically Normal, thus the standard error must decreases as a function of n.

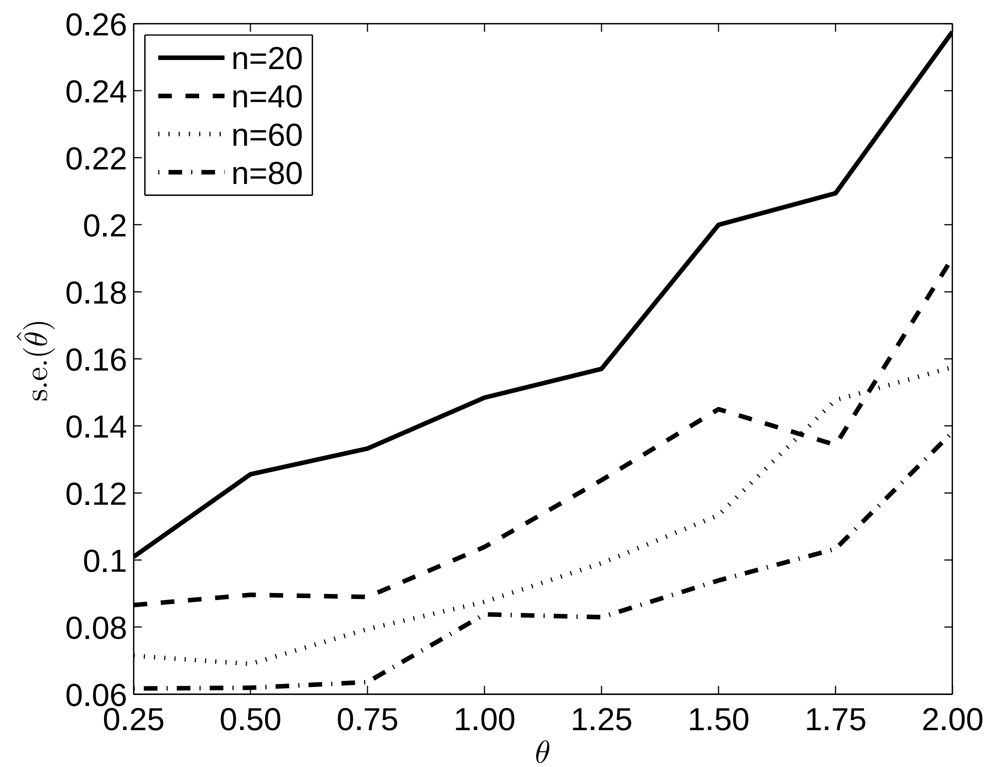
4.4.2 Identically distributed observations

For each of the three estimators in this simplified setup, 100 simulations were run with on four different levels of the sample size ($n = 10$, $n = 50$, $n = 100$ and $n = 500$) and twenty levels of θ ($\theta = 0.1, 0.2, \ldots, 2$). We again can not compute the variance of the estimator so we report the parametric bootstrap standard errors. For the GMM estimator, the preliminary estimate ($\hat{\theta}_p$) was also recorded. The mean estimate as well as sample standard deviation for each case are shown in the tables below.

The simulations indicate that depending on the circumstances, different estima-

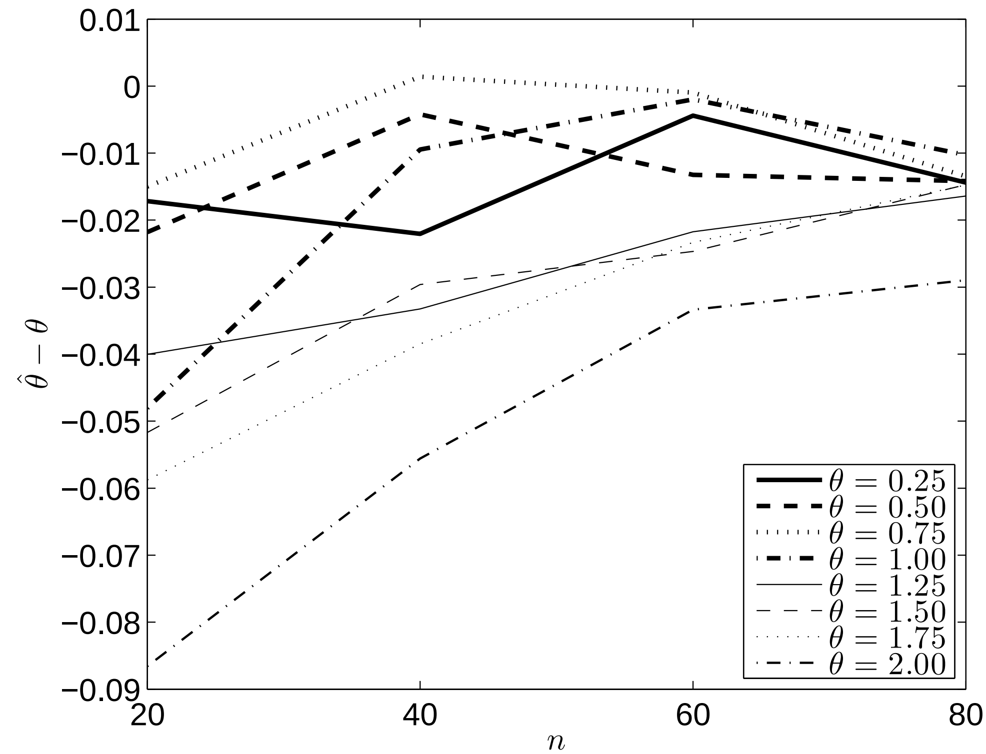


(a) Difference between $\hat{\theta}$ and θ

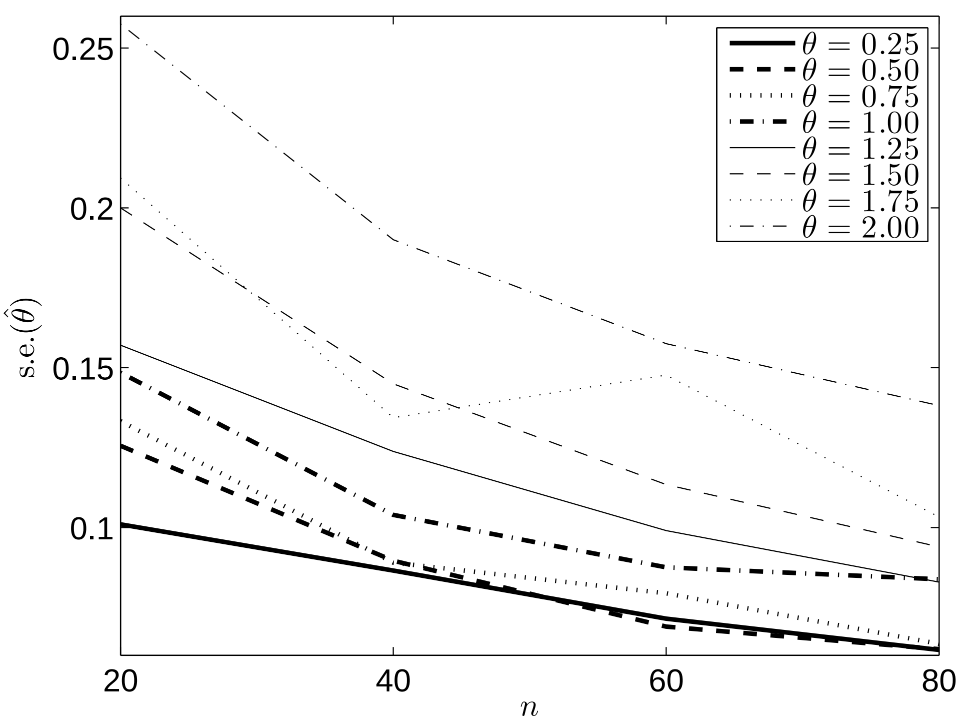


(b) Standard Error of $\hat{\theta}$

Figure 4.2: Simulation results with respect to θ



(a) Difference between $\hat{\theta}$ and θ



(b) Standard Error of $\hat{\theta}$

Figure 4.3: Simulation results with respect to n

θ	$\hat{\theta}_1$		$\hat{\theta}_1$		$\hat{\theta}_p$		$\hat{\theta}_G$	
0.1	0.104	(0.022)	0.105	(0.024)	0.104	(0.022)	0.102	(0.014)
0.2	0.207	(0.059)	0.224	(0.097)	0.204	(0.068)	0.196	(0.041)
0.3	0.323	(0.132)	0.333	(0.148)	0.329	(0.120)	0.304	(0.061)
0.4	0.455	(0.221)	0.689	(0.160)	0.400	(0.149)	0.384	(0.073)
0.5	0.551	(0.236)	0.675	(0.172)	0.656	(0.207)	0.458	(0.099)
0.6	0.598	(0.227)	0.675	(0.201)	0.725	(0.205)	0.531	(0.145)
0.7	0.987	(0.182)	0.714	(0.230)	0.763	(0.218)	0.628	(0.168)
0.8	0.985	(0.210)	0.713	(0.262)	0.810	(0.257)	0.682	(0.197)
0.9	1.055	(0.261)	0.813	(0.313)	0.848	(0.282)	0.751	(0.221)
1.0	1.103	(0.297)	0.855	(0.338)	0.903	(0.305)	0.810	(0.240)
1.1	1.091	(0.331)	0.931	(0.342)	0.953	(0.349)	0.917	(0.276)
1.2	1.186	(0.373)	1.065	(0.345)	1.002	(0.317)	0.945	(0.242)
1.3	1.186	(0.426)	1.092	(0.345)	1.110	(0.331)	1.044	(0.252)
1.4	1.321	(0.415)	1.225	(0.349)	1.201	(0.329)	1.138	(0.261)
1.5	1.410	(0.408)	1.320	(0.341)	1.265	(0.304)	1.194	(0.215)
1.6	1.493	(0.385)	1.372	(0.307)	1.394	(0.293)	1.293	(0.234)
1.7	1.529	(0.394)	1.430	(0.308)	1.461	(0.321)	1.379	(0.241)
1.8	1.644	(0.379)	1.518	(0.290)	1.624	(0.340)	1.516	(0.254)
1.9	1.763	(0.379)	1.651	(0.312)	1.614	(0.278)	1.523	(0.197)
2.0	1.804	(0.376)	1.676	(0.276)	1.698	(0.305)	1.623	(0.220)

Table 4.2: Average moment estimates and (standard errors) for $n = 10$

θ	$\hat{\theta}_1$		$\hat{\theta}_1$		$\hat{\theta}_p$		$\hat{\theta}_G$	
0.1	0.100	(0.008)	0.100	(0.009)	0.100	(0.011)	0.099	(0.006)
0.2	0.201	(0.025)	0.202	(0.032)	0.199	(0.025)	0.198	(0.015)
0.3	0.303	(0.042)	0.315	(0.072)	0.302	(0.060)	0.298	(0.023)
0.4	0.409	(0.081)	0.660	(0.121)	0.394	(0.087)	0.394	(0.031)
0.5	0.530	(0.170)	0.652	(0.143)	0.564	(0.138)	0.493	(0.047)
0.6	0.614	(0.169)	0.664	(0.155)	0.680	(0.147)	0.575	(0.067)
0.7	0.938	(0.159)	0.654	(0.170)	0.741	(0.138)	0.673	(0.095)
0.8	0.999	(0.196)	0.769	(0.199)	0.783	(0.159)	0.758	(0.090)
0.9	1.044	(0.230)	0.881	(0.240)	0.847	(0.187)	0.850	(0.130)
1.0	1.064	(0.247)	0.939	(0.245)	0.954	(0.209)	0.948	(0.143)
1.1	1.120	(0.285)	0.996	(0.267)	1.033	(0.241)	1.043	(0.159)
1.2	1.103	(0.279)	1.104	(0.231)	1.156	(0.222)	1.151	(0.184)
1.3	1.287	(0.306)	1.241	(0.219)	1.221	(0.234)	1.212	(0.199)
1.4	1.308	(0.327)	1.292	(0.243)	1.300	(0.220)	1.272	(0.185)
1.5	1.462	(0.308)	1.401	(0.244)	1.373	(0.240)	1.358	(0.209)
1.6	1.574	(0.280)	1.532	(0.236)	1.479	(0.231)	1.447	(0.194)
1.7	1.672	(0.266)	1.601	(0.218)	1.597	(0.248)	1.517	(0.223)
1.8	1.717	(0.301)	1.668	(0.256)	1.653	(0.249)	1.606	(0.227)
1.9	1.828	(0.314)	1.783	(0.283)	1.757	(0.285)	1.691	(0.257)
2.0	1.844	(0.314)	1.777	(0.253)	1.826	(0.266)	1.759	(0.233)

Table 4.3: Average moment estimates and (standard errors) for $n = 50$

θ	$\hat{\theta}_1$		$\hat{\theta}_1$		$\hat{\theta}_p$		$\hat{\theta}_G$	
0.1	0.099	(0.007)	0.099	(0.008)	0.099	(0.008)	0.100	(0.004)
0.2	0.198	(0.016)	0.198	(0.021)	0.203	(0.018)	0.198	(0.011)
0.3	0.301	(0.036)	0.314	(0.074)	0.297	(0.041)	0.297	(0.017)
0.4	0.403	(0.058)	0.653	(0.109)	0.404	(0.073)	0.394	(0.023)
0.5	0.524	(0.120)	0.625	(0.113)	0.552	(0.138)	0.493	(0.036)
0.6	0.618	(0.151)	0.635	(0.130)	0.642	(0.111)	0.589	(0.048)
0.7	0.975	(0.161)	0.693	(0.169)	0.740	(0.127)	0.686	(0.057)
0.8	0.973	(0.179)	0.767	(0.168)	0.829	(0.141)	0.782	(0.077)
0.9	0.965	(0.190)	0.818	(0.198)	0.858	(0.152)	0.882	(0.096)
1.0	0.974	(0.197)	0.891	(0.208)	0.990	(0.168)	0.995	(0.107)
1.1	1.093	(0.236)	1.066	(0.175)	1.061	(0.160)	1.075	(0.119)
1.2	1.171	(0.270)	1.161	(0.178)	1.152	(0.162)	1.157	(0.123)
1.3	1.266	(0.282)	1.273	(0.180)	1.247	(0.212)	1.247	(0.159)
1.4	1.357	(0.279)	1.352	(0.198)	1.360	(0.203)	1.347	(0.173)
1.5	1.466	(0.284)	1.434	(0.190)	1.433	(0.167)	1.416	(0.164)
1.6	1.513	(0.281)	1.515	(0.211)	1.536	(0.201)	1.517	(0.196)
1.7	1.624	(0.270)	1.624	(0.246)	1.621	(0.194)	1.588	(0.200)
1.8	1.725	(0.250)	1.694	(0.206)	1.719	(0.207)	1.676	(0.202)
1.9	1.816	(0.237)	1.777	(0.205)	1.765	(0.228)	1.699	(0.213)
2.0	1.937	(0.267)	1.880	(0.223)	1.888	(0.208)	1.829	(0.209)

Table 4.4: Average moment estimates and (standard errors) for $n = 100$

θ	$\hat{\theta}_1$		$\hat{\theta}_1$		$\hat{\theta}_p$		$\hat{\theta}_G$	
0.1	0.100	(0.003)	0.100	(0.004)	0.100	(0.003)	0.100	(0.002)
0.2	0.200	(0.007)	0.200	(0.009)	0.200	(0.009)	0.200	(0.004)
0.3	0.299	(0.014)	0.299	(0.021)	0.299	(0.019)	0.300	(0.008)
0.4	0.398	(0.026)	0.660	(0.070)	0.405	(0.032)	0.399	(0.010)
0.5	0.504	(0.046)	0.591	(0.078)	0.497	(0.053)	0.500	(0.015)
0.6	0.621	(0.086)	0.599	(0.084)	0.619	(0.058)	0.600	(0.018)
0.7	0.940	(0.116)	0.673	(0.108)	0.709	(0.069)	0.697	(0.027)
0.8	0.923	(0.123)	0.774	(0.093)	0.792	(0.083)	0.798	(0.037)
0.9	0.935	(0.138)	0.877	(0.118)	0.891	(0.079)	0.889	(0.041)
1.0	0.992	(0.163)	0.996	(0.087)	0.993	(0.080)	0.991	(0.049)
1.1	1.082	(0.147)	1.094	(0.074)	1.091	(0.078)	1.084	(0.057)
1.2	1.171	(0.169)	1.199	(0.088)	1.184	(0.074)	1.181	(0.059)
1.3	1.295	(0.137)	1.293	(0.079)	1.289	(0.079)	1.285	(0.070)
1.4	1.378	(0.127)	1.381	(0.084)	1.392	(0.090)	1.388	(0.084)
1.5	1.491	(0.123)	1.494	(0.102)	1.491	(0.095)	1.478	(0.100)
1.6	1.577	(0.118)	1.569	(0.092)	1.573	(0.083)	1.555	(0.086)
1.7	1.690	(0.107)	1.679	(0.093)	1.678	(0.109)	1.645	(0.108)
1.8	1.792	(0.107)	1.781	(0.110)	1.781	(0.114)	1.750	(0.114)
1.9	1.880	(0.123)	1.862	(0.116)	1.862	(0.115)	1.827	(0.120)
2.0	1.991	(0.111)	1.977	(0.111)	1.985	(0.117)	1.942	(0.126)

Table 4.5: Average moment estimates and (standard errors) for $n = 500$

tors are preferable. As in the more general case, each estimator becomes more accurate and more precise as n increases, as well as for lower values of θ. For the largest sample size simulations (i.e. $n = 500$) all estimators appear to preform equally well. Interestingly, the bias effect is considerably greater in this simplified situation than in the general case. When each path is observed at a different time, a more accurate estimate can be made using significantly less observations.

4.5 Estimation of T_D

As stated before, the primary value of interest is the first time the process X_t passes some fixed threshold. We can set the threshold to represent the level at which some object fails, and then T_D represents the failure time. While this stopping time is a random variable that is a function of θ, the analytic computation of the density of T_D is even more involved than that of $\hat{\theta}$, so again we cannot attempt to find a maximum likelihood estimator even once we have a consistent estimator of θ. However, we can still rely on the strong law of large numbers to develop an estimator of T_D.

If, for a fixed $D > 0$, we consider n independent paths of X_t, since the resulting stopped times $\{T_{D,i}\}_{i=1}^{n}$ are independent and identically distributed draws of the non-negative random variable T_D, we have the following law of large numbers result

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} T_{D,i} = \mathbb{E}\left[T_D\right] \quad \text{a.s.} \tag{4.17}$$

Thus using the strong law of numbers yet again to drive the estimation, we define $\hat{T}_D$ as follows:

Definition IV.14. Once we have obtained a consistent estimator for θ, for a fixed D, we simulate n independent paths of X_t. For each of these processes, we calculate the first time the path crosses the threshold D, $T_{D,i}$ for $i = 1, \ldots, n$. We then define $\hat{T}_D$ as:

$$\frac{1}{n}\sum_{i=1}^{n} T_{D,i} = \hat{T}_D \tag{4.18}$$

As with the asymptotic method of moments estimator of θ, $\hat{\theta}$, by definition of the estimator we directly have almost sure consistency of our estimator.

Since our estimator is the average of n i.i.d. observations of the same random variable, T_D, we can also directly appeal to the central limit theorem. In fact, we have the following result.

Proposition IV.15. *For a fixed $D > 0$, we consider n observations of the random variable T_D. With $\sigma_{T_D}^2 = Var(T_D)$, we have*

$$\sqrt{n}\left(\hat{T}_D - \mathbb{E}\left[T_D\right]\right) \sim N\left(0, \sigma_{T_D}^2\right)$$

Of course there are two practical concerns for our estimator. First, we are estimating θ with $\hat{\theta}$, but since our estimate is almost surely consistent, this presents no problems to the asymptotic Normality of the estimator. Second, we are also estimating σ_{T_D} with the sample standard deviation, but again since n is limited only by computational power and the sample standard deviation is an unbiased consistent estimator of the population standard deviation, our asymptotic central limit theorem result holds.

A second quantity concerning T_D that we are interested in is a prediction interval of T_D. In particular, since there is not only variation in T_D for a fixed θ, but also variation in $\hat{\theta}$, we must incorporate both variations in prediction. To this end, once an estimate of θ is calculated, $\hat{\theta}$, we then simulate data using $\hat{\theta}$ as the true θ many times. We obtain a collection of new estimates $\{\theta_i^*\}$. From each of these estimates we again simulate data, this time estimating T_D. Combining the two estimation problems allows us to combine the variation in both problems to obtain a prediction interval for T_D.

4.6 Simulation Studies on T_D

For this simulation study, we imagine that we have already calculated an estimate of θ. We then simulate n paths of the process X_t and compute our estimate $\hat{T}_D$. Since our method is based on simulations we do not have to limit ourselves to supposing that we have a small number of observed stopped times, and thus the only limit in the number of paths is computational. For the purposes of comparison, we obtain an estimate of T_D for $D = 5$ and $D = 10$, $\theta = 0.5, 1, 1.5$ and 2 for $n = 100$ and $n = 200$. We do this simulation 100 times and report the average estimate as well as the parametric bootstrap standard error. Again, since by using this method of estimation, we not only arrive at an estimate of $\mathbb{E}[T_D]$, but also its distribution, we present the estimates to the distribution as histograms below. For the sample distribution of T_D as well as the sample standard deviation computation, we use 2000 sample paths to compute the histogram of stopped times. Using the sample distribution of the mean stopped time, we construct bootstrap prediction intervals by identifying the values for which 2.5% and 97.5% of the observations fall below. For all the simulations we run the paths to $t = 20$. While this is long enough to ensure over 99% of the processes stop, since there is, for any t, a nonzero probability that $T_D > t$, no matter what value we chose to stop the processes, there will always be a chance that a few do not stop.

For the prediction intervals, we see that when both forms of variation (variation in estimating θ as well as variation in T_D) are combined, the resulting intervals are quite wide. However, the same trend that we observe when predicting $\hat{T}_D$ exists here as well; namely that the interval decreases as θ increases.

	ExT_D (s(T_D))	
	$D = 5$	$D = 10$
θ	$n = 2000$	$n = 2000$
0.5	5.213146 (1.848930)	8.393804 (3.090153)
1.0	4.729977 (2.482036)	6.498354 (3.382244)
1.5	4.195491 (2.577465)	5.456557 (3.135371)
2.0	3.780593 (2.432531)	4.627412 (3.053394)

Table 4.6: Sample means of T_D and sample standard deviations

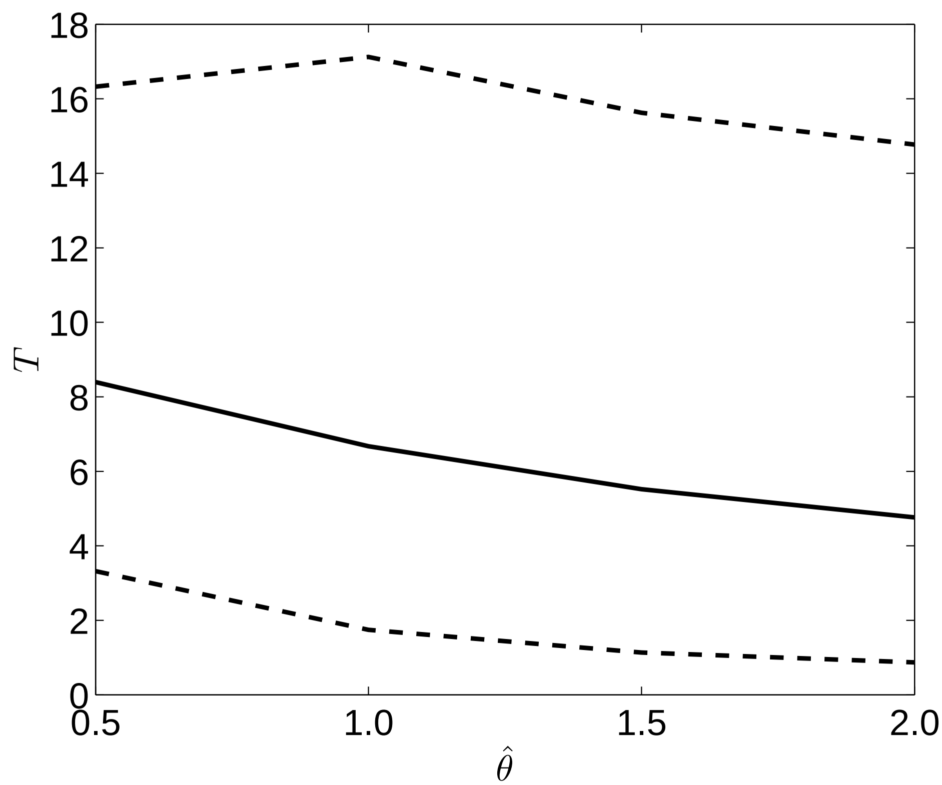


Figure 4.4: Expected value of T_D with 95% prediction intervals, $D = 10$

4.7 Conclusion

There were two main difficulties to overcome when considering estimating θ based on finite observations of paths. The first difficultly, the intractability of the maximum likelihood estimate was able to be solved using moment estimators. Though the

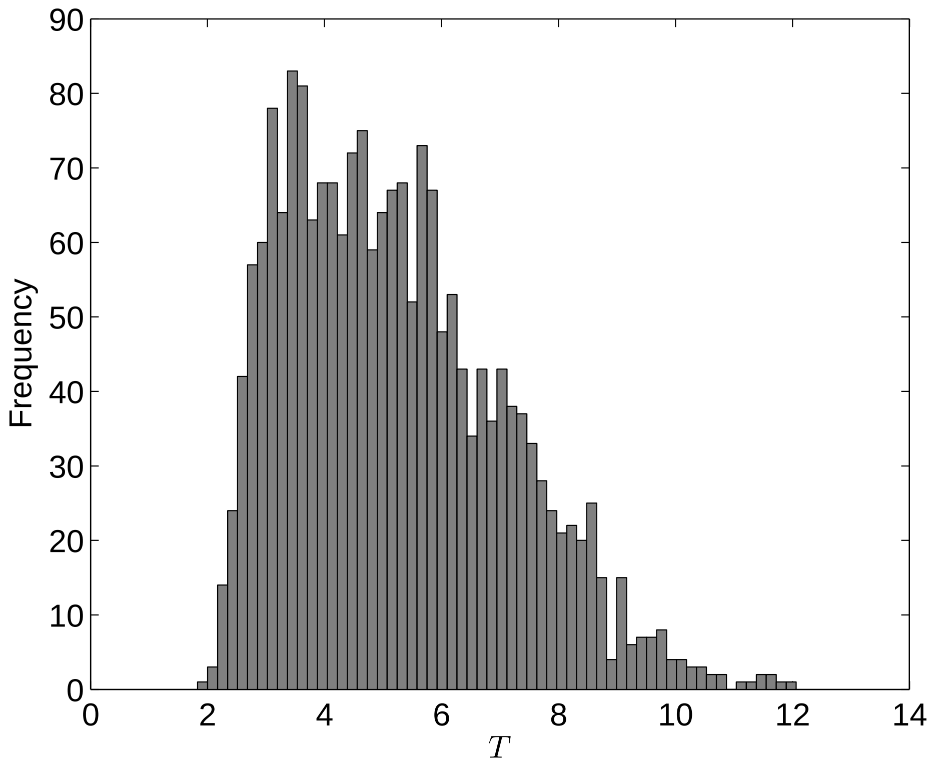


Figure 4.5: Sample distribution of T_D, $D = 5$, $\theta = 0.5$

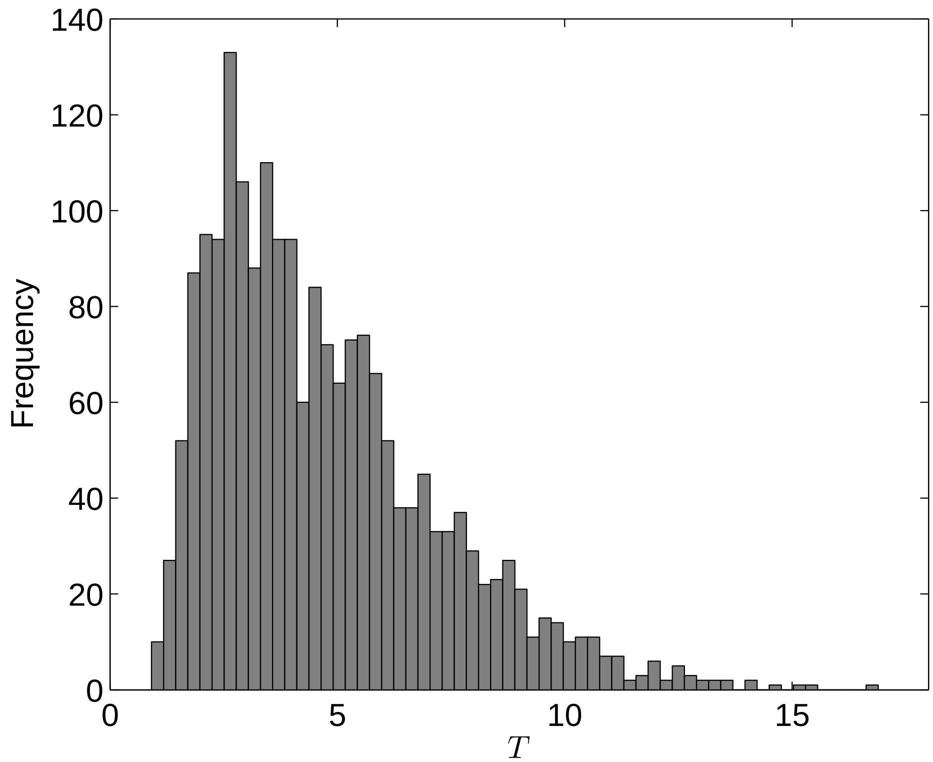


Figure 4.6: Sample distribution of T_D, $D = 5$, $\theta = 1.0$

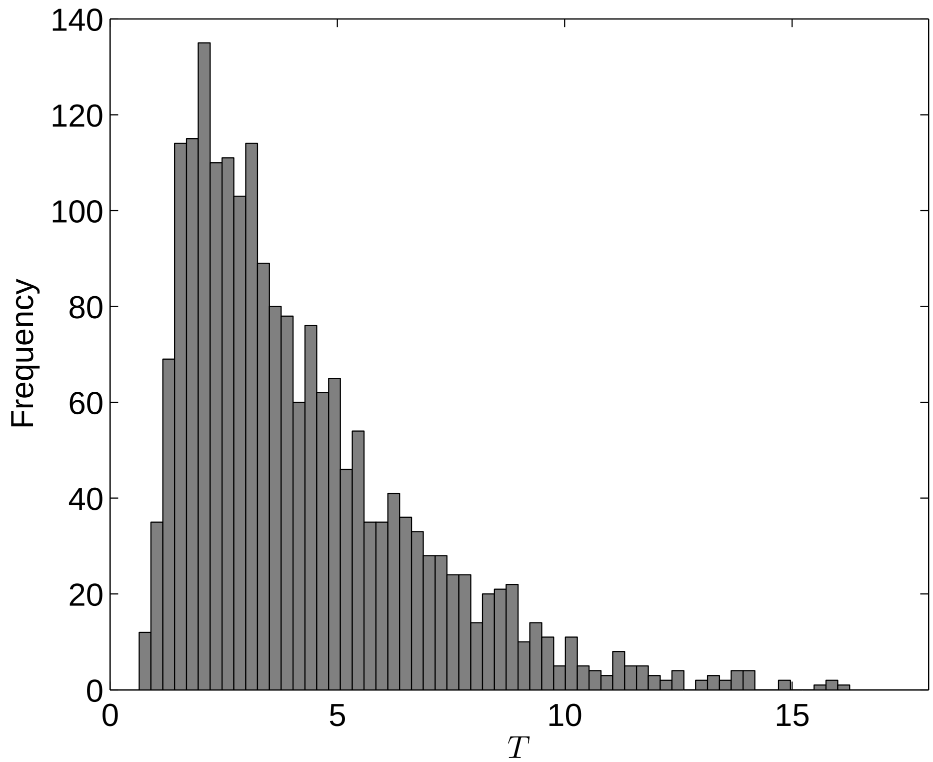


Figure 4.7: Sample distribution of T_D, $D = 5$, $\theta = 1.5$

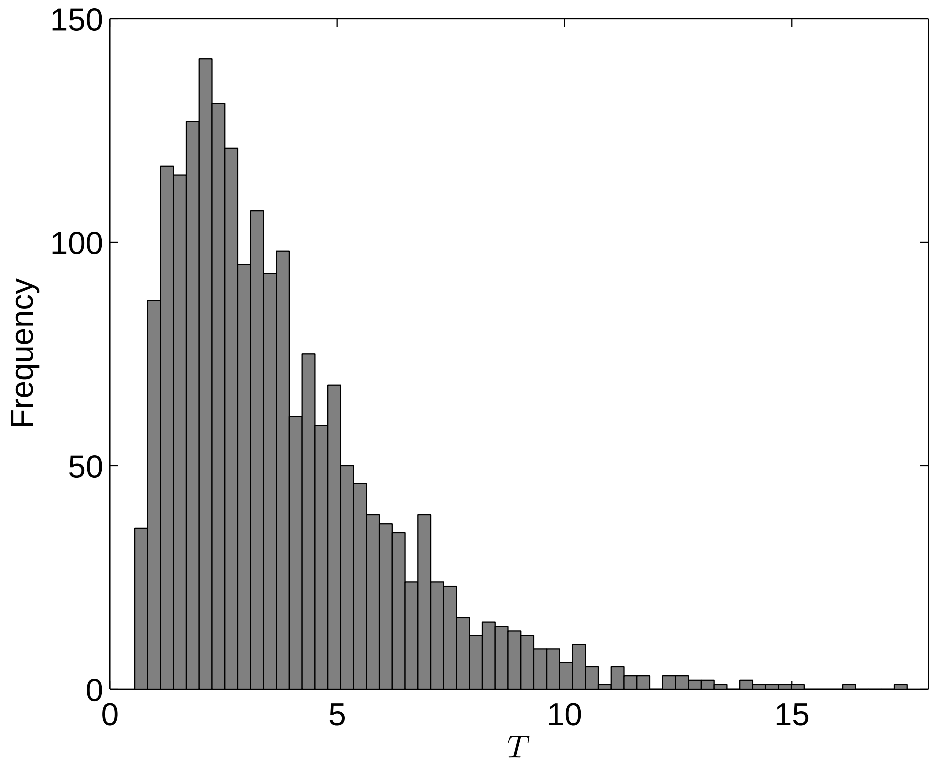


Figure 4.8: Sample distribution of T_D, $D = 5$, $\theta = 2.0$

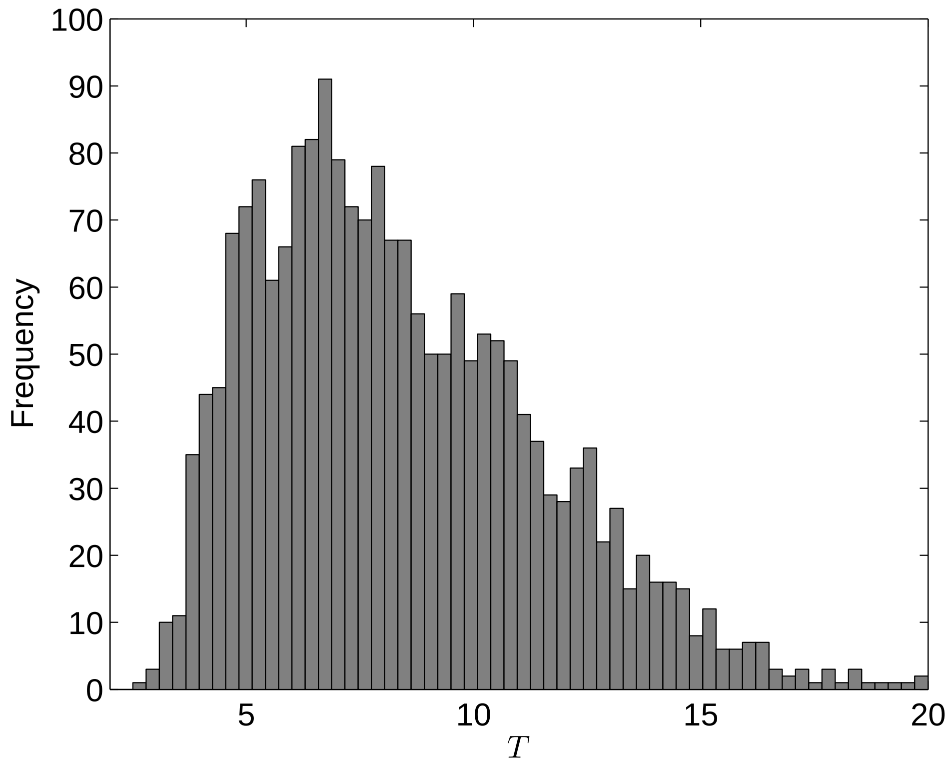


Figure 4.9: Sample distribution of T_D, $D = 10$, $\theta = 0.5$

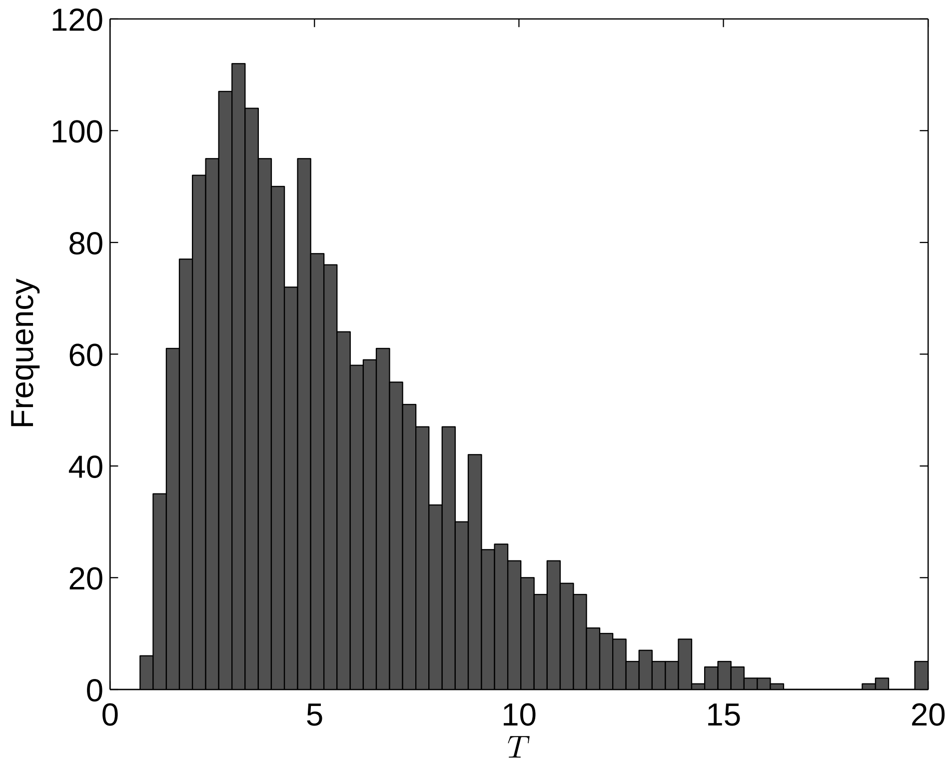


Figure 4.10: Sample distribution of T_D, $D = 10$, $\theta = 1.0$

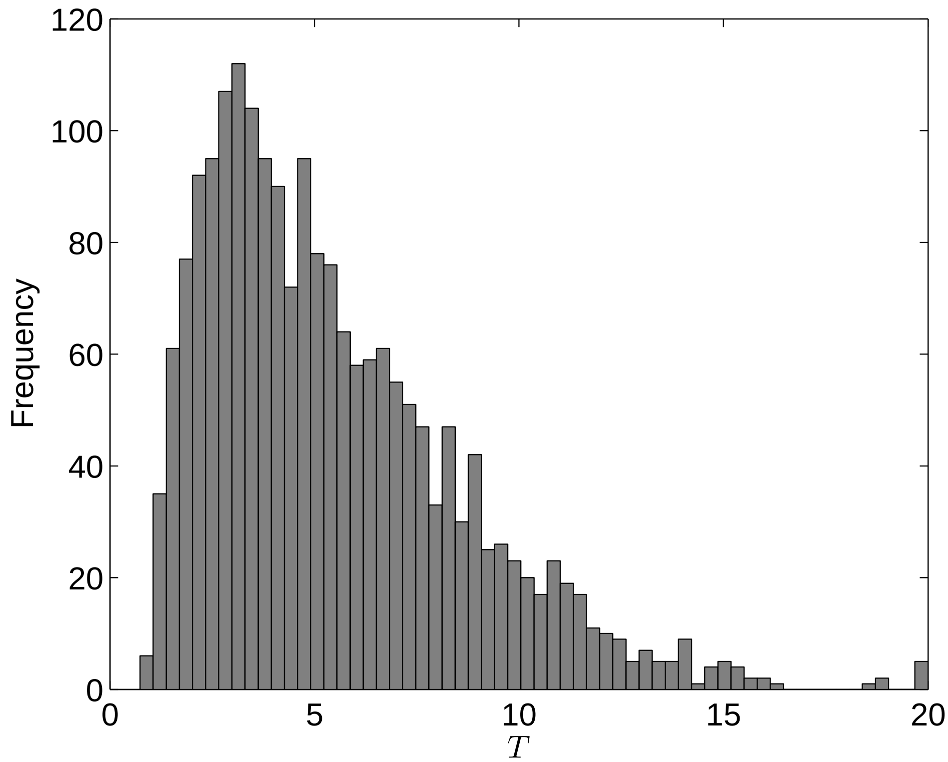


Figure 4.11: Sample distribution of T_D, $D = 10$, $\theta = 1.5$

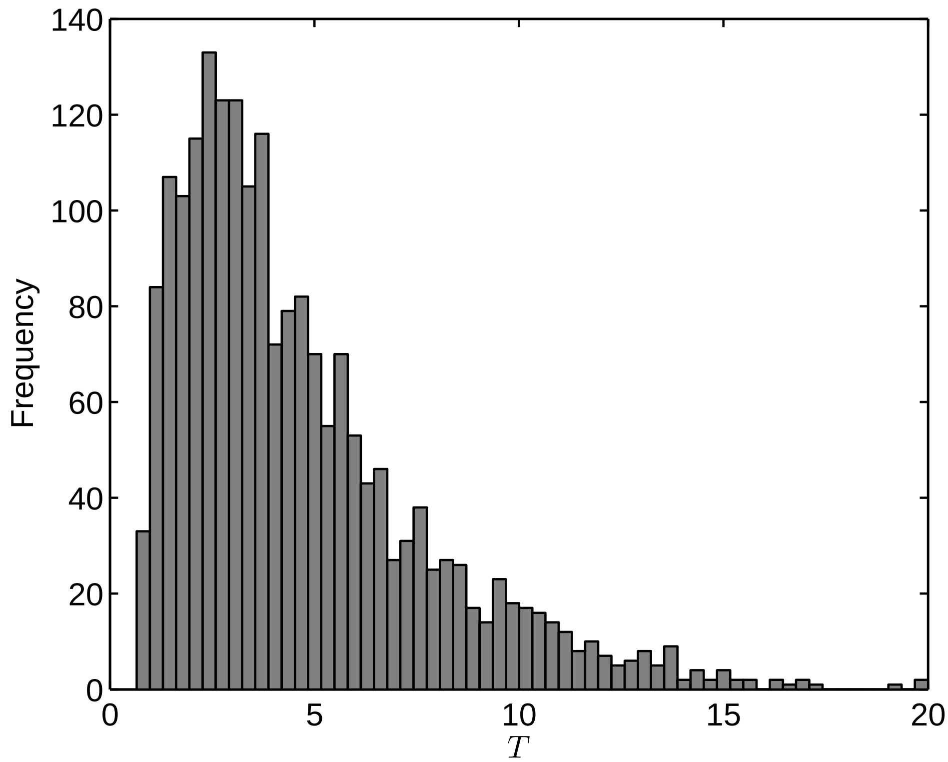


Figure 4.12: Sample distribution of T_D, $D = 10$, $\theta = 2.0$

process is complex, the moments are relatively easy to compute (when compared to attempts to compute the distribution of the process) and thus this opens up various moment methods. The general case where the observations are made on independent paths at different times interestingly produces a better estimate. Were the data to have been collected at the same time, and we were then forced into the identically observed estimation situation, we would have to consider what value we thought θ would be to decide which estimate we should use. From the above tables, there are some clear parameter regimes where the GMM estimator performs significantly better than the simpler method of moment estimators. When $\theta \leq 0.5$ (which corresponds to situations where the original Brownian motion is not being inflated by θ), the GMM estimator has a standard error about half of all the other estimates. Though this difference becomes less important as n increases, it would make a large difference when the number of observations is moderate. Conversely, for large values of θ, the method of moments estimator based on the first moment ($\hat{\theta}_1$) has the smallest bias. While for small n, the GMM estimator still has a smaller standard error, the values are more comparable, and in light of the bias, $\hat{\theta}_1$ would be the recommended estimator. Computationally, $\hat{\theta}_1$ is the easiest to compute, though not significantly more so than $\hat{\theta}_2$ or $\hat{\theta}_p$. Due to the generally good performance of $\hat{\theta}_2$ and $\hat{\theta}_p$, if no a priori range for θ can be established, both $\hat{\theta}_2$ and $\hat{\theta}_p$ can be reliably used for any number of paths and value of θ.

The second difficulty that is presented by this estimation problem is that we can not take advantage of multiple observations on the same path. Some applications lend themselves to more easily providing many observations from few paths. However, the non-Markovian property of the process causes any estimation procedure that is based on the moments to fail. Since all of the above estimates require the observations to be independent, and the definition of the process causes there to be strong dependence of observations on the same path, this third setup can not be solved using moments without knowing something on the covariance between two points on the same path.

CHAPTER V

Summary

Using completely different techniques we have derived estimators for two distinct types of processes that are solutions to stochastic evolution equations driven by Gaussian processes. In both cases, the observed processes were non-Markovian, and in both cases analytical calculation of the distribution of the process was intractable.

Concerning the stochastic differential equations driven by Volterra processes, the first step needed for inference was establishing the stochastic calculus as well as ensuring the existence and uniqueness of solutions to the differential equations. The primary tool used was a Girsanov-type transformation that allowed utilization of the associated martingale to the Volterra process to compute the maximum likelihood estimator.

The generality of Volterra processes allows for much more exotic dependence structures than that of fractional Brownian motion, while these processes still share enough in common so that no matter which Volterra process is driving the stochastic differential equation, we can still define an estimator that has several desirable properties. In the one parameter case, we have not only consistency, but a central limit theorem and a law of iterated logarithm associated with our estimate. In the multi-parameter case, while we can not establish as many properties due to the limitations of the multi-parameter martingale theory, we are still able to establish a type of consistency for our estimator.

The estimators defined in both Chapter 2 and Chapter 3 depend on continuous observation of the process; a practically impossible assumption. In the one parameter case, steps are taken toward the definition of a discretized estimator based on a finite number of observations, but unfortunately a completely discretized estimator can only be established on a case by case basis (i.e. the form of the Volterra process kernel must be known at least asymptotically to ensure convergence of the fully discretized estimator).

Another unrealistic assumption in practice is that we know the form of both the drift function as well as the two kernels associated with the Volterra process. Unfortunately neither of these problems is yet easy to address. As with the study of fractional Brownian motion, the establishment of an estimator is preceding a method to estimate the specific Gaussian noise, and as such we must still assume we have a specific noise if we wish to conduct any inference.

There are several directions that future work will hopefully take. First, the derivation of the associated kernels for a Volterra process other than Brownian motion and fractional Brownian motion is ongoing. The difficulty is finding the kernel from Theorem II.6, which requires solving a complex integral equation. A second direction of future research is concerning the estimation of a volatility parameter. In the inference of Chapter 2 and Chapter 3, the volatility was assumed to be a know function, and in some cases a known constant. Of particular interest would be estimating the volatility function if it is assumed to be a nonconstant function but follows some known form.

For the problem concerning the functional of Brownian motion and the estimation of the failure time, the analytic form of the distribution of T_D is not yet known. However, since we can directly compute the first two moments of the processes, we can establish several different estimators (depending on the setup of the observations) that allow us to then compute a bootstrap estimate of both the expected time to failure as well as a prediction interval associated with it.

As in the previous chapters, we were able to establish properties for our estimators. In all cases we show that our estimates are consistent, and for the estimators of θ, we either prove a central limit theorem or indicate the direction needed to go to prove the asymptotic result. Our primary estimate of θ, $\hat{\theta}$, performs very well even when we only have 20 observations, and in the simplified observation setup, we introduce and analyze the generalized method of moment estimation as a way of getting as much information as possible out of our data using both moment conditions.

There are several observational setups that are realistic but for which we can not develop an estimator. Since it may be practically easier to observe a few devices several times, we would like to be able to estimate θ in a case were we observe few paths of the process, but we observe each path several times. Additionally, we would like to be able to directly estimate the mean of T_D based on the observations instead of having to use bootstrap methods.

As before, there are several directions future work can take for this problem. Of obvious interest is finding any method for estimating the density of the process, and the density of T_D, other than the currently available one of simulation. Were we to have many more moments analytically, we could construct an approximate density as a weighted sum of the moments, but due to the increasing complexity of computing these moments, this seems unrealistic. Since the pair $(X_t^\theta, \mathbb{B}_t)$ is a Markov process in two dimensions, we could explore the methods used in the work by Lachel [32] and try to apply them to our problem. A second important direction of study is to understand the covariance between two points on a single path so that we can again consider the case where we observe few paths many times each.

For both projects, there are both theoretical and practical considerations that must be balanced. While a purely theoretical estimator can be developed in some cases, applying this estimator to real-world data situations forces us to weigh complexity of the estimator with the gain in accuracy our complex estimator adds. As seen in Chapter 4, there are many circumstances where the simpler estimator does not only match the performance of the complex estimator, it in fact beats it. Furthering the theory of stochastic differential equations must be fettered by its applicability. While in the situations presented in Chapters 2 and 3 there is no such "simple" estimator, it is still vital to be sure that the estimators developed are at least somewhat practical.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1965.

[2] E. Alòs, O. Mazet, and D. Nualart. Stochastic calculus with respect to fractional Brownian motion with Hurst parameter lesser than $\frac{1}{2}$. *Stochastic Processes and their Applications*, 86:121–139, 2000.

[3] E. Alòs, O. Mazet, and D. Nualart. Stochastic calculus with respect to Gaussian processes. *The Annals of Probability*, 29(2):766–801, 2001.

[4] E. Alòs and D. Nualart. Stochastic integration with respect to the fractional Brownian motion. *Stochastics and stochastics reports*, 75(3):129–152, June 2003.

[5] F. Baudoin and D. Nualart. Equivalence of Volterra processes. *Stochastic Processes and their Applications*, 107:327–350, 2003.

[6] F. Baudoin and D. Nualart. Corrigendum to "Equivalence of Volterra processes". *Stochastic Processes and their Applications*, 115:701–703, 2005.

[7] A. Benassi, S. Jaffard, and D. Roux. Elliptic Gaussian random processes. *Rev. Mat. Iberoamericana*, 13(1):19–90, 1997.

[8] A. Le Breton. Filtering an d-parameter estimation in a simple linear system driven by fractional Brownian motion. *Statistics and Probability Letters*, (38):263–274, 1998.

[9] R.H. Cameron and W.T. Martin. Evaluation of various Wiener integrals by use of certain Sturm-Liouville differential equations. *Bull. Amer. Math. Soc*, 51:73–90, 1945.

[10] P. Carmona, L. Coutin, and G. Montseny. Stochastic integration with respect to fractional Brownian motion. *A.. I. H. Poincaré*, 39(1):27–68, 2003.

[11] R.S. Chhikara and L. Folks. *The inverse Gaussian distribution: theory, methodology, and applications*. CRC, 1989.

[12] A. Davies. *Handbook of condition monitoring: techniques and methodology*. Springer, 1997.

[13] L. Decreusefond. Regularity properties of some stochastic Volterra integrals with singular kernel. *Potential Analysis*, 16:139–149, 2002.

[14] L. Decreusefond. Stochastic integration with respect to Gaussian processes. *Probability Theory*, 334:903–908, 2002.

[15] L. Decreusefond. Stochastic integration with respect to Volterra processes. *Ann. I. H. Poincaré*, 41:123–149, 2005.

[16] L. Decreusefond and A. S. Üstünel. Stochastic analysis of the fractional Brownian motion. *Potential Analysis*, 10:177–214, 1999.

[17] M. Errami and F. Russo. n-covariation, generalized Dirichlet processes and calculus with respect to finite cubic variation processes. *Stochastic Processes and their Applications*, 104:259–299, 2003.

[18] F. Flandoli and F. Russo. Generalized integration and stochastic ODEs. *The Annals of Probability*, 30(1):270–292, 2002.

[19] F. Flandoli, F. Russo, and J. Wolf. Some SDEs with distributional drift. Part I: General Calculus. *Osaka J. Math.*, 40:493–542, 2003.

[20] M. Gradinaru, I. Nourdin, F. Russo, and P. Vallois. m-order integrals and generalized Itô's formula; the case of a fractional Brownian motion with any Hurst index. *Ann. I. H. Poincaré*, 41:781–806, 2005.

[21] L.P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054, 1982.

[22] C. Houdré and J. Villa. An example of infinite dimensional quasi-helix. *Contempory Mathematics*, 336:195–201, 2003.

[23] H. Hult. Approximating some Volterra type stochastic integrals with applications to parameter estimation. *Stochastic Processes and their Applications*, 105:1–32, 2003.

[24] M. Kac. On the average of a certain Wiener functional and a related limit theorem in calculus of probability. *Transactions of the American Mathematical Society*, 59(3):401–414, 1946.

[25] M. Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

[26] M. L. Kleptsyna and A. Le Breton. Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Statist. Infer. for Stoch. Proc.*, 5:229–248, 2002.

[27] M. L. Kleptsyna, A. Le Breton, and M.-C. Roubaud. An elementary approach to filtering in systems with fractional Brownian observation noise. *Probability Theory and Mathematical Statistics*, 1998.

[28] M. L. Kleptsyna, A. Le Breton, and M.-C. Roubaud. Parameter extimation and optimal filtering for fractional type stochastic systems. *Statist. Infer. for Stoch. Proc.*, 3:173–182, 2000.

[29] F. Klingenhöfer and M. Zähle. Ordinary differential equations with fractal noise. *Proceedings of the American Mathematical Society*, 127(4):1021–1028, 1999.

[30] I. Kruk, F. Russo, and C. A. Tudor. Wiener integrals, Malliavin calculus and covariance measure structure. *Journal of Funtional Analysis*, (249):92–142, 2007.

[31] A. Lachal. First exit time from a bounded interval for a certain class of additive functionals of Brownian motion. *Journal of Theoretical Probability*, 13(3):733–775, 2000.

[32] A. Lachal. Some explicit distributions related to the first exit time from a bounded interval for certain functionals of Brownian motion. *Journal of Theoretical Probability*, 19(4):757–771, 2006.

[33] D. Lepingle. Sur le comportement asymptotique des martingales locales. *Seminaire de Probabilites*, XII:148–161, 1978.

[34] M. Linn and A. Amirdjanova. Representations of the optimal filter in the context of nonlinear filtering of random fields with fractional noise. *Stochastic Processes and their Applications*, 119(8):2481–2500, 2009.

[35] R. S. Lipster and A. N. Shiryayev. *Theory of Martingales*. Kluwer Academic Publishers, 1989.

[36] S. V. Lotosky and K. Stemmann. From random processes to generalized fields: a unified approach to stochastic integration. *To appear.*

[37] B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

[38] L. Mátyás. *Generalized method of moments estimation.* Cambridge Univ Pr, 1999.

[39] O. Mocioalca and F. Viens. Skorohod integreation and stochastic calculus beyond the fractional Brownian scale. *Journal of Functional Analysis*, 222:385–434, 2005.

[40] V. Nair and Wang X. A Class of Inverse-Gaussian Models for Degradation Data. *Preprint.*

[41] A. Neuenkirch and I. Nourdin. Exact rate of convergence of some approximation schemes associated to SDEs driven by a fractional Brownain motion. *J. Theor. Probab*, 20:871–899, 2007.

[42] I. Norros, E. Valkeila, and J. Virtamo. An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli*, 5(4):571–587, 1999.

[43] D. Nualart and Y. Ouknine. Regularization of differential equations by fractional noise. *Stochastic Processes and their Applications*, 102(1):103–116, 2002.

[44] D. Nualart and A. Răşanu. Differential equations driven by fractional Brownian motion. *Collect. Math.*, 53(1):55–81, 2002.

[45] B. Øksendal. *Stochastic Differential Equations*. Springer, 5 edition, 2000.

[46] Y. Ouknine and M. Erraoui. Equivalence of Volterra processes: Degenerate case. *Statistics and Probability Letters*, 78(4):435–444, 2008.

[47] B. L. S. Prakasa Rao. *Semimartingales and Their Statistical Inference.* CRC Press, 1999.

[48] B. L. S. Prakasa Rao. *Statistical inference for diffusion type processes.* Arnold, London and Oxford University Press, New York, 1999.

[49] B. L. S. Prakasa Rao. Parametric estimation for linear stochastic differential equations driven by fractional Brownian motion. *Random Oper. and Stoch. Equ.*, 11(3):229–242, 2003.

[50] S.I. Resnick. *A Probability Path.* Birkhäuser, 1999.

[51] F. Russo. Stochastic calculus with respect to continuous finite quadratic variation processes. *Stochastics and stochastics reports*, 70:1–40, 2000.

[52] F. Russo and G. Trutnau. Some parabolic PDEs whose drift is an irregular random noise in space. *The Annals of Probability*, 35(6):2213–2262, 2007.

[53] F. Russo and C. A. Tudor. On bi-fractional Brownian motion. *Stochastic Processes and their Applications*, 116:830–856, 2006.

[54] F. Russo and P. Vallois. Forward, backward and symmetric stochastic integration. *Probab. Theory Relat. Fields*, 97:403–421, 1993.

[55] A. A. Ruzmaikina. Stieltjes integrals of Hölder continuous finctions with applications to fractional Brownian motion. *Journal of Statistical Physics*, 100(5/6):1049–1069, 2000.

[56] C. Tudor and M. Tudor. On the two-parameter fractional Brownian motion and Stieltjes integrals for Hölder functions. *Journal of Mathematical Analysis and Applications*, 286(2):765–781, 2003.

[57] C. A. Tudor and F. G. Viens. Statistical aspects of the fractional stochastic calculus. *Annals of Statistics*, 35(3):1183–1212, 2007.

[58] C. A. Tudor and Y. Xiao. Sample path properties of bi-fractional Brownian motion. *Bernoulli*, 13(4):1023–1052, 2007.

[59] L.C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica*, 67(1):251–282, 1936.

[60] M. Zähle. Integration with respect to fractal functions and stochastic calculus. I. *Probability Theory and Related Fields*, 111:333–374, 1998.