Unified Stochastic SIR Model with Parameter Estimation and Application to the COVID-19 Pandemic

Terry D. Easlick

A Thesis
in
The Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Mathematics) at Concordia University
Montréal, Québec, Canada

July 2023

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Terry D. Easlick**
Entitled: **Unified Stochastic SIR Model with Parameter Estimation and Application to the COVID-19 Pandemic**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mathematics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.
Signed by the Final Examining Committee:

_____ Chair
Dr. Mariana Frank

_____ External Examiner
Dr. Michael Kouritzin

_____ Examiner
Dr. Yang Lu

_____ Examiner
Dr. Lea Popovic

_____ Examiner
Dr. Arusharka Sen

_____ Supervisor
Dr. Wei Sun

Approved by _____
Dr. Lea Popovic, Graduate Program Director

_____ 2023

_____
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Terry D. Easlick, Ph.D.
Concordia University, 2023

This thesis is comprised of three parts which collectively serve as a study of stochastic epidemiological models, in particular, the susceptible, infected, recovered/removed (SIR) model.

I

We propose a unified stochastic SIR model driven by Lévy noise. The structural model allows for time-dependency, nonlinearity, discontinuity, demography and environmental disturbances. We present concise results on the existence and uniqueness of positive global solutions and investigate the extinction and persistence of the novel model. Examples and simulations are provided to illustrate the main results.

II

This part is twofold; we investigate the parameter estimation and forecasting of two forms of a stochastic SIR model driven by small Lévy noises, and we provide theoretical results on parameter estimation of time-dependent drift for Lévy noise-driven stochastic differential equations. A novel algorithm is introduced for approximating the least-squares estimators, which lack attainable closed-forms; moreover, the presented results ensure the consistency of these approximated estimators.

III

We apply the previous results to study the COVID-19 pandemic using data from New York City, New York. This application yields parameter estimation and predictive analysis, including the unknown period for a periodic transmission function and importation/exportation of infection.

Contribution of Authors

The work herein is based upon [21, 22]. All presented work has been a years-long collaboration between the authors Terry Easlick and Dr. Wei Sun. Both authors contributed questions, ideas, proofs, and edits to the work presented here.

Acknowledgments

I express immense gratitude to my supervisor, Dr. Wei Sun, for all the support you have provided– academic and non-academic. I have learnt much from you, and collaborating with you has been a pleasure.

Thank you, Dr. Lea Popovic, for teaching me much of what I know regarding stochastic differential equations. Thank you, Dr. Yang Lu, for your comments at the early stages of the work for this thesis which helped shape the work.

I would like to express my appreciation to all members of the thesis committee for taking the time to consider the work presented here.

I thank Dr. Lan Hong, the late Dr. Hoat Le and Dr. Nick Slinglend for their encouragement during the early stages of my academic pursuits.

Additionally, thank you to Dr. Arusharka Sen and Dr. Galia Dafni for graciously helping me navigate the Ph.D. program.

I state here the inexplicable gratefulness due to my parents, without whom I would not be here to do this, it saddens me neither of you is here to see me finish, but I know how much you believed in me; thank you. I also profoundly thank my wife, Julia, who has constantly encouraged me. Last but not least, I am grateful for my cat, Tarski, who has been with me through this entire journey.

I wish to thank Calcul Québec, a regional partner of the Digital Research Alliance of Canada, for providing high-performance computing resources needed to accomplish the work presented here. This work was supported by Concordia University Department of Mathematics and Statistics and the Natural Sciences and Engineering Research Council of Canada (No. 4394-2018).

Dedication

Dedicated to my late parents
Harold "Dan" Easlick 1956-2020
LaDean Easlick 1962-2022

Contents

List of Figures

List of Tables

Chapter 1

Motivation

Mathematical and scientific understanding of our world is an incremental endeavour to which we wish to add contributions herein. In this endeavour, there are far too many questions arising to be answered by our work or the work of others alone to be all-encompassing; thus, in the following pages, the reader will find many contributions, and many cited works attempting to answer some of these questions. Care and diligence were taken to ensure our contributions were novel and relevant.

1 Introduction

Real-world phenomena have been extensively modelled using deterministic systems and increasingly by stochastic systems. Uncertainties in the real-world necessitate the use of stochastic systems for modelling problems in many branches of science. One such interdisciplinary field is epidemiology, and there is no end in sight to the usefulness of modelling in this field. The recent COVID-19 pandemic shines a light on the importance of mathematical epidemiology. Moreover, the pandemic exhibited many uncertainties, thus justifying the investigation of stochastic modelling modes. To establish more realistic predictions of such phenomena, it is necessary to go beyond the deterministic models by including random perturbations. Stochastic differential equations (SDEs) have made such modelling possible.

In order to adequately investigate epidemiological phenomena, one must begin by establishing a model. Assuming a suitable model has then been established, it is necessary to study the dynamics of the model; namely, in the case of epidemiology, one particular question arises, "What information does the model provide about the extinction or persistence of disease?".

The epidemiological model we investigate is known as the susceptible, infected, recovered/removed (SIR) model introduced by Kermack and McKendrick [38] in 1927. This model is comprised of three compartments, the aforementioned susceptible, infected and recovered/removed. In its simplest form, a population (nearly in its entirety) begins in the susceptible compartment, then moves to the infected compartment according to some transmission rate β, and finally to the recovered/removed compartment via a recovery rate γ–a visual representation is provided in Figure 1.

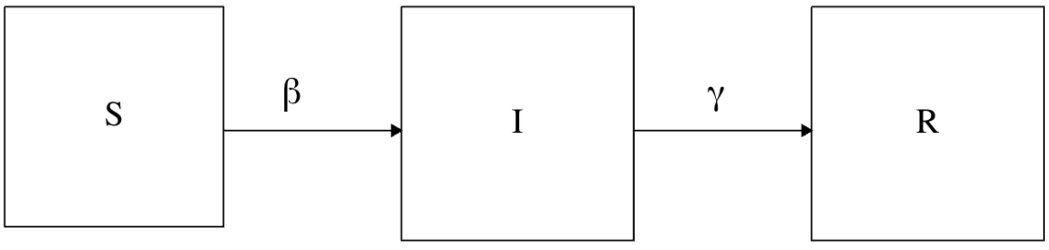

Figure 1: Diagram of compartments in the basic SIR model.

Remark 1.1 *For the recovered/removed compartment, the use of the words recovered or removed may be done exclusively. The former conveys the scenario where recovery is the primary end result, whereas the latter encapsulates the scenario when death and disease recovery is possible. As a real-world example, consider the common cold versus COVID-19.*

This thesis is the culmination of our work to establish a stochastic epidemiological model framework, study the dynamics of the model, form a parameter estimation regime and apply the work to the COVID-19 pandemic. The thesis is structured in the following manner:

Chapter 2: Necessary background

Given the interdisciplinary nature of this thesis, we include all the necessary preliminaries, including a review of epidemiological models, stochastic processes and SDEs with jumps such as

$$\begin{aligned} dX_t &= b(t,X_t)dt + \sigma(t,X_t)dB_t \\ &\quad + \int_{\{|u|\leq 1\}} H(t,X_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G(t,X_{t-},u)N(dt,du), \end{aligned} \tag{1.1}$$

along with important definitions and theorems necessary for the topics in the subsequent chapters.

Chapter 3: Establishing the unified stochastic SIR (USSIR) model
We define a generalized stochastic model of the SIR using coupled SDEs (dX_t, dY_t, dZ_t) driven by Lévy noise

$$\begin{cases} dX_t = b_1(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{1j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ + \int_{\{|u|\leq 1\}} H_1(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_1(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dY_t = b_2(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{2j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ + \int_{\{|u|\leq 1\}} H_2(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_2(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dZ_t = b_3(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{3j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ + \int_{\{|u|\leq 1\}} H_3(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_3(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \end{cases} \tag{1.2}$$

with the convention dX_t corresponds to the suceptible compartment, dY_t to the infected compartment, and dZ_t to the removed/recovered compartment. Stochastic epidemiological models, including the SIR, are an ongoing study (cf., e.g., [6, 13, 14, 16, 17, 18, 23, 24, 28, 29, 30, 35, 36, 43, 44, 51, 59, 64]). It is common in the literature to work with a specific,

explicit form of a stochastic SIR model. Contrast this with the model (1.2); it provides a foundation from which we obtain results concerning the dynamics of the model without the limitations of an explicit form. More specifically, our presented results on extinction and disease persistence are then readily applied to explicit stochastic SIR models. Simulation studies of explicit stochastic SIR models accompany our theoretical results.

Chapter 4: Parameter estimation of a novel stochastic SIR model

With a specific application in mind, parameter estimation of stochastic SIR models with periodic transmission (this work falls under the domain of the USSIR), we build upon the work of Long et al. [45, 46] in estimating the drift parameters of discretely observed small Lévy noise driven SDEs. Namely, given the SDE (1.1) parameterized by θ with the inclusion of a small dispersion coefficient ε

$$\begin{aligned} dX_t^\varepsilon &= b(t, X_t^\varepsilon, \theta)dt + \varepsilon\Bigg(\sigma(t, X_t^\varepsilon)dB_t \\ &\quad + \int_{\{|u|\leq 1\}} H(t, X_{t-}^\varepsilon, u)\tilde{N}(dt, du) + \int_{\{|u|>1\}} G(t, X_{t-}^\varepsilon, u)N(dt, du)\Bigg), \end{aligned}$$

and given finitely many, say n, observations, we perform estimation by the method of least squares via a contrast function of the form

$$\Psi_{n,\varepsilon}(\theta) = n\sum_{k=1}^{n} P_k^*(\theta)P_k(\theta),$$

where $\{t_k = \frac{k}{n}, k = 1, 2, ..., n\}$ and

$$P_k(\theta) = X_{t_k}^\varepsilon - X_{t_{k-1}}^\varepsilon - \frac{1}{n}b(t_{k-1}, X_{t_{k-1}}^\varepsilon, \theta).$$

In the pursuit of estimating time-dependent drift parameters, it is required to have the necessary theoretical underpinnings to support the validity of estimators. We provide an extension of the drift parameter estimations results by Long et al. [45, 46] to the scenario of time-dependent drift. Moreover, we take a more general approach to the driving noise; in particular, it may be singular or have an unobtainable explicit form. We include a novel approach to approximating the sought-after estimators for those without a readily available closed-form. The study of small noise-driven SDEs is an active domain of research (cf., e.g., [27], [37], [39], [45], [46], [55], [56], [60], [62]).

Chapter 5: Application to the COVID-19 pandemic using real-world data

Indeed, events such as COVID-19 will spur further epidemiological research, and SDEs serve as remarkable tools for such research. Some notable recent works are available in the literature cf. [1, 2, 4, 10, 33, 48, 50, 57, 58, 63]. In fact, the various *waves* occurring during the COVID-19 pandemic have also been studied (cf. [2, 48, 50]).

Data availability has increased immensely in recent years, including data for the COVID-19 pandemic. Collecting COVID-19 data is far from a perfected task; that is, a lot of noise arises, given the complexity and sheer scale. However, with valiant efforts made by many to ensure adequate data collection, we applied our theoretical results to real data. John Hopkins University was a leading institution in the collection and distribution of pandemic-related data for a wide variety of geographic locations. More specifically, COVID-19 Data Repository by the Centre for Systems Science and Engineering (CSSE) at John Hopkins University is the primary source of data utilized in this chapter. For our application, we honed in on New York City for reasons such as the density and population size, the availability of quality data, and it is a commonly visited geographic location. We introduce a stochastic SIR model with attributes such as the periodic transmission of infection and the importation/exportation of infection. Moreover, both attributes are natural to consider as COVID-19 occurred in waves, and there is no shortage of far-reaching travel in the modern world. This model is given by

$$\begin{bmatrix} dX_t \\ dY_t \\ dZ_t \end{bmatrix} = \begin{bmatrix} -\beta(t)X_t(Y_t+\varrho) \\ \beta(t)X_t(Y_t+\varrho)-\gamma Y_t \\ \gamma Y_t \end{bmatrix} dt \; + \; \varepsilon\sigma(t, X_{t-}, Y_{t-}, Z_{t-}) \begin{bmatrix} dL_t^{(1)} \\ dL_t^{(2)} \\ dL_t^{(3)} \end{bmatrix},$$

where $L_t = [L_t^{(1)}, L_t^{(2)}, L_t^{(3)}]$ is Lévy noise.

Chapter 6: Discussion

The thesis closes with a discussion of the contributions of the work completed herein. Included is a mention of future directions for the work as new questions have arisen during these studies.

Chapter 2

Preliminaries

2.1 Epidemiological models

The earliest known mathematical model of epidemiology dates back to the 18th century with the advent of a model for smallpox by Daniel Bernoulli. Many of the foundations for the modern study of mathematical epidemiology were developed in the early to mid-20th century. These foundations remain useful in recent times for the study of diseases, especially in the case of epidemics and pandemics such as the SARS epidemic of 2002-03 and the recent COVID-19 pandemic. As mathematical epidemiology is invaluable to society for the formulation of new policies and measures to curtail disease spread; moreover, mathematical models alleviate the absence of validation by experiment–an absence given such experiments would have questionable ethical standings. Nonetheless, there has been no shortage of disease outbreaks; hence there is no shortage of models and, subsequently, questions to consider. For a thorough introduction to mathematical epidemiology, we refer the reader to the exploratory texts of Brauer et al. [11, 12].

Communicable diseases are a fact of reality and the method of transmission is generally done via viral infection. These infections may occur horizontally (e.g., human-to-human interaction via biotic/abiotic methods) or vertically (e.g., parent-progeny relationship). Our focus is on COVID-19 which spreads primarily horizontally but may spread vertically during childbirth. Kermack and McKendrick [38] laid much of the groundwork in 1927 for epidemiological compartmental models, that is, a disease model in which the population is segmented into different compartments depending on their health status in relation to the disease. The starting point of much of this work on compartmental models was the introduction of the susceptible, infected, and recovered/removed (SIR) by Kermack and McKendrick [38]; moreover, the study of this model continues today. Commonly the SIR model given in the following version uses (deterministic) differential equations

$$\begin{cases} \frac{dX_t}{dt} = -\beta X_t Y_t, \\ \frac{dY_t}{dt} = (\beta X_t - \gamma)\, Y_t, \\ \frac{dZ_t}{dt} = \gamma Y_t, \end{cases} \tag{2.1}$$

where β is the transmission rate and γ the recovery rate. Additionally, vital demographics may be introduced to include birth rate Λ and mortality rate μ as:

$$\begin{cases} \frac{dX_t}{dt} = \Lambda - \mu X_t - \beta X_t Y_t, \\ \frac{dY_t}{dt} = [\beta X_t - (\mu + \gamma)]\, Y_t, \\ \frac{dZ_t}{dt} = \gamma Y_t - \mu Z_t. \end{cases} \tag{2.2}$$

The basic SIR models (2.1) and (2.2) have many variations or related models (cf. e.g., [5], [12]). The most notable related models which are also of interest to those active in the field are the susceptible, infected, susceptible (SIS); susceptible, infected, recovered/removed, susceptible

(SIRS); susceptible, exposed, infected, removed/recovered (SEIR). This is a non-exhaustive list, and the specific model considered will be disease dependent. Our focus has been the extension of the (stochastic) SIR model, but our work can be extended further to include related models.

The introduction of stochastics to mathematical epidemiology gave way to a currently active research area. Namely, the study of these models using SDEs as opposed to deterministic differential equations has a growing sector of active research (cf. e.g., [6, 13, 14, 16, 17, 18, 23, 24, 28, 29, 30, 35, 36, 43, 44, 51, 59, 64])–including for a stochastic SIR model commonly given in a basic form

$$\begin{cases} \frac{dX_t}{dt} = -\beta X_t Y_t - \sigma X_t Y_t dB_t, \\ \frac{dY_t}{dt} = (\beta X_t - \gamma) Y_t + \sigma X_t Y_t dB_t, \\ \frac{dZ_t}{dt} = \gamma Y_t, \end{cases} \tag{2.3}$$

and as before, the model (2.3) also yields many variations and related models (cf. e.g., [5], [12]).

The mathematical necessities are given in the following section prior to the presentation of our results.

2.2 Stochastic differential equations driven by Lévy noises

We refer the reader to [3] for the notation and terminology used below. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a complete probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (i.e., it is increasing, right continuous and $\mathcal{F}_0$ contains all P-null sets).

Definition 2.2 *An adapted continuous-time $\mathbb{R}^n$-valued stochastic process $X = (X_t)_{t\geq 0}$ with $X_0 = 0$ a.s. is a Lévy process if*

1. *X has increments independent of the past, i.e., $X_t - X_s$ is independent of $\mathcal{F}_s$ for any $0 \leq s < t$.*
2. *X has stationary increments, i.e., $X_t - X_s \sim X_{t-s}$ for any $0 \leq s < t$.*
3. *X is stochastically continuous, i.e., $X_s \to X_t$ in probability $s \to t$ for any $t \geq 0$.*

Remark 2.3 *It can be shown that any Lévy process has a unique modification which is càdlàg, i.e., the paths $t \to X_t$ are all right-continuous with left limits. In this thesis, we will only consider càdlàg stochastic processes; hence, we will often omit the word càdlàg with the understanding it is implied. The standard n-dimensional Brownian motion (Wiener process) $(B_t)_{t\geq 0}$ and the Poisson process $(N_t)_{t\geq 0}$ with intensity $\lambda > 0$ are two typical examples of Lévy processes.*

Let X be a Lévy process. We define the jump process $\Delta X = (\Delta X_t)_{t\geq 0}$ of X by

$$\Delta X_t = X_t - X_{t-}, \qquad t > 0.$$

Hereafter X_{t-} denotes the left limit at the point t. For $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R}^n - \{0\})$, define

$$N(t, A) = \sum_{0<s\leq t} 1_A(\Delta X_s),$$

and

$$\mu(A) = E[N(1, A)].$$

μ is called the Lévy measure (intensity measure) of X. We have

$$\int_{\mathbb{R}^n - \{0\}} (1 \wedge |y|^2)\mu(dy) < \infty.$$

Hereafter $|y| := \left(\sum_{i=1}^n y_i^2 \right)^{\frac{1}{2}}$ denotes the Euclidean norm. We define the compensated Poisson random measure by

$$\widetilde{N}(t, A) = N(t, A) - t\mu(A).$$

For $A \in \mathcal{B}(\mathbb{R}^n - \{0\})$, $(N(t, A))_{t\geq 0}$ is a Poisson process with intensity $\mu(A)$.

Theorem 2.4 (Lévy-Itô decomposition) *(cf. [3, Theorem 2.4.16]) Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^n$. Then, there exist $b \in \mathbb{R}^n$, a Brownian motion B^A with covariance matrix A, and an independent Poisson random measure N on $\mathbb{R}_+ \times (\mathbb{R}^n - \{0\})$ with compensated Poisson measure $\widetilde{N}$ and Lévy measure μ, such that for each $t \geq 0$,*

$$X_t = bt + B_t^A + \int_{\{|x|\leq 1\}} x\tilde{N}(t, dx) + \int_{\{|x|>1\}} xN(t, dx).$$

An adapted stochastic process $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ is called a semi-martingale if it may be written as

$$X_t = X_0 + M_t + A_t,$$

where $(M_t)_{t\geq 0}$ is a local martingale and $(A_t)_{t\geq 0}$ is an adapted process possessing finite variation.

Proposition 2.5 *(cf. [3, Propostion 2.7.1]) Every Lévy process is a semi-martingale.*

Let B be a k-dimensional standard Brownian motion and N an independent Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}^l - \{0\})$ with intensity measure μ. The SDEs driven by Lévy noise considered herein are of the form

$$\begin{aligned} dX_t &= b(t, X_t)dt + \sigma(t, X_t)dB_t \\ &\quad + \int_{\{|u|\leq 1\}} H(t, X_{t-}, u)\tilde{N}(dt, du) + \int_{\{|u|>1\}} G(t, X_{t-}, u)N(dt, du), \end{aligned} \tag{2.4}$$

with $X_0 \in \mathcal{F}_0$. We assume that the coefficient functions $b(t, x) : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma(t, x) : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times k}$, $H(t, x, u) : [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n$ and $G(t, x, u) : [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n$ are all Borel measurable.

Let $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n; \mathbb{R})$ be the space of all functions $f(t,x)$ from $\mathbb{R}_+ \times \mathbb{R}^n$ to $\mathbb{R}$ such that they are continuously differentiable and twice continuously differentiable with respect to t and x, respectively. Hereafter, we set $f_t = \frac{\partial f}{\partial t}$, $f_x = \nabla_x f$ and $f_{xx} = (\frac{\partial^2 f}{\partial x_i \partial x_j})_{n\times n}$. A vital theorem used in virtually any discussion or application of these SDEs is the following

Theorem 2.6 (Itô's formula for SDE driven by Lévy noises) *(cf. [3, Theorem 4.4.7]) Let $\{X_t\}$ be the solution to the SDE (2.4). Then, for $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n; \mathbb{R})$, we have*

$$\begin{aligned} df(t,X_t) &= f_t(t,X_t)dt + \langle f_x(t,X_t), b(t,X_t)\rangle dt + \frac{1}{2}\text{trace}\left[\sigma^T(t,X_t) f_{xx}(t,X_t)\sigma(t,X_t)\right] dt \\ &\quad + \int_{\{|u|\le 1\}} \left[f(t,X_{t-} + H(t,X_{t-},u)) - f(t,X_{t-}) - \langle f_x(t,X_{t-}), H(t,X_{t-},u)\rangle\right] \mu(du)dt \\ &\quad + \langle f_x(t,X_t), \sigma(t,X_t)\rangle dB_t + \int_{\{|u|\le 1\}} \left[f(t,X_{t-} + H(t,X_{t-},u)) - f(t,X_{t-})\right] \tilde{N}(dt,du) \\ &\quad + \int_{\{|u|>1\}} \left[f(t,X_{t-} + G(t,X_{t-},u)) - f(t,X_{t-})\right] N(dt,du). \end{aligned}$$

Moreover, consider a function $V \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n; \mathbb{R})$, define the operator $\mathcal{L}$ on V such that

$$\begin{aligned} \mathcal{L}V(t,x) &:= V_t(t,x) + \langle V_x(t,x), b(t,x)\rangle + \frac{1}{2}\text{trace}(\sigma^T(t,x)V_{xx}(t,x)\sigma(t,x)) \\ &\quad + \int_{\{|u|\le 1\}} [V(t,x + H(t,x,u)) - V(t,x) - \langle V_x(t,x), H(t,x,u)\rangle]\mu(\mathrm{d}u) \\ &\quad + \int_{\{|u|>1\}} [V(t,x + G(t,x,u)) - V(t,x)]\mu(\mathrm{d}u). \end{aligned}$$

The following result gives an inequality which proves invaluable when dealing with SDEs–as will be seen in later chapters.

Proposition 2.7 (Gronwall's inequality) *(cf. [3, Proposition 6.1.4]) Let $[a,b] \subset \mathbb{R}$ be a closed interval and $\alpha, \beta : [a,b] \to \mathbb{R}$ be non-negative such that α is locally bounded and β is integrable. If there exists $C \ge 0$ for which all $t \in [a,b]$,*

$$\alpha(t) \le C + \int_a^t \alpha(s)\beta(s)ds$$

then we have

$$\alpha(t) \le C \exp\left(\int_a^t \beta(s)ds\right).$$

The following result provides a valuable inequality in the study of the long-time average limit; namely, it is helpful in our study of the persistence of disease.

Lemma 2.8 *(cf. [36, Lemma 5.1]) Let $f \in C([0,\infty)\times\Omega;(0,\infty))$. If there exist positive constants λ_0, λ such that*

$$\log(f(t)) \geq \lambda t - \lambda_0 \int_0^t f(s)ds + F(t), \ a.s.$$

for all $t \geq 0$, where $F \in C([0,\infty)\times\Omega;\mathbb{R})$ and $\lim_{t\to\infty} \frac{F(t)}{t} = 0$ a.s., then

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t f(s)ds \geq \frac{\lambda}{\lambda_0} \, a.s..$$

2.3 Parameter estimation of SDEs and optimization

The study and subsequent parameter estimation of discretely observed diffusion processes began in earnest in the 1970s with the works of Dorogovtsev ([20]) and Le Breton ([41]). The former focused on approximations of the maximum likelihood, and the latter turned their attention to (conditional) least-squares estimators. Given some diffusion process $(X_t)_{t\geq 0}$ along with a sequence $\{X_k\}_{k\leq n}$ of observations, the (conditional) least-squares estimation regime set out to minimize a function

$$Q_n(\theta) = \sum_{k=1}^{n} \frac{(X_{t_k} - X_{t_{k-1}} - \mu(\theta, t_{k-1}, X_{t_{k-1}})\Delta_{t_k})^2}{\sigma^2(t_{k-1}, X_{t_{k-1}})\Delta_{t_k}},$$

where the timesteps are evenly-spaced such that $\Delta_{t_k} = \frac{T}{n}, T > 0$. Dorogovtsev (1976) [20] proved weak consistency of the obtained (conditional) LSE estimator $\hat{\theta}_{n,T}$ for $T \to \infty, \frac{T}{n} \to 0$. Later in 1988, Kasonga [37] proved strong consistency of the same estimator. Later in the 1990s, the work of Genon-Catalot [26] and Laredo [40] established results for drift parameter estimation for discretely observed diffusions with a small dispersion coefficient ε tending to 0. Moving forward to the previous decade, the work of Long et al. [45, 46] extended the drift parameter estimation of small diffusions to small jump-diffusions, namely, SDEs driven by Lévy noises. Their work directly inspired our work; namely, we are still considering discrete observations, yet now our processes are time-dependent jump-diffusions or time-dependent SDEs driven by Lévy noises with a small dispersion rate ε tending to 0. For a thorough treatise of the history and many of the available methods in SDE parameter estimation, we refer the reader to [9, Bishwal].

The following two theorems are necessary for our results on parameter estimation given in Chapter 4.

Given a random variable sequence $\{\xi_n\}_{n\geq 0}$ which converges to 0 in probability we will then write

$$\xi_n = o_P(1).$$

Theorem 2.9 *(cf. [61, Theorem 5.7]) Let M_n be random functions and M be a fixed function of θ such that for every $\varepsilon > 0$,*

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0,$$

$$\sup_{\theta : d(\theta,\theta_0)\geq\varepsilon} M(\theta) < M(\theta_0).$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_P(1)$ converges in probability to θ_0.

Theorem 2.10 *(cf. [61, Theorem 5.9]) Let Ψ_n be random vector-valued functions and Ψ be a fixed vector-valued function of θ such that for every $\varepsilon > 0$,*

$$\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{P} 0,$$

$$\inf_{\theta : d(\theta, \theta_0) \geq \varepsilon} \|\Psi(\theta)\| > 0 = \|\Psi(\theta_0)\|.$$

Then any sequence of estimators $\hat{\theta}_n$ with $\Psi_n(\hat{\theta}_n) = o_P(1)$ converges in probability to θ_0.

In an attempt to estimate parameters, it is reasonable to view the estimation problem as an optimization problem. Namely, given some function $f(\theta)$ of the parameters we wish to estimate, we can approach this by asking *"what value of θ optimizes the function f?"*. Moreover, as is the case for our studies, finding a closed-form of our sought estimator $\hat{\theta}$ is not possible; thus, an approximation $\hat{\theta}^*$ is the next best option. A concept from optimization we make use of is the method of gradient descent (cf. [7, Chapter 8]). Briefly, suppose we are given an $\mathbb{R}$-valued, convex objective function $f(\theta)$ and wish to minimize to find

$$\arg\min_{\theta \in \Theta} f(\theta).$$

The method of gradient descent accomplishes this by using an update rule

$$\theta^{(k+1)} = \theta^{(k)} - \eta_k \nabla f(\theta^{(k)})$$

for some initial estimate $\theta^{(0)}$ and learning rate (stepsize) $\eta > 0$, where k denotes the current iterations. The number of iterations varies depending on the complexity and accuracy desired/-possible. Gradient descent is widely used in unconstrained optimization; however, as will be seen in our work, constraints are often present, and the use of gradient descent is still possible. In the presence of constraints, a specialized version of gradient descent called projected gradient descent (PGD) is utilized (cf. [7, Chapter 10]). Again, given an $\mathbb{R}$-valued, convex function f along with a constraint set $\mathcal{C} \subset \mathbb{R}^n$ we can proceed via a modified update rule

$$\xi^{(k+1)} = \theta^{(k)} - \eta_k \nabla f(\theta^{(k)}); \quad \theta^{(k+1)} = \arg\min_{\theta \in \mathcal{C}} \|\xi^{(k+1)} - \theta\|^2. \tag{2.5}$$

Namely, it is very similar to the (unconstrained) gradient step but the inclusion of one more step where we obtain the updated parameter value $\xi^{(k+1)}$ then project to obtain an updated $\theta^{(k+1)}$.

The method of determining η_k is known as a *line search* algorithm (cf. [7, Chapter 8]), the most common method of this is called *steepest descent* and in this η_k is chosen according to the rule

$$\eta_k = \arg\min_{\eta \geq 0} f(\theta^{(k)} - \eta \nabla f(\theta^{(k)})).$$

The following lemma and theorem are known results regarding the convergence analysis of this methodology.

Lemma 2.11 (Descent lemma) *(cf. [7, Lemma 5.7]) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying*

$$\|\nabla f(\theta_1) - \nabla f(\theta_2)\| \leq L\|\theta_1 - \theta_2\|,\ L > 0.$$

Then for a convex set $\mathcal{D}$ and for any $\theta_1, \theta_2 \in \mathcal{D}$ we have

$$f(\theta_2) \leq f(\theta_1) + \langle \nabla f(\theta_1), \theta_2 - \theta_1 \rangle + \frac{L}{2}\|\theta_1 - \theta_2\|^2.$$

Given a function f on $\mathbb{R}^n$ for which we wish to optimize over a constraint set $\mathcal{C}$, we note

$$\arg\min_{\theta \in \mathcal{C}} f(\theta) \equiv \arg\min_{\theta \in \mathbb{R}^n} \{f(\theta) + \chi(\theta)\},$$

where $\chi(\theta)$ takes value 0 for $\theta \in \mathcal{C}$; otherwise, takes value ∞.

Theorem 2.12 ($O(k^{-1})$ rate of convergence of PGD) *(cf. [7, Theorem 10.21]) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function satisfying*

$$\|\nabla f(\theta_1) - \nabla f(\theta_2)\| \leq L\|\theta_1 - \theta_2\|,\ L > 0.$$

Assume the following

(i) the problem $\min_{\theta \in \mathbb{R}^n}\{f(\theta) + \chi(\theta)\}$ has a non-empty optimal set denoted Θ^*,

(ii) $\{\theta^{(k)}\}$ is a sequence of values generated by the application of the PGD method (cf. (2.5)).

Then for fixed stepsize $\eta \leq \frac{1}{L}$, any $\theta^* \in \Theta^*$, and $k > 0$,

$$f(\theta^{(k)}) - f_{opt} \leq \frac{\|\theta^{(0)} - \theta^*\|^2}{2\eta k},$$

where f_{opt} is the optimal value.

This has been a brief introduction which is sufficient for the understanding of the material contained in Chapter 4. We refer the reader to [7, Beck] for a more in-depth exploration.

2.4 Scientific programming with Julia

With the increase in data, there is also an increased need for computational performance. For those possessing (scientific) computational tasks, there are many options available, including *Fortran*, *Julia*, *MATLAB*, *Python*, and *R*. For the work of this thesis, being able to perform computational tasks such as simulation studies is of great importance–both for the theoretical and applied. The *Julia* programming language was explicitly designed for mathematical and scientific computing (cf. [8, Bezanson et al.]). The primary centre for the development of *Julia* is the Julia Lab at MIT; however, active research and development are ongoing at numerous institutions worldwide–it is open-source with 1000s of active contributors. A strong benefit of *Julia* is the intentional development for use in the computational sciences; it comes with immense capability from the start, and there is a growing ecosystem of packages (supplementary functionality) to suit any conceivable computational need. *Julia* is

- a high-level programming language; namely, it affords the user strong abstraction while being easily readable by users;
- general purpose meaning that it applies to building software in a wide variety of domains;
- dynamically-typed allowing users to write more concise code which is versatile and flexible;
- fast; namely, when its code is optimized, it competes with some of the fastest languages such as *C++* and *Fortran*.

Beyond the "out-of-the-box ready-to-use" *Julia*, there exists the *Julia* SciML ecosystem which is a robust collection of *Julia* packages for computational needs. For instance, needs such as ours in the simulations of the USSIR model simulation or our parameter estimation efforts. Two vital packages in this ecosystem are *DifferentialEquations.jl* and *Optim.jl* (cf. [47] and [53]).

2.4.1 *DifferentialEquations.jl*

In short, *DifferentialEquations.jl* is a package with the sole intent of solving differential equations; moreover, it covers many varieties of DEs, from ordinary (ODE) to partial (PDE) to random (RDE) to stochastic (SDE/SPDE) (cf. [53]). Our domain of use is for SDEs, and given the robustness of this package, we were not constrained in our ability to model and simulate the SDEs describing the USSIR. Namely, everything from defining an SDE problem, including the customizability of the driving noise to the ability to plot simulations, is provided. As a brief example of use, consider a basic stochastic SIR taking the form

$$\begin{cases} dX_t = -\beta X_t Y_t - \sigma X_t Y_t dB_t, \\ dY_t = \beta X_t Y_t - \gamma Y_t + \sigma X_t Y_t dB_t, \\ dZ_t = \gamma Y_t. \end{cases}$$

We can then use *DifferentialEquations.jl* to simulate this model using the following code:

```
using DifferentialEquations

function drift(du,u,p,t)
du[1] = -β*u[1]*u[2]
du[2] = β*u[1]*u[2] -γ*u[2]
du[3] = γ*u[2]
end

function noise(du,u,p,t)
du[1] = -u[1]*u[2]
du[2] = u[1]*u[2]
end

u₀ = [x_0,y_0,z_0]
dt = 0.001
tspan = (0.0, 1.0)
prob = SDEProblem(drift, noise, u₀, (0.0, 1.0))
sol = solve(prob, EM(), dt = dt)
```

2.4.2 *Optim.jl*

Optim.jl provided all the functionality necessary for our parameter estimation efforts, recall we approached this problem from the perspective of optimization (cf. [47]). This alleviated any need to write and test code for the execution of gradient-descent necessary in our methodology; furthermore, also made writing the necessary code less cumbersome for executing the LS-GD algorithm introduced in Chapter 4. Consider the example immediately above, and say we would like to estimate the parameters β and γ by use of our contrast function $\Psi_{n,\varepsilon}$ (cf. Chapter 4); we can then use gradient descent via *Optim.jl*. First, assume we have n-many observations of the model, then take this collection of observations and save it as *sol* in the form of a readily-useable data structure (commonly a CSV file), then define our problem in code as

```
using Optim; Using Dataframes
function contrast(par)
        sol = DataFrame(CSV.File("/pathname_to_sol"))
        Φ = []
        sampN = \# of observations
        for i in 1:sampN-1
         xk = sol[i+1,2];yk = sol[i+1,3];zk = sol[i+1,4]
          xj = sol[i,2];yj = sol[i,3];zj = sol[i,4]
          t  = sol[i,1]
          β = par[1]; γ = par[2]

          pk = [xk - xj - (-β*xj*yj/sampN)
                  yk - yj - (β*xj*yj-γ*yj)/sampN)
                 zk - zj - γ*yj)/sampN ]
          Φκ = transpose(pk)*pk
           push!(Φ, Φκ)
        end
        sampN*sum(Φ)
    end

    initial_x = [β_0, γ_0]
    results = Optim.optimize(contrast,initial_x,GradientDescent())
```

The returned results will be saveable for future uses (e.g., simulations using the found parameter values).

Ultimately, the ease of use and performance of *Julia* with these two packages allowed for less time writing and testing code and more time completing the necessary simulation studies presented here.

Chapter 3

A unified stochastic SIR model driven by Lévy noise with time-dependency

Introduction

Epidemiological compartment models have garnered much attention from researchers in an attempt to understand better and control the spread of infectious diseases. Mathematical analysis of such models aids decision-making regarding public health policy changes – especially in the event of a pandemic (e.g., COVID-19). Recall the aforementioned basic SIR by Kermack and McKendrick [38]

$$\begin{cases} \frac{dX_t}{dt} = -\beta X_t Y_t, \\ \frac{dY_t}{dt} = (\beta X_t - \gamma) Y_t, \\ \frac{dZ_t}{dt} = \gamma Y_t, \end{cases} \tag{3.1}$$

where β is the transmission rate and γ the recovery rate. As mentioned prior, the basic SIR model (3.1) has many variations, including the SIRD, SIRS, SIRV, SEIR, MSIR, etc. (cf., e.g., [5], [12] and [54]). The deterministic model has proven a useful tool, and to instill further realistic behaviours, there are been varying stochastic framework extensions. (cf. e.g., [6, 13, 14, 16, 17, 18, 23, 24, 28, 29, 30, 35, 36, 43, 44, 51, 59, 64]). The existing models are often analyzed with a focus on specific diseases or parameters. Such studies have been very successful in achieving new results; however, often, it is the case that structural variability is lacking in these models. To overcome the drawbacks inherent in traditional approaches, we propose and investigate in this chapter the unified stochastic SIR (USSIR) model.

Hereafter, $\mathbb{R}_+$ denotes the set of all positive real numbers, $\left(B_t^{(1)}, \ldots, B_t^{(r)}\right)_{t\geq 0}$ is a standard r-dimensional Brownian motion, N is a Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}^l - \{0\})$ with intensity measure μ satisfying $\int_{\mathbb{R}^l - \{0\}} (1 \wedge |u|^2)\mu(du) < \infty$ and $\tilde{N}(dt, du) = N(dt, du) - \mu(du)dt$, $\left(B_t^{(1)}, \ldots, B_t^{(r)}\right)_{t\geq 0}$ and N are independent, $b_i, \sigma_{ij} : [0, \infty) \times \mathbb{R}_+^3 \mapsto \mathbb{R}$, $H_i, G_i : [0, \infty) \times \mathbb{R}_+^3 \times (\mathbb{R}^l - \{0\}) \mapsto \mathbb{R}$, $i = 1, 2, 3,\ j = 1, 2, \ldots, n$, are measurable functions.

The USSIR model is defined by

$$\begin{cases} dX_t = \ b_1(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{1j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_1(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_1(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dY_t = \ b_2(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{2j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_2(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_2(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dZ_t = \ b_3(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r} \sigma_{3j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_3(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_3(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du). \end{cases} \tag{3.2}$$

In this chapter, based on [21, E. and Sun], results are given of a model that allows for time-dependency, nonlinearity (of drift, diffusion and jump) and demography. Environmental disturbances can profoundly affect transmission, recovery, mortality and population growth, and the above model encapsulates such stochastic perturbations driven by Brownian motions $(B_t^{(j)})$ with intensities $\sigma_{ij}(t,X_t,Y_t,Z_t)$ and a Poisson random measure $N(dt,du)$ with small jumps $H_i(t,X_{t-},Y_{t-},Z_{t-},u)$ and large jumps $G_i(t,X_{t-},Y_{t-},Z_{t-},u)$. An important structural feature we emphasize is time-dependency; this feature can capture the progression of a disease insofar as mutations/transmissibility (e.g., Delta and Omicron variants of COVID-19, vaccination programs).

In the following sections, we establish results on the existence and uniqueness of positive global solutions, extinction and persistence of diseases, and provide illustrative examples and simulations. Section 3.1 is concerned with the USSIR model (3.2) for population proportions, whereas Section 3.2 covers the USSIR model (3.2) for population numbers. Both approaches are commonly found in studies of SIR models; hence, the importance of investigation for a unifying model. In section 3.3, we present simulations which correspond to examples given in the previous two sections. At the time of writing [21], there was no existing work on the USSIR model, and the aim was to add a novel model as such to the existing literature.

3.1 Model for population proportions

In this section, we let X_t, Y_t and Z_t denote respectively the proportions of susceptible, infected and recovered populations at time t. Define

$$\Delta := \{(x,y,z) \in \mathbb{R}^3_+ : x+y+z = 1\}.$$

For $t \in [0,\infty)$, $(x,y,z) \in \Delta$ and $u \in (\mathbb{R}^l - \{0\})$, define

$$k(t,x,y,z,u) := \frac{H_1(t,x,y,z,u)}{x} + \frac{H_2(t,x,y,z,u)}{y} + \frac{H_3(t,x,y,z,u)}{z} - \ln\left\{\left(1+\frac{H_1(t,x,y,z,u)}{x}\right)\left(1+\frac{H_2(t,x,y,z,u)}{y}\right)\left(1+\frac{H_3(t,x,y,z,u)}{z}\right)\right\}. \quad (3.3)$$

For any $T > 0$, let $L^1_+[0,T]$ denote the set of positive, integrable functions on $[0,T]$. We make the following assumptions.

(A1) There exists $(x_0,y_0,z_0) \in \Delta$ such that for any $T \in (0,\infty)$ and $i = 1,2,3$,

$$b_i(\cdot,x_0,y_0,z_0),\ \sum_{j=1}^{n}|\sigma_{ij}(\cdot,x_0,y_0,z_0)| \in L^2[0,T],\ \int_{\{|u|\le 1\}}|H_i(\cdot,x_0,y_0,z_0,u)|^2\mu(du) \in L^1[0,T].$$

(A2) For any $T \in (0,\infty)$ and $N \in \mathbb{N}$, there exists $K_{N,T} \in L^1_+[0,T]$ such that

$$\begin{aligned} &\sum_{i=1}^{3}|b_i(t,x_1,y_1,z_1) - b_i(t,x_2,y_2,z_2)|^2 + \sum_{i=1}^{3}\sum_{j=1}^{r}|\sigma_{ij}(t,x_1,y_1,z_1) - \sigma_{ij}(t,x_2,y_2,z_2)|^2 \\ &+\sum_{i=1}^{3}\int_{\{|u|\le 1\}}|H_i(t,x_1,y_1,z_1,u) - H_i(t,x_2,y_2,z_2,u)|^2\mu(du) \\ \le\ & K_{N,T}(t)[(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2], \\ & \forall t \in [0,T],\ (x_1,y_1,z_1),(x_2,y_2,z_2) \in \left[\frac{1}{N}, 1-\frac{1}{N}\right]^3. \end{aligned}$$

(A3) For any $t \in (0,\infty)$, $(x,y,z) \in \Delta$ and $u \in (\mathbb{R}^l - \{0\})$,

$$\begin{aligned} &\sum_{i=1}^{3} b_i(t,x,y,z) = 0,\ \ \sum_{i=1}^{3}\sigma_{ij}(t,x,y,z) = 0 \text{ for } j = 1,2,\ldots,n, \\ &\sum_{i=1}^{3} H_i(t,x,y,z,u) = 0,\ \ \sum_{i=1}^{3} G_i(t,x,y,z,u) = 0. \end{aligned}$$

(A4) For any $(x,y,z) \in \Delta$, $t \in (0,\infty)$ and $u \in (\mathbb{R}^l - \{0\})$,

$$\left(1+\frac{H_1(t,x,y,z,u)}{x}\right),\ \left(1+\frac{H_2(t,x,y,z,u)}{y}\right),\ \left(1+\frac{H_3(t,x,y,z,u)}{z}\right) > 0,$$

and

$$\left(1+\frac{G_1(t,x,y,z,u)}{x}\right),\ \left(1+\frac{G_2(t,x,y,z,u)}{y}\right),\ \left(1+\frac{G_3(t,x,y,z,u)}{z}\right) > 0.$$

(A5) For any $T \in (0,\infty)$,

$$\inf_{t\in[0,T],\,(x,y,z)\in\Delta}\left\{\frac{b_1(t,x,y,z)}{x}+\frac{b_2(t,x,y,z)}{y}+\frac{b_3(t,x,y,z)}{z}\right\}>-\infty,$$

$$\sum_{j=1}^{r}\sup_{t\in[0,T],\,(x,y,z)\in\Delta}\left\{\frac{|\sigma_{1j}(t,x,y,z)|}{x}+\frac{|\sigma_{2j}(t,x,y,z)|}{y}+\frac{|\sigma_{3j}(t,x,y,z)|}{z}\right\}<\infty,$$

and there exists $\eta_T \in \mathcal{B}(\mathbb{R}^l)$ such that

$$\eta_T>0,\qquad \int_{\{|u|\le 1\}}\eta_T(u)\mu(du)<\infty, \tag{3.4}$$

and

$$\sup_{t\in[0,T],\,(x,y,z)\in\Delta,u\in(\mathbb{R}^l-\{0\})}\frac{k(t,x,y,z,u)}{\eta_T(u)}<\infty.$$

First, we discuss the existence and uniqueness of solutions to the system (3.2).

Theorem 3.13 *Suppose that Assumptions* **(A1)**–**(A5)** *hold. Then, for any given initial value* $(X_0,Y_0,Z_0)\in\Delta$*, the system (3.2) has a unique, strong solution taking values in* Δ*.*

Proof. By **(A3)**, **(A4)** and the interlacing technique (cf. [3]), to complete the proof, we need only consider the case that $G_i\equiv 0$, $i=1,2,3$. Then, equation (3.2) becomes

$$\begin{cases} dX_t = b_1(t,X_t,Y_t,Z_t)+\sum_{j=1}^{r}\sigma_{1j}(t,X_t,Y_t,Z_t)dB_t^{(j)}+\int_{\{|u|\le1\}}H_1(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du),\\ dY_t = b_2(t,X_t,Y_t,Z_t)+\sum_{j=1}^{r}\sigma_{2j}(t,X_t,Y_t,Z_t)dB_t^{(j)}+\int_{\{|u|\le1\}}H_2(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du),\\ dZ_t = b_3(t,X_t,Y_t,Z_t)+\sum_{j=1}^{r}\sigma_{3j}(t,X_t,Y_t,Z_t)dB_t^{(j)}+\int_{\{|u|\le1\}}H_3(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du). \end{cases} \tag{3.5}$$

By **(A1)** and **(A2)**, similar to [31, Lemma 2.1], we can show that there exists a unique local strong solution to equation (3.5) on $[0,\tau)$, where τ is the explosion time. We will show below that $\tau=\infty$ a.s.. Define

$$\tau_N=\inf\left\{t\in[0,\tau):(X_t,Y_t,Z_t)\notin\left[\frac{1}{N},1-\frac{1}{N}\right]^3\right\},\qquad N\in\mathbb{N},$$

and

$$\tau_\infty=\lim_{N\to\infty}\tau_N.$$

We have that $\tau_\infty\le\tau$ so it suffices to show $\tau_\infty=\infty$ a.s.. Hence assume the contrary that there exist $\varepsilon>0$ and $T>0$ such that

$$P(\tau_\infty<T)>\varepsilon,$$

which implies that

$$P(\tau_N < T) > \varepsilon, \quad \forall N \in \mathbb{N}. \tag{3.6}$$

Define

$$V(x,y,z) = -\ln(xyz), \quad (x,y,z) \in (0,1)^3,$$

and

$$W_t = (X_t, Y_t, Z_t).$$

By Itô's formula, we obtain that for $t \le \tau_N$,

$$\begin{aligned} V(W_t) &= V(W_0) + \int_0^t \mathcal{L}V(s,W_s)ds + \int_0^t \langle V_w(W_s), \sigma(s,W_s)\rangle dB_s \\ &\quad + \int_0^t \int_{\{|u|\le 1\}} [V(W_{s-} + H(s,W_{s-},u)) - V(W_{s-})]\,\tilde{N}(ds,du), \end{aligned}$$

Hereafter, for $t \ge 0$ and $w = (w_1, w_2, w_2) \in \mathbb{R}^3$,

$$\begin{aligned} \mathcal{L}V(t,w) &= \langle V_w(w), b(t,w)\rangle + \frac{1}{2}\mathrm{trace}(\sigma^T(t,w)V_{ww}(w)\sigma(t,w)) \\ &\quad + \int_{\{|u|\le 1\}} [V(w + H(t,w,u)) - V(w) - \langle V_w(w), H(t,w,u)\rangle]\mu(\mathrm{d}u), \end{aligned}$$

with $V_w = \nabla_w V = (\frac{\partial V}{\partial w_1}, \frac{\partial V}{\partial w_2}, \frac{\partial V}{\partial w_3})$ and $V_{ww} = (\frac{\partial^2 V}{\partial w_i \partial w_j})_{1\le i,j\le 3}$. Then, by **(A5)**, there exist $C_T > 0$ and $\eta_T \in \mathcal{B}(\mathbb{R}^d)$ such that (3.4) holds and

$$\begin{aligned} & (\ln N)P(\tau_N < T) - V(X_0,Y_0,Z_0) \\ \le\ & E\left[V(X_{T\wedge\tau_N}, Y_{T\wedge\tau_N}, Z_{T\wedge\tau_N})\right] - V(X_0,Y_0,Z_0) \\ =\ & E\left[\int_0^{T\wedge\tau_N} \mathcal{L}V(s,X_s,Y_s,Z_s)ds\right] \\ =\ & -E\left[\int_0^{T\wedge\tau_N} \left\{\frac{b_1(s,X_s,Y_s,Z_s)}{X_s} + \frac{b_2(s,X_s,Y_s,Z_s)}{Y_s} + \frac{b_3(s,X_s,Y_s,Z_s)}{Z_s}\right\} ds\right] \\ & + \frac{1}{2}\sum_{j=1}^r E\left[\int_0^{T\wedge\tau_N} \left\{\frac{\sigma_{1j}^2(s,X_s,Y_s,Z_s)}{X_s^2} + \frac{\sigma_{2j}^2(s,X_s,Y_s,Z_s)}{Y_s^2} + \frac{\sigma_{3j}^2(s,X_s,Y_s,Z_s)}{Z_s^2}\right\} ds\right] \\ & + E\left[\int_0^{T\wedge\tau_N} \int_{\{|u|\le 1\}} k(s,X_s,Y_s,Z_s,u)\mu(du)ds\right] \\ \le\ & C_T T + \frac{C_T^2 T}{2} + C_T T \int_{\{|u|\le 1\}} \eta_T(u)\mu(du), \end{aligned}$$

which contradicts with (3.6). Therefore, $\tau = \infty$ a.s. and the proof is complete. □

Now we consider the extinction and persistence of diseases. Namely, we investigate whether a disease will extinct at an exponential rate or will be persistent in mean. The system (3.2) is called persistent in mean if

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds > 0 \quad a.s..$$

Theorem 3.14 *Suppose that Assumptions* **(A1)**–**(A5)** *hold. Let* (X_t, Y_t, Z_t) *be a solution to equation* (3.2) *with* $(X_0, Y_0, Z_0) \in \Delta$. *We assume that*

$$\int_0^\infty \frac{\varphi(t)}{(1+t)^2} dt < \infty, \tag{3.7}$$

where

$$\begin{aligned}\varphi(t) \;:=\; \sup_{(x,y,z)\in\Delta} \Bigg\{ & \frac{\sum_{j=1}^r \sigma_{2j}^2(t,x,y,z)}{y^2} + \int_{\{|u|\le 1\}} \left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right)\right]^2 \mu(du) \\ & + \int_{\{|u|>1\}} \left[\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\right]^2 \mu(du)\Bigg\}.\end{aligned}$$

(i) If

$$\begin{aligned}\alpha \;:=\; \limsup_{t\to\infty} \Bigg\{ & \sup_{(x,y,z)\in\Delta} \left[\frac{b_2(t,x,y,z)}{y} - \frac{\sum_{j=1}^r \sigma_{2j}^2(t,x,y,z)}{2y^2}\right] \\ & + \int_{\{|u|\le 1\}} \sup_{(x,y,z)\in\Delta} \left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right) - \frac{H_2(t,x,y,z,u)}{y}\right] \mu(du) \\ & + \int_{\{|u|>1\}} \sup_{(x,y,z)\in\Delta} \left[\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\right] \mu(du) \\ <\; & 0,\end{aligned} \tag{3.8}$$

then

$$\limsup_{t\to\infty} \frac{\ln Y_t}{t} \le \alpha \ a.s.. \tag{3.9}$$

(ii) If there exist positive constants λ_0 *and* λ *such that*

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{t}\int_0^t \Bigg\{ & \lambda_0 Y_s + \frac{b_2(s,X_s,Y_s,Z_s)}{Y_s} - \frac{\sum_{j=1}^r \sigma_{2j}^2(s,X_s,Y_s,Z_s)}{2Y_s^2} \\ & + \int_{\{|u|\le 1\}} \left[\ln\left(1+\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right) - \frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right] \mu(du) \\ & + \int_{\{|u|>1\}} \ln\left(1+\frac{G_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right) \mu(du)\Bigg\} ds \\ \ge\; & \lambda,\end{aligned} \tag{3.10}$$

then

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds \ge \frac{\lambda}{\lambda_0} \ a.s.. \tag{3.11}$$

(iii) If there exist positive constants λ_0 and λ such that

$$\begin{aligned}\liminf_{t\to\infty}\inf_{(x,y,z)\in\Delta}\Bigg\{&\lambda_0 y+\frac{b_2(t,x,y,z)}{y}-\frac{\sum_{j=1}^r\sigma_{2j}^2(t,x,y,z)}{2y^2}\\&+\int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right)-\frac{H_2(t,x,y,z,u)}{y}\right]\mu(du)\\&+\int_{\{|u|>1\}}\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\mu(du)\Bigg\}\\ \ge\ &\lambda,\end{aligned}\tag{3.12}$$

then

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t Y_s ds\ge\frac{\lambda}{\lambda_0}\ a.s..$$

Proof. (i) By Itô's formula, we get

$$\begin{aligned}\ln Y_t\ =\ &\ln Y_0+\int_0^t\left[\frac{b_2(s,X_s,Y_s,Z_s)}{Y_s}-\frac{\sum_{j=1}^r\sigma_{2j}^2(s,X_s,Y_s,Z_s)}{2Y_s^2}\right]ds\\&+\int_0^t\int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)-\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right]\mu(du)ds\\&+\int_0^t\int_{\{|u|>1\}}\ln\left(1+\frac{G_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\mu(du)ds\\&+\int_0^t\frac{\sum_{j=1}^r\sigma_{2j}(s,X_s,Y_s,Z_s)}{Y_s}dB_s^{(j)}\\&+\int_0^t\int_{\{|u|\le 1\}}\ln\left(1+\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\tilde N(ds,du)\\&+\int_0^t\int_{\{|u|>1\}}\ln\left(1+\frac{G_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\tilde N(ds,du).\end{aligned}\tag{3.13}$$

Denote the martingale part of $\ln Y_t$ by M_t. Then, by (3.13), we get

$$\begin{aligned}\langle M\rangle_t\ =\ &\int_0^t\frac{\sum_{j=1}^r\sigma_{2j}^2(s,X_s,Y_s,Z_s)}{Y_s^2}ds\\&+\int_0^t\int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\right]^2\mu(du)ds\\&+\int_0^t\int_{\{|u|>1\}}\left[\ln\left(1+\frac{G_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\right]^2\mu(du)ds.\end{aligned}$$

By (3.7) and the strong law of large numbers for martingales (see [42, Theorem 10, Chapter 2]), we get

$$\lim_{t\to\infty}\frac{M_t}{t}=0\ a.s..\tag{3.14}$$

Then, (3.9) holds by (3.8), (3.13) and (3.14).

(ii) By (3.10) and (3.13), if we take $\eta \in (0, \lambda)$ then there exists $T_\eta > 0$ such that for $t \geq T_\eta$,

$$\begin{aligned}
\ln Y_t \;\geq\; & \ln Y_0 + (\lambda - \eta)t - \lambda_0 \int_0^t Y_s ds + \int_0^t \frac{\sum_{j=1}^r \sigma_{2j}(s, X_s, Y_s, Z_s)}{Y_s} dB_t^{(j)} \\
& + \int_0^t \frac{\sum_{j=1}^r \sigma_{2j}(s, X_s, Y_s, Z_s)}{Y_s} dB_s^{(j)} \\
& + \int_0^t \int_{\{|u| \leq 1\}} \ln\left(1 + \frac{H_2(s, X_{s-}, Y_{s-}, Z_{s-}, u)}{Y_{s-}}\right) \tilde{N}(ds, du) \\
& + \int_0^t \int_{\{|u| > 1\}} \ln\left(1 + \frac{G_2(s, X_{s-}, Y_{s-}, Z_{s-}, u)}{Y_{s-}}\right) \tilde{N}(ds, du).
\end{aligned}$$

Thus, by following the argument of the proof of [36, Lemma 5.1], we can show that (3.11) holds by (3.14).

(iii) Obviously, condition (3.12) implies condition (3.10). Hence, the assertion is a direct consequence of assertion (ii). □

Remark 3.15 *If we take the following assumption of our model:*

$$b_2(t, x, y, z) = b_{2,1}(t, x, y, z) - b_{2,2}(t, x, y, z),$$

where $b_{2,i}(t, x, y, z) \geq 0$ *for any* $(t, x, y, z) \in [0, \infty) \times \Delta$, $i = 1, 2$, *then condition (3.8) can be strengthened to*

$$\begin{aligned}
\alpha^* \;:=\; & \limsup_{t \to \infty} \left\{ \sup_{(x,y,z) \in \Delta} \left[\frac{b_{2,1}^2(t, x, y, z)}{2 \sum_{j=1}^r \sigma_{2j}^2(t, x, y, z)} - \frac{b_{2,2}(t, x, y, z)}{y} \right] \right. \\
& + \int_{\{|u| \leq 1\}} \sup_{(x,y,z) \in \Delta} \left[\ln\left(1 + \frac{H_2(t, x, y, z, u)}{y}\right) - \frac{H_2(t, x, y, z, u)}{y} \right] \mu(du) \\
& \left. + \int_{\{|u| > 1\}} \sup_{(x,y,z) \in \Delta} \left[\ln\left(1 + \frac{G_2(t, x, y, z, u)}{y}\right) \right] \mu(du) \right\} \\
<\; & 0, \qquad (3.15)
\end{aligned}$$

In fact, we have $\alpha \leq \alpha^*$ *and hence condition (3.15) implies that*

$$\limsup_{t \to \infty} \frac{\ln Y_t}{t} \leq \alpha^* \quad a.s..$$

Denote by $L_+^\infty[0, \infty)$ the set of all bounded, non-negative, measurable functions on $[0, \infty)$. For $f \in L_+^\infty[0, \infty)$, define

$$\overline{f} := \sup_{t \in [0, \infty)} f(t), \qquad \underline{f} := \inf_{t \in [0, \infty)} f(t).$$

Example 3.16 *In the following examples, we let $d = 1$ and the intensity measure μ of the Poisson random measure N be given by*

$$d\mu = \mathbb{1}_{[-2,2]}(x)dx,$$

where dx is the Lebesgue measure.

(a) Let $\beta, \gamma, \xi, \sigma_1, \sigma_2, \varphi_1, \varphi_2, \varphi_3 \in L^\infty_+[0,\infty)$ and $H_1, H_2, G_1, G_2 \in L^\infty_+(-\infty,\infty)$. Define

$$\varphi(t,x,y) = \varphi_1(t)x + \varphi_2(t)y + \varphi_3(t)xy, \qquad (t,x,y) \in [0,\infty) \times \mathbb{R}^2_+.$$

We consider the system

$$\begin{cases} dX_t = -\frac{\beta(t)X_t^{\xi(t)}Y_t}{1+\varphi(t,X_t,Y_t)}dt - \frac{\sigma_1(t)X_tY_t}{1+\varphi(t,X_t,Y_t)}dB_t^{(1)} - \int_{\{|u|\le 1\}} H_1(u)X_{t-}Y_{t-}\tilde{N}(dt,du) \\ \qquad\quad - \int_{\{|u|>1\}} G_1(u)X_{t-}Y_{t-}N(dt,du), \\ dY_t = \left[\frac{\beta(t)X_t^{\xi(t)}Y_t}{1+\varphi(t,X_t,Y_t)} - \gamma(t)Y_t\right]dt + \frac{\sigma_1(t)X_tY_t}{1+\varphi(t,X_t,Y_t)}dB_t^{(1)} + \sigma_2(t)Y_tZ_tdB_t^{(2)} \\ \qquad\quad + \int_{\{|u|\le 1\}}[H_1(u)X_{t-}Y_{t-} - H_2(u)Y_{t-}Z_{t-}]\tilde{N}(dt,du) \\ \qquad\quad + \int_{\{|u|>1\}}[G_1(u)X_{t-}Y_{t-} - G_2(u)Y_{t-}Z_{t-}]N(dt,du), \\ dZ_t = \gamma(t)Y_tdt - \sigma_2(t)Y_tZ_tdB_t^{(2)} + \int_{\{|u|\le 1\}} H_2(u)Y_{t-}Z_{t-}\tilde{N}(dt,du) \\ \qquad\quad + \int_{\{|u|>1\}} G_2(u)Y_{t-}Z_{t-}N(dt,du). \end{cases} \tag{3.16}$$

Suppose that

$$\underline{\xi} \ge 1, \quad \overline{H_1}, \overline{H_2}, \overline{G_1}, \overline{G_2} < 1.$$

We have $\frac{dX_t}{dt} + \frac{dY_t}{dt} + \frac{dZ_t}{dt} = 0$. Hence, by Theorem 3.13, the system (3.16) has a unique, strong solution taking values in Δ. If

$$\overline{\beta} + 2\overline{G_1} < \underline{\gamma},$$

then by Theorem 3.14(i) and noting that $\ln(1+x) - x \le 0$ for $x > -1$, we obtain that the disease will go extinct at an exponential rate:

$$-\alpha \ge \underline{\gamma} - \overline{\beta} - 2\overline{G_1}.$$

Additionally, a vital feature of the system (3.16) to note is that the transmission function is in the form of a power function which differs from the often-seen bilinear form.

(b) Let $\beta, \gamma_1, \gamma_2, \sigma \in L^\infty_+[0,\infty)$ and $H_1, H_2, G_1, G_2 \in L^\infty_+(-\infty,\infty)$. We consider the system

$$\begin{cases} dX_t = -\beta(t)X_tY_tdt - \sigma(t)X_tY_tZ_tdB_t - \int_{\{|u|\le 1\}} H_1(u)X_{t-}Y_{t-}Z_{t-}\tilde{N}(dt,du) \\ \qquad - \int_{\{|u|>1\}} G_1(u)X_{t-}Y_{t-}Z_{t-}N(dt,du), \\ dY_t = [\beta(t)X_t - \gamma_1(t) + \gamma_2(t)Z_t]\,Y_tdt + 2\sigma(t)X_tY_tZ_tdB_t \\ \qquad + \int_{\{|u|\le 1\}}[H_1(u) - H_2(u)]X_{t-}Y_{t-}Z_{t-}\tilde{N}(dt,du) \\ \qquad + \int_{\{|u|>1\}}[G_1(u) - G_2(u)]X_{t-}Y_{t-}Z_{t-}N(dt,du), \\ dZ_t = [\gamma_1(t) - \gamma_2(t)Z_t]Y_tdt - \sigma(t)X_tY_tZ_tdB_t + \int_{\{|u|\le 1\}} H_2(u)X_{t-}Y_{t-}Z_{t-}\tilde{N}(dt,du) \\ \qquad + \int_{\{|u|>1\}} G_2(u)X_{t-}Y_{t-}Z_{t-}N(dt,du). \end{cases} \tag{3.17}$$

We have $\frac{dX_t}{dt} + \frac{dY_t}{dt} + \frac{dZ_t}{dt} = 0$. Hence, by Theorem 3.13, the system (3.17) has a unique, strong solution taking values in Δ.

Suppose that

$$\overline{H_1}, \overline{H_2}, \overline{G_1}, \overline{G_2} < 1, \quad \overline{\gamma_1} < \underline{\beta} \le \underline{\gamma_2},$$

and

$$\overline{\sigma}^2 + \overline{H_1} - \ln\{(1-\overline{H_2})(1-\overline{G_2})\} < \frac{\underline{\gamma_2} - \overline{\gamma_1}}{2}.$$

Set

$$\lambda_0 = \underline{\gamma_2}, \qquad \lambda = \underline{\gamma_2} - \overline{\gamma_1} - 2[\overline{\sigma}^2 + \overline{H_1} - \ln\{(1-\overline{H_2})(1-\overline{G_2})\}].$$

Then, condition (3.12) is satisfied. Therefore, by Theorem 3.14(iii), we obtain that the disease is persistent and

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds \ge \frac{\lambda}{\lambda_0} \quad a.s..$$

3.2 Model for population numbers

In this section, we let X_t, Y_t and Z_t denote respectively the numbers of susceptible, infected and recovered individuals at time t. We make the following assumptions.

(B1) There exists $(x_0, y_0, z_0) \in \mathbb{R}^3_+$ such that for any $T \in (0,\infty)$ and $i = 1,2,3$,

$$b_i(\cdot, x_0, y_0, z_0), \quad \sum_{j=1}^n |\sigma_{ij}(\cdot, x_0, y_0, z_0)| \in L^2[0,T], \quad \int_{\{|u|\le 1\}} |H_i(\cdot, x_0, y_0, z_0, u)|^2 \mu(du) \in L^1[0,T].$$

(B2) For any $T \in (0,\infty)$ and $N \in \mathbb{N}$, there exists $K_{N,T} \in L^1_+[0,T]$ such that

$$\begin{aligned}
&\sum_{i=1}^{3}|b_i(t,x_1,y_1,z_1)-b_i(t,x_2,y_2,z_2)|^2+\sum_{i=1}^{3}\sum_{j=1}^{r}|\sigma_{ij}(t,x_1,y_1,z_1)-\sigma_{ij}(t,x_2,y_2,z_2)|^2\\
&+\sum_{i=1}^{3}\int_{\{|u|\le 1\}}|H_i(t,x_1,y_1,z_1,u)-H_i(t,x_2,y_2,z_2,u)|^2\mu(du)\\
\le\ & K_{N,T}(t)[(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2],\\
&\forall t\in[0,T],\ (x_1,y_1,z_1),(x_2,y_2,z_2)\in\left[\frac{1}{N},N\right]^3.
\end{aligned}$$

(B3) For any $(x,y,z)\in\mathbb{R}^3_+$, $t\in(0,\infty)$ and $u\in(\mathbb{R}^l-\{0\})$,

$$\left(1+\frac{H_1(t,x,y,z,u)}{x}\right),\ \left(1+\frac{H_2(t,x,y,z,u)}{y}\right),\ \left(1+\frac{H_3(t,x,y,z,u)}{z}\right)>0,$$

and

$$\left(1+\frac{G_1(t,x,y,z,u)}{x}\right),\ \left(1+\frac{G_2(t,x,y,z,u)}{y}\right),\ \left(1+\frac{G_3(t,x,y,z,u)}{z}\right)>0.$$

(B4) For any $T\in(0,\infty)$,

$$\sup_{t\in[0,T],\,(x,y,z)\in\mathbb{R}^3_+}\left\{\frac{(x-1)b_1(t,x,y,z)}{x}+\frac{(y-1)b_2(t,x,y,z)}{y}+\frac{(z-1)b_3(t,x,y,z)}{z}\right\}<\infty,$$

$$\sum_{j=1}^{r}\sup_{t\in[0,T],\,(x,y,z)\in\mathbb{R}^3_+}\left\{\frac{|\sigma_{1j}(t,x,y,z)|}{x}+\frac{|\sigma_{2j}(t,x,y,z)|}{y}+\frac{|\sigma_{3j}(t,x,y,z)|}{z}\right\}<\infty,$$

and there exists $\eta_T\in\mathcal{B}(\mathbb{R}^l)$ such that (3.4) holds and

$$\sup_{t\in[0,T],\,(x,y,z)\in\mathbb{R}^3_+,u\in(\mathbb{R}^l-\{0\})}\frac{k(t,x,y,z,u)}{\eta_T(u)}<\infty,$$

where $k(t,x,y,z,u)$ is defined by (3.3).

Now we present the result on the existence and uniqueness of solutions to the system (3.2).

Theorem 3.17 *Suppose that Assumptions* **(B1)**–**(B4)** *hold. Then, for any given initial value* $(X_0,Y_0,Z_0)\in\mathbb{R}^3_+$, *the system (3.2) has a unique, strong solution taking values in* $\mathbb{R}^3_+$.

Proof. By **(B3)** and the interlacing technique, to complete the proof, we need only consider the case that $G_i\equiv 0$, $i=1,2,3$. Then, equation (3.2) becomes equation (3.5). By **(B1)** and **(B2)**, similar to [45, Lemma 2.1], we can show that there exists a unique local strong solution

to equation (3.5) on $[0,\tau)$, where τ is the explosion time. We will show below that $\tau=\infty$ a.s.. Define

$$\tau_N=\inf\left\{t\in[0,\tau):(X_t,Y_t,Z_t)\notin\left[\frac{1}{N},N\right]^3\right\},\qquad N\in\mathbb{N},$$

and

$$\tau_\infty=\lim_{N\to\infty}\tau_N.$$

We have that $\tau_\infty\le\tau$ so it suffices to show $\tau_\infty=\infty$ a.s.. Hence assume the contrary that there exist $\varepsilon>0$ and $T>0$ such that

$$P(\tau_\infty<T)>\varepsilon,$$

which implies that

$$P(\tau_N<T)>\varepsilon,\qquad\forall N\in\mathbb{N}. \tag{3.18}$$

Define

$$V(x,y,z)=(x-1-\ln x)+(y-1-\ln y)+(z-1-\ln z),\qquad (x,y,z)\in(0,\infty)^3.$$

By Itô's formula and **(B4)**, there exist $C_T>0$ and $\eta_T\in\mathcal{B}(\mathbb{R}^d)$ such that (3.4) holds and for $t\le\tau_N$,

$$\begin{aligned}
&E\left[V(X_{T\wedge\tau_N},Y_{T\wedge\tau_N},Z_{T\wedge\tau_N})\right]-V(X_0,Y_0,Z_0)\\
=\;&E\left[\int_0^{T\wedge\tau_N}\left\{\frac{(X_s-1)b_1(s,X_s,Y_s,Z_s)}{X_s}+\frac{(Y_s-1)b_2(s,X_s,Y_s,Z_s)}{Y_s}+\frac{(Z_s-1)b_3(s,X_s,Y_s,Z_s)}{Z_s}\right\}ds\right]\\
&+\frac{1}{2}\sum_{j=1}^{r}E\left[\int_0^{T\wedge\tau_N}\left\{\frac{\sigma_{1j}^2(s,X_s,Y_s,Z_s)}{X_s^2}+\frac{\sigma_{2j}^2(s,X_s,Y_s,Z_s)}{Y_s^2}+\frac{\sigma_{3j}^2(s,X_s,Y_s,Z_s)}{Z_s^2}\right\}ds\right]\\
&+E\left[\int_0^{T\wedge\tau_N}\int_{\{|u|\le1\}}k(s,X_s,Y_s,Z_s,u)\mu(du)ds\right]\\
\le\;&C_TT+\frac{C_T^2T}{2}+C_TT\int_{\{|u|\le1\}}\eta_T(u)\mu(du).
\end{aligned}$$

However, by (3.18), we get

$$E[V(X_{T\wedge\tau_N},Y_{T\wedge\tau_N},Z_{T\wedge\tau_N})]>\varepsilon\left[\left(\frac{1}{N}-1+\ln N\right)\wedge(N-1-\ln N)\right]\to\infty\ \text{ as } N\to\infty.$$

We have arrived at a contradiction. Therefore, $\tau=\infty$ a.s. and the proof is complete. □

Similar to Theorem 3.14, we can prove the following result on the extinction and persistence of diseases.

Theorem 3.18 *Suppose that Assumptions* **(B1)**–**(B4)** *hold. Let* (X_t,Y_t,Z_t) *be a solution to equation* (3.2) *with* $(X_0,Y_0,Z_0)\in\mathbb{R}^3_+$. *We assume that*

$$\int_0^\infty\frac{\varphi(t)}{(1+t)^2}dt<\infty,$$

where

$$\varphi(t) := \sup_{(x,y,z)\in\mathbb{R}^3_+}\left\{\frac{\sum_{j=1}^r \sigma_{2j}^2(t,x,y,z)}{y^2} + \int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right)\right]^2\mu(du) + \int_{\{|u|>1\}}\left[\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\right]^2\mu(du)\right\}.$$

(i) If

$$\begin{aligned}\alpha &:= \limsup_{t\to\infty}\left\{\sup_{(x,y,z)\in\mathbb{R}^3_+}\left[\frac{b_2(t,x,y,z)}{y} - \frac{\sum_{j=1}^r \sigma_{2j}^2(t,x,y,z)}{2y^2}\right]\right.\\ &\quad + \int_{\{|u|\le 1\}}\sup_{(x,y,z)\in\mathbb{R}^3_+}\left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right) - \frac{H_2(t,x,y,z,u)}{y}\right]\mu(du)\\ &\quad \left.+ \int_{\{|u|>1\}}\sup_{(x,y,z)\in\mathbb{R}^3_+}\left[\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\right]\mu(du)\right\}\\ &< 0, \end{aligned} \tag{3.19}$$

then

$$\limsup_{t\to\infty}\frac{\ln Y_t}{t} \le \alpha \ a.s..$$

(ii) If there exist positive constants λ_0 and λ such that

$$\begin{aligned}&\liminf_{t\to\infty}\frac{1}{t}\int_0^t\left\{\lambda_0 Y_s + \frac{b_2(s,X_s,Y_s,Z_s)}{Y_s} - \frac{\sum_{j=1}^r \sigma_{2j}^2(s,X_s,Y_s,Z_s)}{2Y_s^2}\right.\\ &\quad + \int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right) - \frac{H_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right]\mu(du)\\ &\quad \left.+ \int_{\{|u|>1\}}\ln\left(1+\frac{G_2(s,X_{s-},Y_{s-},Z_{s-},u)}{Y_{s-}}\right)\mu(du)\right\}ds\\ &\ge \lambda,\end{aligned}$$

then

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t Y_s ds \ge \frac{\lambda}{\lambda_0} \ a.s..$$

(iii) If there exist positive constants λ_0 and λ such that

$$\begin{aligned}&\liminf_{t\to\infty}\inf_{(x,y,z)\in\mathbb{R}^3_+}\left\{\lambda_0 y + \frac{b_2(t,x,y,z)}{y} - \frac{\sum_{j=1}^r \sigma_{2j}^2(t,x,y,z)}{2y^2}\right.\\ &\quad + \int_{\{|u|\le 1\}}\left[\ln\left(1+\frac{H_2(t,x,y,z,u)}{y}\right) - \frac{H_2(t,x,y,z,u)}{y}\right]\mu(du)\\ &\quad \left.+ \int_{\{|u|>1\}}\ln\left(1+\frac{G_2(t,x,y,z,u)}{y}\right)\mu(du)\right\}\\ &\ge \lambda,\end{aligned}$$

then

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t Y_s ds \geq \frac{\lambda}{\lambda_0}\ a.s..$$

Example 3.19 *Let $\Lambda, \nu, \beta, \gamma, \varepsilon, \sigma \in L^\infty_+[0,\infty)$. We consider the system*

$$\begin{cases} dX_t = [\Lambda(t) - \nu(t)X_t - \beta(t)X_tY_t]dt - \sigma(t)X_tY_tdB_t, \\ dY_t = [\beta(t)X_tY_t - (\nu(t)+\gamma(t)+\varepsilon(t))Y_t]dt + \sigma(t)X_tY_tdB_t, \\ dZ_t = [\gamma(t)Y_t - \nu(t)Z_t]dt. \end{cases} \tag{3.20}$$

Suppose that

$$\underline{\nu} > 0.$$

By (3.20), we get

$$d(X_t+Y_t+Z_t) \leq [\overline{\Lambda} - \underline{\nu}(X_t+Y_t+Z_t)]dt,$$

which implies that

$$\Gamma := \left\{(x,y,z)\in\mathbb{R}^3_+ : x+y+z \leq \frac{\overline{\Lambda}}{\underline{\nu}}\right\}$$

is an invariant set of the system (3.20). Hence, the system (3.20) has a unique, strong solution taking values in Γ by Theorem 3.17.

Define

$$\alpha^* := \sup_{x\in\left(0,\frac{\overline{\Lambda}}{\underline{\nu}}\right)}\left[\overline{\beta}x - (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) - \frac{\underline{\sigma}^2x^2}{2}\right].$$

Then, we have that

$$\begin{aligned} & \textit{condition (3.19)} \\ \Leftrightarrow\ & \alpha = \limsup_{t\to\infty}\sup_{x\in\left(0,\frac{\overline{\Lambda}}{\underline{\nu}}\right)}\left[\beta(t)x - (\nu(t)+\gamma(t)+\varepsilon(t)) - \frac{\sigma^2(t)x^2}{2}\right] < 0 \\ \Leftarrow\ & \alpha \leq \alpha^* < 0 \\ \Leftrightarrow\ & \alpha^* = \max\left\{\frac{\overline{\beta}\,\overline{\Lambda}}{\underline{\nu}} - (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) - \frac{\underline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2},\ \frac{\overline{\beta}^2}{2\underline{\sigma}^2} - (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})\right\} < 0 \\ \Leftrightarrow\ & \begin{cases} \alpha^* = \frac{\overline{\beta}\,\overline{\Lambda}}{\underline{\nu}} - (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) - \frac{\underline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2} < 0, & \text{if } \underline{\sigma}^2 \leq \frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}}, \\ \alpha^* = \frac{\overline{\beta}^2}{2\underline{\sigma}^2} - (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) < 0, & \text{if } \underline{\sigma}^2 > \frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}}. \end{cases} \end{aligned}$$

Thus, by Theorem 3.18(i), we obtain that if

$$\underline{\sigma}^2 \leq \frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}} \quad\text{and}\quad \tilde{R}_0 := \frac{\overline{\beta}\,\overline{\Lambda}}{\underline{\nu}(\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})} - \frac{\underline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2(\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})} < 1,$$

then the disease will go extinct at an exponential rate:

$$-\alpha \geq (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})\left(1-\tilde{R}_0\right);$$

if

$$\underline{\sigma}^2 > \max\left\{\frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}},\ \frac{\overline{\beta}^2}{2(\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})}\right\},$$

then the disease becomes extinct at an exponential rate:

$$-\alpha \geq (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) - \frac{\overline{\beta}^2}{2\underline{\sigma}^2}.$$

This result generalizes the results in [36, Theorem 2.1].

By (3.20), we get

$$X_t + Y_t = X_0 + Y_0 + \int_0^t [\Lambda(s) - \nu(s)X_s - (\nu(s)+\gamma(s)+\varepsilon(s))Y_s]ds.$$

Since

$$X_t + Y_t \leq \frac{\overline{\Lambda}}{\underline{\nu}}, \qquad \forall t \geq 0, \tag{3.21}$$

we get

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t [\Lambda(s) - \nu(s)X_s - (\nu(s)+\gamma(s)+\varepsilon(s))Y_s]ds = 0,$$

which implies that

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t X_s ds \geq \frac{\underline{\Lambda}}{\overline{\nu}} - \lim_{t\to\infty}\frac{\overline{\nu}+\overline{\gamma}+\overline{\varepsilon}}{\overline{\nu}t}\int_0^t Y_s ds. \tag{3.22}$$

Suppose that

$$\tilde{R}_0 := \frac{\underline{\beta}\,\underline{\Lambda}}{\overline{\nu}(\overline{\nu}+\overline{\gamma}+\overline{\varepsilon})} - \frac{\overline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2(\overline{\nu}+\overline{\gamma}+\overline{\varepsilon})} > 1.$$

Recall the quantity $b_2(s, X_s, Y_s, Z_s) = \beta(s)X_sY_s - (\nu(s)+\gamma(s)+\varepsilon(s))Y_s$. Then, by (3.20)–(3.22), we get

$$\begin{aligned}
&\liminf_{t\to\infty}\frac{1}{t}\int_0^t \left[\frac{\underline{\beta}(\overline{\nu}+\overline{\gamma}+\overline{\varepsilon})}{\overline{\nu}}\cdot Y_s + \frac{b_2(s,X_s,Y_s,Z_s)}{Y_s} - \frac{\sigma^2(s)X_s^2}{2}\right]ds \\
\geq\ & \frac{\underline{\beta}\,\underline{\Lambda}}{\overline{\nu}} - (\overline{\nu}+\overline{\gamma}+\overline{\varepsilon}) - \frac{\overline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2}.
\end{aligned}$$

Therefore, by Theorem 3.18(ii), we obtain that the disease is persistent and

$$\liminf_{t\to\infty}\frac{1}{t}\int_0^t Y_s ds \geq \frac{\overline{\nu}(\tilde{R}_0-1)}{\underline{\beta}} \quad a.s..$$

This result generalizes those in [36, Theorem 3.1].

Let $M > 0$ be a fixed constant. For $x \geq 0$, define

$$x^\star := x \wedge 1, \qquad x^\dagger := x \wedge M.$$

Example 3.20 *We revisit Example 3.16 with some changes for the population numbers model. Let $d = 1$ and the intensity measure μ of the Poisson random measure N be given by*

$$d\mu = \mathbb{1}_{[-2,2]}(x)dx.$$

(a) Let $\Lambda, \nu, \beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \xi, \sigma_1, \sigma_2, \varphi_1, \varphi_2, \varphi_3 \in L^\infty_+[0,\infty)$ and $H_1, H_2, H_3, G_1, G_2 \in L^\infty_+(-\infty,\infty)$. Define

$$\varphi(t,x,y) = \varphi_1(t)x + \varphi_2(t)y + \varphi_3(t)xy, \qquad (t,x,y) \in [0,\infty) \times \mathbb{R}^2_+.$$

We consider the system

$$\begin{cases} dX_t = \left[\Lambda(t) - \nu(t)X_t^\dagger - \frac{\beta(t)X_t^\dagger Y_t^{\dagger\xi(t)}}{1+\varphi(t,X_t,Y_t)} + \gamma_1(t)Z_t^\dagger\right]dt - \frac{\sigma_1(t)X_t^\dagger Y_t^\dagger}{1+\varphi(t,X_t,Y_t)}dB_t^{(1)} \\ \qquad - \int_{\{|u|\leq 1\}}[H_1(u)X^\star_{t-}Y^\star_{t-} - H_3(u)X^\star_{t-}Z^\star_{t-}]\tilde{N}(dt,du) - \int_{\{|u|>1\}} G_1(u)X^\star_{t-}Y^\star_{t-}N(dt,du), \\ dY_t = \left[\frac{\beta(t)X_t^\dagger Y_t^{\dagger\xi(t)}}{1+\varphi(t,X_t,Y_t)} + (\gamma_2(t) - \nu(t) - \gamma_3(t)Y_t^\dagger)Y_t^\dagger\right]dt + \frac{\sigma_1(t)X_t^\dagger Y_t^\dagger}{1+\varphi(t,X_t,Y_t)}dB_t^{(1)} + \sigma_2(t)Y_t^\dagger Z_t^\dagger dB_t^{(2)} \\ \qquad + \int_{\{|u|\leq 1\}}[H_1(u)X^\star_{t-}Y^\star_{t-} - H_2(u)Y^\star_{t-}Z^\star_{t-}]\tilde{N}(dt,du) \\ \qquad + \int_{\{|u|>1\}}[G_1(u)X^\star_{t-}Y^\star_{t-} - G_2(u)Y^\star_{t-}Z^\star_{t-}]N(dt,du), \\ dZ_t = \left[\gamma_4(t)Y_t^\dagger - (\nu(t) + \gamma_1(t))Z_t^\dagger\right]dt - \sigma_2(t)Y_t^\dagger Z_t^\dagger dB_t^{(2)} \\ \qquad + \int_{\{|u|\leq 1\}}[H_2(u)Y^\star_{t-}Z^\star_{t-} - H_3(u)X^\star_{t-}Z^\star_{t-}]\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_2(u)Y^\star_{t-}Z^\star_{t-}N(dt,du). \end{cases} \tag{3.23}$$

Suppose that

$$\underline{\xi} \geq 1, \ \overline{H_i}, \overline{G_j} < 1, \qquad i = 1,2,3, \ j = 1,2.$$

Then, Assumptions (B1)–(B4) hold. Thus, by Theorem (3.17), the system (3.23) has a unique, strong solution in $\mathbb{R}^3_+$. Assume that

$$\overline{\nu} < \underline{\gamma_2}, \qquad \overline{\sigma_1}^2 + \overline{\sigma_2}^2 + 2[\overline{H_1} - \ln\{(1-\overline{H_2})(1-\overline{G_2})\}] < 2\min\{M, \underline{\gamma_2} - \overline{\nu}\}.$$

Set

$$\lambda_0 = \overline{\gamma_3} + 1, \qquad \lambda = \min\{M, \underline{\gamma_2} - \overline{\nu}\} - \left[\frac{\overline{\sigma_1}^2 + \overline{\sigma_2}^2}{2} + \overline{H_1} - \ln\{(1-\overline{H_2})(1-\overline{G_2})\}\right].$$

Then, by Theorem (3.18)(iii), we obtain that the disease is persistent and

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds \geq \frac{\lambda}{\lambda_0} \quad a.s..$$

(b) Let $\Lambda, \nu, \beta, \gamma_1, \gamma_2, \sigma \in L^\infty_+[0,\infty)$ and $H_1, H_2, H_3, G_1, G_2 \in L^\infty_+(-\infty,\infty)$. We consider the system

$$\begin{cases} dX_t = \left[\Lambda(t) - \nu(t)X_t^\dagger - \beta(t)X_t^\dagger Y_t^\dagger + \gamma_1(t)Z_t^\dagger\right]dt - \sigma(t)X_t^\dagger Y_t^\dagger Z_t^\dagger dB_t \\ \qquad - \int_{\{|u|\le 1\}}[H_1(u) - H_3(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}\tilde{N}(dt,du) \\ \qquad - \int_{\{|u|>1\}}[G_1(u) - G_3(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}N(dt,du), \\ dY_t = \left[\beta(t)X_t^\dagger Y_t^\dagger - (\nu(t)+\gamma_2(t))Y_t^\dagger\right]dt + 2\sigma(t)X_t^\dagger Y_t^\dagger Z_t^\dagger dB_t \\ \qquad + \int_{\{|u|\le 1\}}[H_1(u) - H_2(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}\tilde{N}(dt,du) \\ \qquad + \int_{\{|u|>1\}}[G_1(u) - G_2(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}N(dt,du), \\ dZ_t = [\gamma_2(t)Y_t^\dagger - (\nu(t)+\gamma_1(t))Z_t^\dagger]dt - \sigma(t)X_t^\dagger Y_t^\dagger Z_t^\dagger dB_t \\ \qquad + \int_{\{|u|\le 1\}}[H_2(u) - H_3(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}\tilde{N}(dt,du) \\ \qquad + \int_{\{|u|>1\}}[G_2(u) - G_3(u)]X^\star_{t-}Y^\star_{t-}Z^\star_{t-}N(dt,du). \end{cases} \tag{3.24}$$

Suppose that

$$\overline{H_i}, \overline{G_j} < 1, \quad i = 1,2,3,\ j = 1,2.$$

Then, Assumptions **(B1)**–**(B4)** hold. Thus, by Theorem (3.17), the system (3.24) has a unique, strong solution in $\mathbb{R}^3_+$. If

$$\overline{\beta} + 2\overline{G_1} < \underline{\gamma_2} + \underline{\nu},$$

then the disease will extinct at an exponential rate:

$$-\alpha > \underline{\gamma_2} + \underline{\nu} - \overline{\beta} - 2\overline{G_1}.$$

3.3 Simulations

We now present simulations corresponding to Examples 3.16, 3.19 and 3.20. Simulations are completed using the Euler scheme with a time step $\Delta t = 0.001$. We include stochastic and deterministic results to demonstrate the effect of noise on such systems. The time t is epidemiological time without a specific unit; however, we may imagine the time units represent days, weeks or months.

(i) We assume that the system (3.16) in Example 3.16(a) has initial values $(X_0, Y_0, Z_0) = (0.8, 0.19, 0.01)$ and set the parameters in Table 1:

$f(t)$	$\underline{f}$	$\overline{f}$
$\beta(t) = 0.3 + 0.1\sin(4t)$	0.2	0.4
$\gamma(t) = 0.8 + 0.04\cos(7t)$	0.76	0.84
$\xi(t) = 1 + \frac{t}{1+t}$	1	2
$\varphi_i(t) = 0.01 + 0.005\cos(t), i = 1, 2$	0.005	0.015
$\varphi_3(t) = 1 + 0.5\sin(15t)$	0.5	1.5
$\sigma_1(t) = 0.5 + 0.01\cos(7t)$	0.49	0.51
$\sigma_2(t) = 0.4 + 0.01\sin(7t)$	0.39	0.41
$H_1(u) = 0.01$	—	—
$H_2(u) = 0.025$	—	—
$G_1(u) = 0.1$	—	—
$G_2(u) = 0.12$	—	—

Table 1: Parameters for simulation of the system (3.16).

In Figure 2 below, it is illustrated that the extinction of the disease occurs at an exponential rate. In accordance with Example 3.16(a), the disease will extinct with exponential rate

$$-\alpha \geq \underline{\gamma} - \overline{\beta} - 2\overline{G_1} \geq 0.16.$$

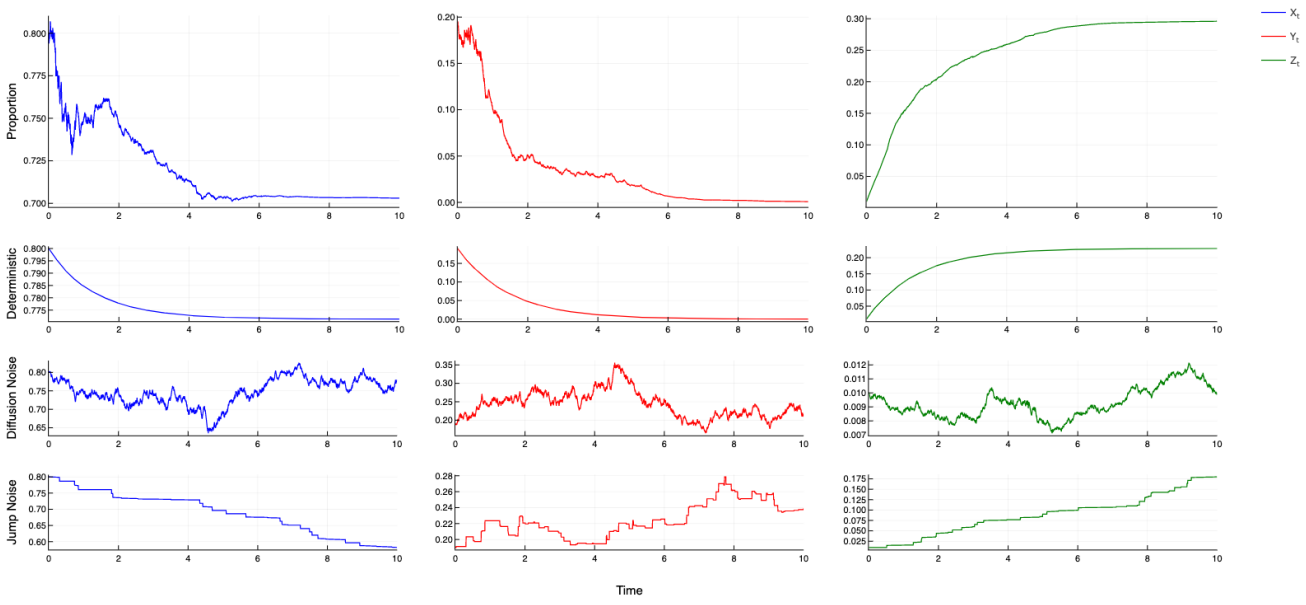

Figure 2: Row 1 gives a simulation of the stochastic simulation using the discretization scheme of the system (3.16), row 2 is the corresponding deterministic simulation, and rows 3 and 4 are the drift-free simulations of the diffusion and jump noise, respectively.

(ii) We assume that the system (3.17) in Example 3.16(b) has initial values $(X_0, Y_0, Z_0) = (0.85, 0.1, 0.05)$ and set the parameters in Table 2:

$f(t)$	$\underline{f}$	$\overline{f}$
$\beta(t) = 0.17 + 0.01\cos(20t)$	0.16	0.18
$\gamma_1(t) = 0.12 + 0.01\cos(t)$	0.11	0.13
$\gamma_2(t) = 0.56 + 0.01\sin(t)$	0.55	0.57
$\sigma(t) = 0.141 + 0.02(\sin(t) + \cos(t))$	$0.141 - 0.02\sqrt{2}$	$0.141 + 0.02\sqrt{2}$
$H_1(u) = 0.019$	—	—
$H_2(u) = 0.018$	—	—
$G_1(u) = 0.11$	—	—
$G_2(u) = 0.1$	—	—

Table 2: Parameters for simulation of system (3.17).

We achieve results that illustrate the disease's persistence, as is displayed in Figure 3 below. Furthermore, we have $\lambda_0 = 0.55$, $\lambda = 0.0776367$ and

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds \geq \frac{0.0776367}{0.55} \geq 0.14115.$$

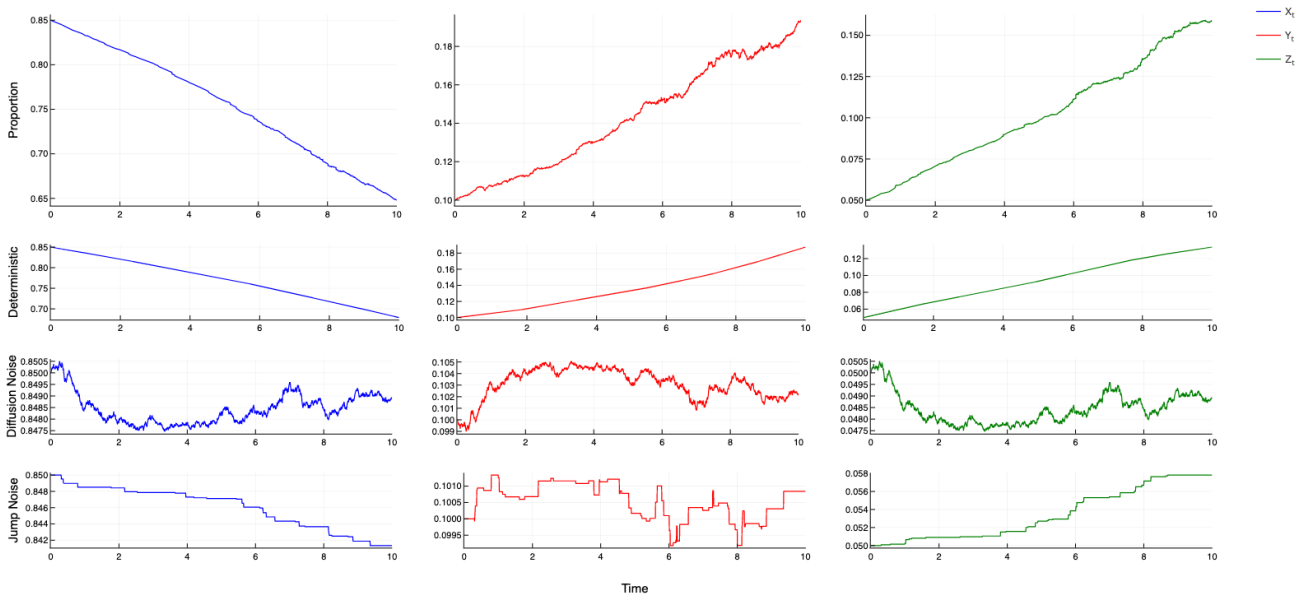

Figure 3: Row 1 gives a simulation of the stochastic simulation using the discretization scheme of the system (3.17), row 2 is the corresponding deterministic simulation, and rows 3 and 4 are the drift-free simulations of the diffusion and jump noise, respectively.

(iii) The simulations are concerned with Example 3.19, system (3.20). The initial condition is set to $(X_0, Y_0, Z_0) = (2.0, 0.8, 1)$, where the starting population is 3.8 million. In the following simulations, the parameters will change to demonstrate their effects on a system with unchanging initial conditions. The first two simulations illustrate the extinction of the disease, and the final simulation will illustrate the persistence of the disease. We initially set the parameters in Table 3:

$f(t)$	$\underline{f}$	$\overline{f}$
$\beta(t) = 0.13 + 0.01\sin(t)$	0.12	0.14
$\gamma(t) = 0.9 + 0.02\sin(t)$	0.88	0.92
$\varepsilon(t) = 0.15 + 0.07\sin(t)$	0.08	0.22
$\sigma(t) = 0.12 + 0.01(\sin(t) + \cos(t))$	$0.12 - 0.01\sqrt{2}$	$0.12 + 0.01\sqrt{2}$
$\Lambda(t) = 0.5 + 0.06\sin(t)$	0.44	0.56
$\nu(t) = 0.07 + 0.004\cos(t)$	0.066	0.074

Table 3: Parameters for simulation 1 of the system (3.20).

Since the initial condition is unchanging, it is essential to note that it forces two parameters, namely $\Lambda(t)$ and $\nu(t)$, to remain unchanged for these simulation purposes. Moreover, we have that

$$\Gamma = \left\{(x, y, z) \in \mathbb{R}^3_+ : x + y + z \leq \frac{\overline{\Lambda}}{\underline{\nu}} = 8.484848\right\}$$

as the invariant set for the system (3.20), that is, this system has a unique, strong solution taking values in Γ per Theorem 3.17. Given these parameters and following Example 3.19, we have

$$\tilde{R}_0 = \frac{\overline{\beta}\,\overline{\Lambda}}{\underline{\nu}(\underline{\nu} + \underline{\gamma} + \underline{\varepsilon})} - \frac{\underline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2(\underline{\nu} + \underline{\gamma} + \underline{\varepsilon})} \leq 0.7646 < 1, \qquad \underline{\sigma}^2 < 0.0121 < 0.0165 = \frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}}.$$

As demonstrated below in Figure 4, the disease will go extinct at an exponential rate:

$$-\alpha \geq (\underline{\nu} + \underline{\gamma} + \underline{\varepsilon})\left(1 - \tilde{R}_0\right) \geq 0.241.$$

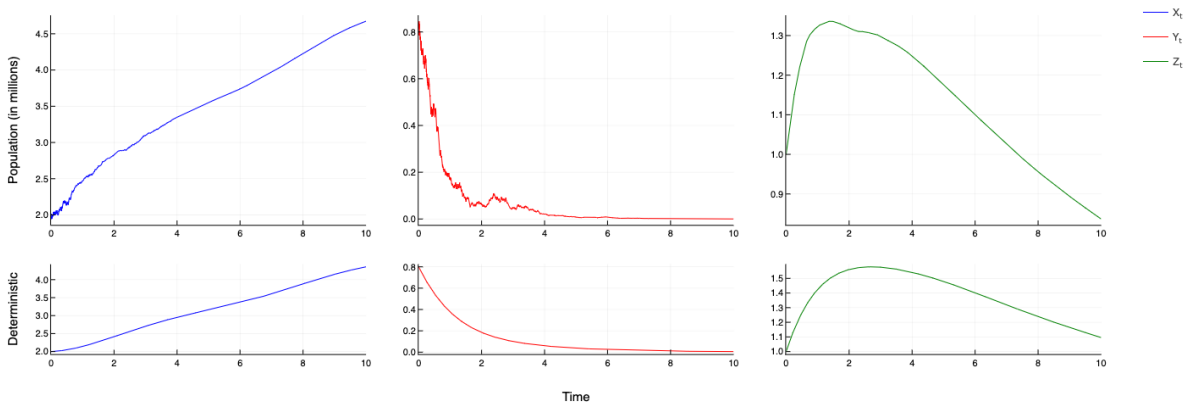

Figure 4: Row 1 is the discretization scheme simulation 1 of the system (3.20) illustrating disease extinction, and row 2 is the corresponding deterministic system.

We now only alter a single parameter in the system (3.20). Assume that $\sigma(t)$ has the form given in Table 4:

$f(t)$	$\underline{f}$	$\overline{f}$
$\sigma(t) = 0.55 + 0.003(\sin(t) + \cos(t))$	$0.55 - 0.003\sqrt{2}$	$0.55 + 0.003\sqrt{2}$

Table 4: Parameters for simulation 2 of the system (3.20).

This alteration yields

$$\underline{\sigma}^2 \geq 0.29 > 0.0165 \geq \max\left\{\frac{\underline{\nu}\overline{\beta}}{\overline{\Lambda}},\ \frac{\overline{\beta}^2}{2(\underline{\nu}+\underline{\gamma}+\underline{\varepsilon})}\right\}.$$

Thus, we have a scenario in which the disease goes extinct at an exponential rate:

$$-\alpha \geq (\underline{\nu}+\underline{\gamma}+\underline{\varepsilon}) - \frac{\overline{\beta}^2}{2\underline{\sigma}^2} \geq 0.993.$$

Moreover, if we compare Figure 5 to the above Figure 4, although initially, the disease dynamics appear more volatile, we notice the disease appears to go extinct at a faster rate which is as expected given the above results.

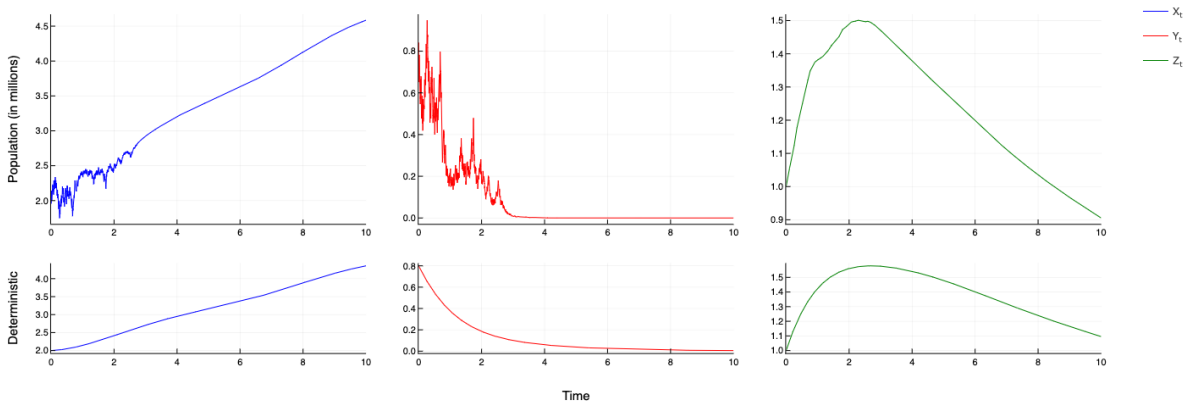

Figure 5: Row 1 is the discretization scheme simulation 2 of the system (3.20) illustrating disease extinction, and row 2 is the corresponding deterministic system.

Now assume that the parameters for the system (3.20) are given in Table 5:

$f(t)$	$\underline{f}$	$\overline{f}$
$\beta(t) = 0.56 + 0.01\sin(4t)$	0.55	0.57
$\gamma(t) = 0.25 + 0.1\cos(5t)$	0.15	0.35
$\sigma(t) = 0.24 + 0.01(\sin(t) + \cos(t))$	$0.24 - 0.01\sqrt{2}$	$0.24 + 0.01\sqrt{2}$

Table 5: Parameters for simulation 3 of system (3.20).

This modification yields

$$\tilde{R}_0 = \frac{\underline{\beta}\,\underline{\Lambda}}{\overline{\nu}(\overline{\nu}+\overline{\gamma}+\overline{\varepsilon})} - \frac{\overline{\sigma}^2\overline{\Lambda}^2}{2\underline{\nu}^2(\overline{\nu}+\overline{\gamma}+\overline{\varepsilon})} \geq 1.7 > 1.$$

In Figure 6 below, we see such a modification yields disease persistence as opposed to disease extinction achieved in the previous two simulations for the system (3.20).

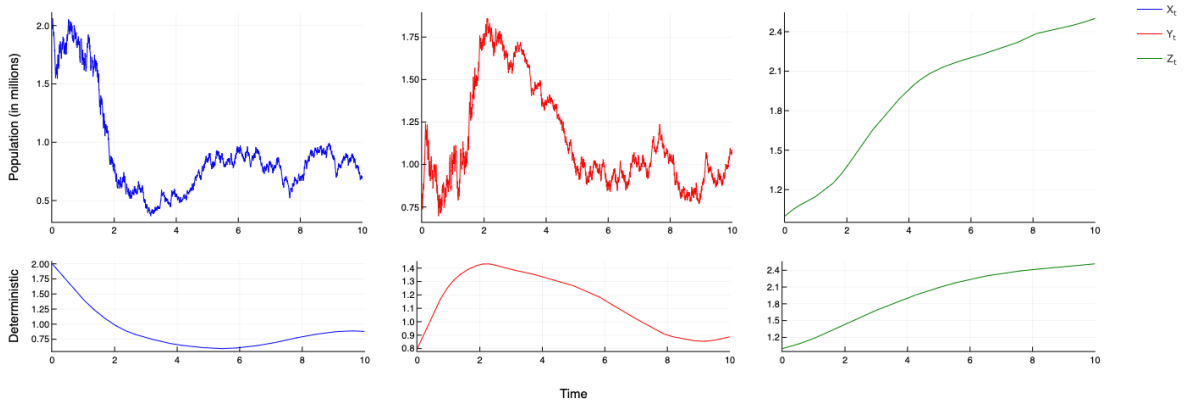

Figure 6: Row 1 is the discretization scheme simulation 3 of the system (3.20) illustrating disease persistence, and row 2 is the corresponding deterministic system.

(iv) We assume the system (3.23) has initial conditions $(3.75, 1.15, 1.1)$, where the values are taken to be in millions. Set the parameters in Table 6:

$f(t)$	$\underline{f}$	$\overline{f}$
$M = 2$	—	—
$\Lambda(t) = 0.15 + 0.006\sin(t)$	0.144	0.156
$\nu(t) = 0.002 + 0.0001\cos(t)$	0.0019	0.0021
$\beta(t) = 0.18 + 0.01\sin(2t)$	0.17	0.19
$\gamma_1(t) = 0.15 + 0.004\cos(t)$	0.146	0.154
$\gamma_2(t) = 0.12 + 0.02\cos(t)$	0.1	0.14
$\gamma_3(t) = 0.12 + 0.04\cos(2t)$	0.08	0.16
$\gamma_4(t) = 0.1 + 0.04\sin(4t)$	0.06	0.14
$\xi(t) = 1 + \ln(1 + \lvert\sin(t)\rvert)$	1	$1 + \ln 2$
$\varphi_i(t) = 0.01 + 0.005\cos(t), i = 1, 2$	0.005	0.015
$\varphi_3(t) = 1 + 0.25\sin(15t)$	0.75	1.25
$\sigma_1(t) = 0.15 + 0.01\cos(t)$	0.14	0.16
$\sigma_2(t) = 0.12 + 0.01\sin(t)$	0.11	0.13
$H_1(u) = 0.0001$	—	—
$H_2(u) = 0.00025$	—	—
$H_3(u) = 0.0009$	—	—
$G_1(u) = 0.001$	—	—
$G_2(u) = 0.0012$	—	—

Table 6: Parameters for simulation of system (3.23).

We have $\lambda_0 = 1.16$ and $\lambda \geq 0.075$ and

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t Y_s ds \geq \frac{0.075}{1.16} \geq 0.064.$$

The resulting persistence of the disease is illustrated below in Figure 7.

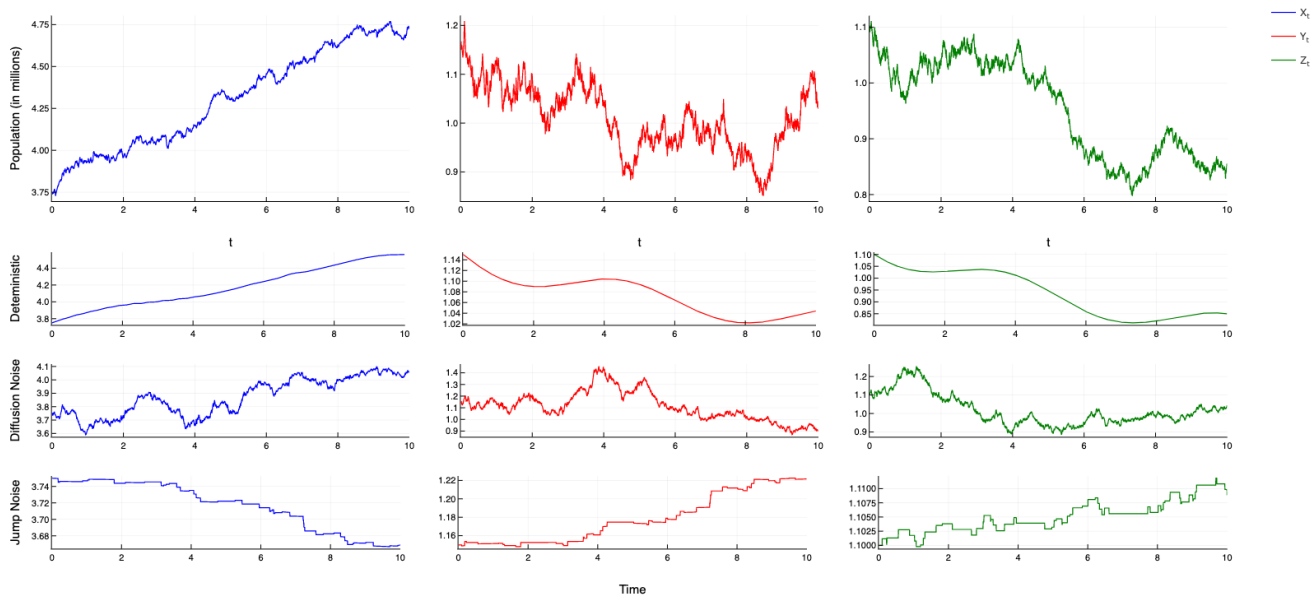

Figure 7: Row 1 gives discretization scheme simulation of the system (3.23) illustrating disease persistence, and subsequent rows give corresponding deterministic, diffusion noise and jump noise simulations, respectively.

(v) We assume the system (3.24) has initial conditions $(7.27, 1.5, 1.11)$, where the values are taken to be in millions. Set the parameters in Table 7:

$f(t)$	$\underline{f}$	$\overline{f}$
$M = 1.5$	—	—
$\Lambda(t) = 0.09 + 0.01\cos(t)$	0.08	0.1
$\nu(t) = 0.003 + 0.001\sin(t)$	0.002	0.004
$\beta(t) = 0.14 + 0.005\cos(10t)$	0.135	0.145
$\gamma_1(t) = 0.002 + 0.002\cos(25t)$	0	0.004
$\gamma_2(t) = 0.35 + 0.04\cos(15t)$	0.31	0.39
$\sigma(t) = 0.3125 + 0.002(\sin(t) + \cos(t))$	$0.3125 - 0.002\sqrt{2}$	$0.3125 + 0.002\sqrt{2}$
$H_1(u) = 0.0001$	—	—
$H_2(u) = 0.0004$	—	—
$H_3(u) = 0.0009$	—	—
$G_1(u) = 0.001$	—	—
$G_2(u) = 0.007$	—	—
$G_3(u) = 0.005$	—	—

Table 7: Parameters for simulation of system (3.24).

The extinction of the disease is illustrated below in Figure 8. Moreover, as in Example (3.20)(b), the disease will go extinct with rate $-\alpha \geq \underline{\gamma_2} + \underline{\nu} - \overline{\beta} - 2\overline{G_1} \geq 0.165$.

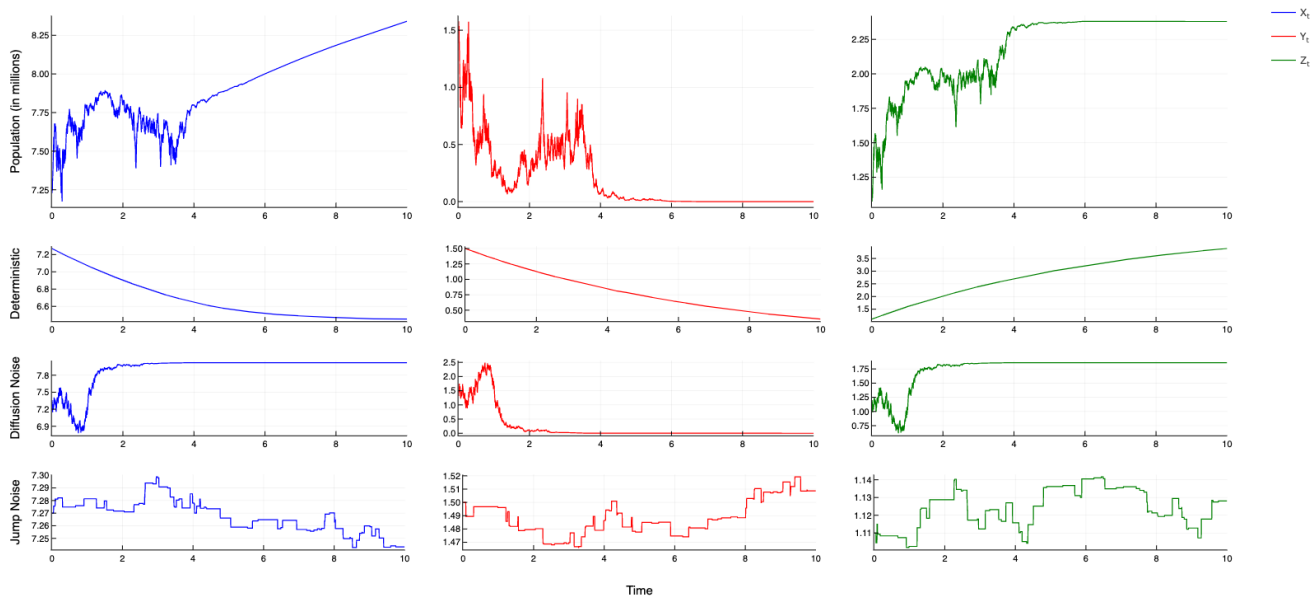

Figure 8: Row 1 gives discretization scheme simulation of the system (3.24) illustrating disease extinction, and subsequent rows give corresponding deterministic, diffusion noise and jump noise simulations, respectively.

3.4 Conclusion

In this chapter, we propose and investigate the USSIR model given by the system (3.2). We have presented two forms of the novel model–one for population proportions and the other for population numbers. We note that either form of the USSIR model could be studied in isolation and it would be possible to derive the results regarding the absent form. However, from the viewpoint of applications, it makes sense to consider both forms since differing scenarios may be more suited to utilize one form over the other.

For both forms of the model, we have given results on the extinction and persistence of diseases; moreover, we have shown that these results still hold with time-dependent, nonlinear parameters and multiple Lévy noise sources. Notably, we give examples and simulations that agree with the theoretical results and illustrate noise's impact on a given SIR model system. Moreover, the ability to allow time-dependency and multiple noises coincides with real-world occurrences of infectious disease spread due to environmental noises or time-dependent events such as temperature, climates, seasons, and so forth. The more general nature of the USSIR model allows for many tailored use cases, some of which are explored in upcoming chapters.

Chapter 4

Parameter estimation of stochastic SIR model driven by small Lévy noises with time-dependent periodic transmission

Introduction

The advantage of small noise is that we can maintain the realistic behaviour of random perturbations in observations while circumventing the necessity of placing stringent conditions on our SDEs to estimate drift parameters (e.g., no ergodicity or moment conditions are necessary). This aligns well with real-world circumstances, as we are able to consider randomness without assuming too many hypothetical properties of that randomness. Indeed, we go as far as to take the assumption we need only know minimal information regarding the noise; in fact, all we require is the noise is of the Lévy type and that there is this small dispersion ε tending towards 0. The former assumption is reasonable given the generality of Lévy processes. The latter arises naturally; for instance, considering the scenario of obtaining more and more observations of a phenomenon involving humans, from a macro point-of-view, one should not expect the intensity of the noise to become increasingly erratic. That is, with an increasing number of observations of a sizeable population, time and time again, we have found detectable patterns despite the presence of random perturbations.

This chapter, based upon [22, E. and Sun], is concerned with the parameter estimation of small noise Lévy-driven SDEs as applied to the USSIR model. Recall the USSIR model (3.2) is defined as

$$\begin{cases} dX_t = b_1(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r}\sigma_{1j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_1(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_1(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dY_t = b_2(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r}\sigma_{2j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_2(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_2(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du), \\ dZ_t = b_3(t,X_t,Y_t,Z_t)dt + \sum_{j=1}^{r}\sigma_{3j}(t,X_t,Y_t,Z_t)dB_t^{(j)} \\ \qquad + \int_{\{|u|\le 1\}} H_3(t,X_{t-},Y_{t-},Z_{t-},u)\tilde{N}(dt,du) + \int_{\{|u|>1\}} G_3(t,X_{t-},Y_{t-},Z_{t-},u)N(dt,du). \end{cases} \tag{4.1}$$

It is a natural question to ask if disease transmission is periodic, and understanding this periodicity dramatically aids in predicting possible outcomes. The specific application is the study of periodicity in the COVID-19 pandemic, as well as future pandemics. However, similar to the previous chapter, it is now subjected to a small dispersion coefficient ε.

In our study, we lay a solid theoretical foundation upon which we can estimate parameters and forecast future disease dynamics. Additionally, given the complexity of the model, optimization techniques are harnessed to solve for approximations to the estimators iteratively. The methodology presented herein is directly inspired by the results of Long et al. in [45] and [46]. We also use the least-squares method to estimate the true value of parameter θ_0 of a discretely observed stochastic process. We introduce a contrast function from which we are able to derive the least-squares estimators (LSEs) when the drift is time-dependent and the explicit expression of the noise is unknown–also given are proofs on the consistency and results on the rate of convergence of the estimators.

Let $(B_t)_{t\geq 0} = (B_t^{(1)}, \ldots, B_t^{(r)})_{t\geq 0}$ be a standard r-dimensional Brownian motion and N a Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}^l - \{0\})$ with intensity measure μ satisfying $\int_{\mathbb{R}^l - \{0\}}(1 \wedge |u|^2)\mu(du) < \infty$. We assume that $(B_t)_{t\geq 0}$ and N are independent. Define $\tilde{N}(dt, du) = N(dt, du) - \mu(du)dt$. Consider a stochastic process $(S_t^\varepsilon)_{t\geq 0}$ which satisfies the SDE:

$$\begin{aligned} dS_t^\varepsilon &= b(t, S_t^\varepsilon, \theta)dt \\ &\quad + \varepsilon \left\{ \sigma(t, S_t^\varepsilon)dB_t + \int_{\{|u|\leq 1\}} H(t, S_{t-}^\varepsilon, u)\widetilde{N}(dt, du) + \int_{\{|u|>1\}} G(t, S_{t-}^\varepsilon, u)N(dt, du) \right\}, \end{aligned} \tag{4.2}$$

where $0 < \varepsilon < 1$, $t \in [0, 1]$, $S_0^\varepsilon = s \in \mathbb{R}^d$, $\theta \in \overline{\Theta}$, the closure of an open convex bounded subset Θ of $\mathbb{R}^p$, $b(\cdot, \cdot, \cdot) : [0, \infty) \times \mathbb{R}^d \times \Theta \to \mathbb{R}^d$, $\sigma(\cdot, \cdot) : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^{d\times r}$, $H(\cdot, \cdot, \cdot)$, $G(\cdot, \cdot, \cdot) : [0, \infty) \times \mathbb{R}^d \times (\mathbb{R}^l - \{0\}) \to \mathbb{R}^d$ are Borel measurable functions. Suppose $(S_t^\varepsilon)_{t\geq 0}$ is observed at regularly spaced time points $\{t_k = \frac{k}{n},\ k = 1, 2, \ldots, n\}$. For a matrix M, we denote its transpose by M^*. Define the contrast function

$$\Psi_{n,\varepsilon}(\theta) = n \sum_{k=1}^{n} P_k^*(\theta) P_k(\theta),$$

where

$$P_k(\theta) = S_{t_k}^\varepsilon - S_{t_{k-1}}^\varepsilon - \frac{1}{n} b(t_{k-1}, S_{t_{k-1}}^\varepsilon, \theta).$$

Let $\hat{\theta}_{n,\varepsilon}$ be a minimum contrast estimator, i.e., a random variable satisfying

$$\hat{\theta}_{n,\varepsilon} := \arg\min_{\theta\in\Theta} \Psi_{n,\varepsilon}(\theta).$$

As will be seen in the following pages, given the complexity of the USSIR model, finding a closed form for the LSE $\hat{\theta}_{n,\varepsilon}$ can be quite difficult; thus, we look for suitable approximations $\hat{\theta}^*_{n,\varepsilon}$ of $\hat{\theta}_{n,\varepsilon}$. In the vernacular for optimization, we may refer to our contrast function $\Psi_{n,\varepsilon}(\theta)$ as an objective function. Recall the discussion of estimation as optimization in Chapter 2; namely, given an objective function that one wishes to minimize for parameter estimation purposes, a commonly utilized method is gradient descent (GD) (cf. [7, Chapter 8]) and in our case projected gradient descent (PGD) (cf. [7, Chapter 10]).

4.1 Least-squares estimators for time-dependent SDEs driven by small Lévy noises

4.1.1 General time-dependent SDEs driven by small Lévy noises

In this subsection, we investigate LSEs for discretely observed stochastic processes driven by small Lévy noises. The results presented here generalize the results of Long et al. in [45] and [46]. Our contributions are the generalization to a time-dependent jump-diffusion model, which is more general to include different (singular and non-singular) coefficient functions for the Brownian motion, small jump and large jump portions. Also noteworthy, although we use a contrast function which is different from that in [46], we make mention of a result extending the asymptotics of their contrast function in the presence of time-dependency and non-singular noise. These results validate our use case, and the proofs are given at the end of this chapter.

4.1.2 Statement of main results

Consider an underlying deterministic (ordinary) differential equation denoted as

$$dS_t^0 = b(t, S_t^0, \theta_0),\ t \in [0,1],\ S_0^0 = s,$$

where θ_0 is the true value of the drift parameter. Denote by $C_{\uparrow}^{1,k,l}([0,1]\times\mathbb{R}^d\times\Theta;\mathbb{R}^d)$ the class of functions $f \in C^{1,k,l}([0,1]\times\mathbb{R}^d\times\Theta;\mathbb{R}^d)$ which satisfy

$$\sup_{t\in[0,1]}\sup_{\theta\in\Theta}|\partial_\theta^{\nu_3}\partial_x^{\nu_2}\partial_t^{\nu_1} f(t,x,\theta)| \le C(1+|x|)^\lambda$$

for some constants C and λ where ν_1, ν_2, and ν_3 are non-negative integer-valued multi-indices satisfying $0 \le \nu_1 \le 1, \sum_{i=1}^d \nu_2^{(i)} \le k$ and $\sum_{i=1}^d \nu_3^{(i)} \le l$.

We take the following assumptions, which are modifications of those given in [46].

(A1) For any $\varepsilon \in (0,1)$, the SDE (4.2) admits a unique, strong solution S^ε (cf. [31, Theorem 2.2] for concrete sufficient conditions).

(A2) There exist $K > 0$ and $\eta, \xi \in \mathcal{B}_+(\mathbb{R}^l)$ such that for any $t \in [0,1]$, $x, y \in \mathbb{R}^d$, $u \in \mathbb{R}^l - \{0\}$ and $\theta \in \overline{\Theta}$,

$$|b(t,x,\theta) - b(t,y,\theta)| \le K|x-y|, \qquad \int_{\{|v|\le 1\}} \eta^2(v)\mu(dv) < \infty,$$
$$|b(t,x,\theta)| + |\sigma(t,x)| + \frac{1_{\{|u|\le 1\}}|H(t,x,u)|}{\eta(u)} + \frac{1_{\{|u|>1\}}|G(t,x,u)|}{\xi(u)} \le K(1+|x|).$$

(A3) $b(\cdot,\cdot,\cdot) \in C_{\uparrow}^{1,1,3}([0,1]\times\mathbb{R}^d\times\Theta;\mathbb{R}^d)$.

(A4) $\theta \ne \theta_0 \Leftrightarrow \exists t \in [0,1]$ such that $b(t, S_t^0, \theta) \ne b(t, S_t^0, \theta_0)$.

(A5) $\sigma(\cdot,\cdot)$ is continuous on $[0,1]\times\mathbb{R}^d$ and $H(\cdot,\cdot,u)$, $G(\cdot,\cdot,u)$ are continuous on $[0,1]\times\mathbb{R}^d$ for any $u \in \mathbb{R}^l - \{0\}$.

(A6) $\varepsilon = \varepsilon_n \to 0$ and $n\varepsilon \to \infty$ as $n \to \infty$.

Due to the nature of the work herein obtaining a closed-form of an estimator $\hat{\theta}_{n,\varepsilon}$ is not always possible hence it makes sense to consider approximations of these estimators. We denote these approximations by $\hat{\theta}^*_{n,\varepsilon}$.

Recall the notation $o_p(1)$, which is short for a sequence of random vectors that converges to zero in probability, and the notation $\xrightarrow{P_{\theta_0}}$ which is short for convergence in probability under P_{θ_0}. By virtue of [61, Theorem 5.7], similar to [45, 46, Theorem 2.1], we can prove the following result on the consistency of the LSEs.

Theorem 4.21 *Let $\hat{\theta}^*_{n,\varepsilon}$ be any sequence of estimators with $\Psi_{n,\varepsilon}(\hat{\theta}^*_{n,\varepsilon}) \leq \Psi_{n,\varepsilon}(\theta_0) + o_p(1)$. Then, under conditions (A1)-(A4), we have*

$$\hat{\theta}^*_{n,\varepsilon} \xrightarrow{P_{\theta_0}} \theta_0 \text{ as } \varepsilon \to 0 \text{ and } n \to \infty.$$

Define the matrix $I(\theta) = (I^{ij}(\theta))_{1\leq i,j\leq p}$ by

$$I^{ij}(\theta) = \int_0^1 (\partial_{\theta_i} b)^*(r, S^0_r, \theta)\partial_{\theta_j} b(r, S^0_r, \theta)dr.$$

Similar to [45, 46, Theorem 2.2], we can prove the following result on the rate of convergence of the LSEs.

Theorem 4.22 *Assume that conditions (A1)-(A6) hold and $I(\theta_0)$ is positive definite. Then,*

$$\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) \xrightarrow{P_{\theta_0}} I^{-1}(\theta_0)\left(\int_0^1 (\partial_{\theta_i} b)^*(r, S^0_r, \theta)\Big\{\sigma(r, S^0_r)dB_r + \int_{\{|u|\leq 1\}} H(r, S^0_r, u)\widetilde{N}(dr, du) + \int_{\{|u|>1\}} G(r, S^0_r, u)N(dr, du)\Big\}\right)^*_{1\leq i\leq p}$$

as $\varepsilon \to 0$ and $n \to \infty$.

Remark 4.23 *Consider the following SDE:*

$$dS^\varepsilon_t = b(t, S^\varepsilon_t, \theta)dt + \varepsilon\sigma(t, S^\varepsilon_t)\left\{dB_t + \int_{\{|u|\leq 1\}} u\widetilde{N}(dt, du) + \int_{\{|u|>1\}} uN(dt, du)\right\}.$$

In the event the diffusion matrix $\sigma\sigma^*$ is invertible, we may use the following contrast function from Long et al. [46]:

$$\Psi_{n,\varepsilon}(\theta) = n\left(\sum_{k=1}^n P^*_k(\theta)\Lambda^{-1}_{k-1}P_k(\theta)\right)1_{\{D>0\}}, \tag{4.3}$$

where

$$P_k(\theta) = S^{\varepsilon}_{t_k} - S^{\varepsilon}_{t_{k-1}} - \frac{1}{n} b(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta), \quad \Lambda_{k-1} = [\sigma\sigma^*](t_{k-1}, S_{t_{k-1}}), \quad D = \inf_{k=0,\dots,n-1} \det \Lambda_k.$$

Define the matrix $I(\theta) = (I^{ij}(\theta))_{1\le i,j\le p}$ *by*

$$I^{ij}(\theta) = \int_0^1 (\partial_{\theta_i} b)^*(r, S^0_r, \theta)[\sigma\sigma^*]^{-1}(r, S^0_r)\partial_{\theta_j} b(r, S^0_r, \theta) dr. \tag{4.4}$$

We make the following additional assumption.

(A7) *There exists an open convex subset* $\mathcal{U} \subset \mathbb{R}^d$ *such that* $S^0_t \in \mathcal{U}$ *for all* $t \in [0,1]$, σ *is smooth on* $[0,1] \times \mathcal{U}$, *and* $\sigma\sigma^*$ *is invertible on* $[0,1] \times \mathcal{U}$.

Following the arguments of [45, 46, Theorems 2.1 and 2.2], we can prove the following result.

Corollary 4.24 *Assume that conditions (A1)-(A7) hold and* $I(\theta_0)$ *defined by (4.4) is positive definite. Then, the assertions of Theorems 4.21 and 4.22 hold for the LSE derived from the contrast function (4.3).*

4.1.3 Application to USSIR model with periodic transmission

Setting $d = 3$, we use equation (4.2) to write the USSIR model as

$$\begin{aligned}\begin{bmatrix} dX^{\varepsilon}_t \\ dY^{\varepsilon}_t \\ dZ^{\varepsilon}_t \end{bmatrix} &= \begin{bmatrix} b_1(t, X^{\varepsilon}_t, Y^{\varepsilon}_t, Z^{\varepsilon}_t, \theta) \\ b_2(t, X^{\varepsilon}_t, Y^{\varepsilon}_t, Z^{\varepsilon}_t, \theta) \\ b_3(t, X^{\varepsilon}_t, Y^{\varepsilon}_t, Z^{\varepsilon}_t, \theta) \end{bmatrix} dt + \varepsilon \Big\{ \sigma(t, X^{\varepsilon}_t, Y^{\varepsilon}_t, Z^{\varepsilon}_t) dB_t \\ &\quad + \int_{\{|u|\le 1\}} H(t, X^{\varepsilon}_{t-}, Y^{\varepsilon}_{t-}, Z^{\varepsilon}_{t-}, u)\widetilde{N}(dt, du) \\ &\quad + \int_{\{|u|>1\}} G(t, X^{\varepsilon}_{t-}, Y^{\varepsilon}_{t-}, Z^{\varepsilon}_{t-}, u) N(dt, du) \Big\},\end{aligned} \tag{4.5}$$

such that all previously stated assumptions hold. Above in equation (4.5), the drift function $b = (b_1, b_2, b_3)$ is given in more general terms but in the subsequent sections, explicit instances will be given. Moreover, at the centre of our focus in this chapter is the presence of a periodic transmission function. Since any well-behaved periodic function may be approximated using a Fourier series; we consider the drift function to contain a periodic transmission function of the form

$$\beta(t) = \alpha_0 + \sum_{k=1}^{K} \alpha_{1,k} \cos\left(\frac{2\pi k t}{\theta}\right) + \alpha_{2,k} \sin\left(\frac{2\pi k t}{\theta}\right), \tag{4.6}$$

where $\theta, \alpha_0, \alpha_{1,k}, \alpha_{2,k} > 0, 1 \le k \le K$. This leads to the contrast function being denoted by $\Psi_{n,\varepsilon}(\theta, \alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K)$ so that our parameters are represented by the vector $(\theta, \alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K)$

$\in \mathbb{R}_+^{2K+2}$. Moreover, assume that $\theta \in [0,1]$, a natural assumption to make given the estimation happens from time $t=0$ to $t=1$. The contrast function $\Psi_{n,\varepsilon}(\theta, \alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K)$ is not globally convex; however, for a fixed θ, it becomes convex in the remaining parameters. Hence, we introduce the following algorithm:

Algorithm 1 Linear-Search Gradient-Descent (LS-GD)

For i in 1 to M,

Fix a test value of θ: $\theta_i \in \left(\frac{i-1}{M}, \frac{i}{M}\right)$;

Set the initial value $(\alpha_0^{(0)}, (\alpha_{1,k}^{(0)}, \alpha_{2,k}^{(0)})_{k=1}^K)$;

Run Gradient Descent on the function $\Psi_{n,\varepsilon}(\theta_i, \alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K)$ with update rule $(\alpha_0^{(l)}, (\alpha_{1,k}^{(l+1)}, \alpha_{2,k}^{(l+1)})_{k=1}^K) = (\alpha_0^{(l)}, (\alpha_{1,k}^{(l)}, \alpha_{2,k}^{(l)})_{k=1}^K) - \eta \nabla_{\{\alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K\}} \Psi_{n,\varepsilon}(\theta_i, \alpha_0^{(l)}, (\alpha_{1,k}^{(l)}, \alpha_{2,k}^{(l)})_{k=1}^K)$;

Store $\Psi_i^* := \underset{\{\alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K\}}{\arg\min} \Psi_{n,\varepsilon}(\theta_i, \alpha_0, (\alpha_{1,k}, \alpha_{2,k})_{k=1}^K)$.

End

Return $\min_{i \in \{1,\ldots,M\}} \Psi_i^*$.

The returned value $\min_{i \in \{1,\ldots,M\}} \Psi_i^*$ will be the approximation $(\theta^*, \alpha_0^*, (\alpha_{1,k}^*, \alpha_{2,k}^*)_{k=1}^K)$ of the LSE $(\hat{\theta}, \hat{\alpha}_0, (\hat{\alpha}_{1,k}, \hat{\alpha}_{2,k})_{k=1}^K)$ we seek.

Remark 4.25 *The above algorithm can be modified for PGD, which we do indeed use, as will be seen in the simulation studies.*

For the remainder, we take $K=1$, that is we consider the case when $\beta(t) = \alpha_0 + \alpha_1 \cos(2\pi t/\theta) + \alpha_2 \sin(2\pi t/\theta)$. Choosing $K > 1$ does not contribute any more insights. Additionally, choosing $K > 1$ would be more computationally expensive to evaluate in our simulation studies.

Recall that one of our primary goals is the estimation of the unknown period θ and the coefficients $(\alpha_0, \alpha_1, \alpha_2)$ of the periodic transmission function $\beta(t)$. Noise ensures this is a more realistic model; moreover, we make no assumptions about the exact forms of the noise coefficient functions σ, H, G.

4.2 Simulation study of SIR model for population proportions

For this study, the unknown parameters are represented by the vector $(\theta, \alpha_0, \alpha_1, \alpha_2) \in \mathbb{R}_+^4$ and our objective function is denoted $\Psi_{n,\varepsilon}(\theta, \alpha_0, \alpha_1, \alpha_2)$, where n is the number of observations we have and $\varepsilon \in (0,1)$. For a fixed θ, recall the convexity of $\Psi_{n,\varepsilon}(\theta, \alpha_0, \alpha_1, \alpha_2)$ with respect to $(\alpha_0, \alpha_1, \alpha_2)$ one only needs to check the Hessian matrix is positive semi-definite.

As mentioned before, we use the *Julia* programming language for our simulation and numerical estimation results. Using *Julia* and utilizing the package *DifferentialEquations.jl*, we are able to generate synthetic data for our testing purposes. Namely, a discretization of the USSIR model is used, from which we take 100 observations to utilize in the estimation of the unknown parameters. We next construct the contrast function and approximate the LSE by use of the LS-GD algorithm. We accomplish the task of implementing the LS-GD algorithm with the aid of the Julia package

Optim.jl. *Optim.jl* handles the projection to the constraint set for us; namely, as we are not computing these estimates by hand, we are able to reasonably circumvent the difficulty presented by projecting to the constraint set–all that is required of us is passing the constraint set $\mathcal{C}$ as an argument to the suitable method from *Optim.jl*. The learning rate η, present in the update rule in (Algorithm 1), is chosen by a Hager-Zhang line-search algorithm. For a more in-depth exploration, we refer the reader to the *Optim.jl* documentation in [47] and also to the paper by Hager and Zhang [32].

Noteworthy is that the contrast function $\Psi_{n,\varepsilon}$ is non-linear in the unknown parameters. This is clear by the definition of $\beta(t)$. Hence, there is an inter-play or sort of dependence of optimal estimators between α_i and θ for $i = 1, 2$. Moreover, this leads to the interesting result that we are able to estimate two or more unknown parameters which are present in a dependent setting of a multiplicative form.

Remark 4.26 *Through some heuristics, we arrived at the initial value* $(\alpha_0^{(0)}, \alpha_1^{(0)}, \alpha_2^{(0)}) = (0.5, 0.31, 0.21)$*; which is used in all implementations of the LS-GD algorithm for this chapter. There is nothing particularly significant about this value; hence we have no reason to believe it had any impact on our findings.*

For choosing the test values of θ_i*, we chose them such that* $\theta_i \sim$ *Uniform*$(\frac{i-1}{M}, \frac{i}{M})$*. As the value* M *increases, it does not make a difference if we individually choose the value or allow it to be randomly assigned within each subinterval.*

4.2.1 Model for population proportions

Consider the following model for population proportions:

$$\begin{bmatrix} dX_t \\ dY_t \\ dZ_t \end{bmatrix} = \begin{bmatrix} -\beta(t) X_t Y_t \\ \beta(t) X_t Y_t - \gamma Y_t \\ \gamma Y_t \end{bmatrix} dt + \varepsilon \begin{bmatrix} -\sigma X_{t-} Y_{t-} Z_{t-} \\ 2\sigma X_{t-} Y_{t-} Z_{t-} \\ -\sigma X_{t-} Y_{t-} Z_{t-} \end{bmatrix} dL_t, \tag{4.7}$$

where

$$L_t = B_t + \int_0^t \int_{\{|u|>0.1\}} u N(ds, du),$$

and as before,

$$\beta(t) = \alpha_0 + \alpha_1 \cos\left(\frac{2\pi t}{\theta}\right) + \alpha_2 \sin\left(\frac{2\pi t}{\theta}\right), \quad \theta \in [0,1], \alpha_0, \alpha_1, \alpha_2 > 0,$$

$\gamma, \sigma > 0$ are constants, and $N = N_1 + N_2$ such that N_1, N_2 are independent Poisson random measures with respective intensity measures μ_1 and μ_2:

$$\begin{cases} N_1 \sim \frac{2}{3}\lambda dt \mu_1(du) \text{ with } \mu_1 = \delta(\{-0.1, 0.1, 0.0\}), \\ N_2 \sim \frac{1}{3}\lambda dt \mu_2(du) \text{ with } \mu_2 = \delta(\{0.0, -0.1, 0.1\}), \end{cases} \tag{4.8}$$

where $\lambda \sim \text{Uniform}(\{1,2,3,4\})$.

By [21, Theorem 2.1], the SDE (4.7) has a unique, strong solution taking values in the set $\Delta = \{(x,y,z) \in \mathbb{R}^3_+ : x+y+z=1\}$. It is easy to verify that all assumptions of Theorems 2.1 and 2.2 hold. We use the following contrast function:

$$\Psi_{\varepsilon,n}(\theta,\alpha_0,\alpha_1,\alpha_2) = 100\left(\sum_{k=1}^{100} P_k^*(\theta,\alpha_0,\alpha_1,\alpha_2)P_k(\theta,\alpha_0,\alpha_1,\alpha_2)\right),$$

where

$$P_k(\theta,\alpha_0,\alpha_1,\alpha_2) = \begin{bmatrix} X_{t_k} - X_{t_{k-1}} - \frac{1}{100}\left[-\beta(t_{k-1})X_{t_{k-1}}Y_{t_{k-1}}\right] \\ Y_{t_k} - Y_{t_{k-1}} - \frac{1}{100}\left[\beta(t_{k-1})X_{t_{k-1}}Y_{t_{k-1}} - \gamma Y_{t_{k-1}}\right] \\ Z_{t_k} - Z_{t_{k-1}} - \frac{1}{100}\gamma Y_{t_{k-1}} \end{bmatrix}.$$

.

4.2.2 Parameter estimation

We generate the synthetic data 10000 times and randomly choose 1000 such data sets to perform estimation upon. Set $\gamma = 0.07142$, $\sigma = 0.5$ and $(X_0,Y_0,Z_0) = (0.84,0.07,0.11)$. For the unknown parameters, the true values for use in generating the synthetic data are randomly decided upon as follows

$$\theta \sim \text{Uniform}(0,1), \quad \alpha_0 \sim \text{Uniform}(0.1,0.8), \quad \alpha_1,\alpha_2 \sim \text{Uniform}\left(0,\frac{\alpha_0}{\sqrt{2}}\right). \tag{4.9}$$

It is noteworthy that although the distributions of α_1,α_2 are dependent on the distribution of α_0, their specific values are not. The reasoning for this dependency is we require $\beta(t) \geq 0$. Also note that for each generation of synthetic data, the jump parameter λ is randomly chosen as described in (4.8).

The subsequent plots and table contain our findings, and the metric we use to test our methodology is

$$MAE = \sum_{i=1}^{1000} \frac{|y_i - \hat{y}_i|}{1000},$$

where y_i is the true value and $\hat{y}_i$ is the estimated value for $\theta,\alpha_0,\alpha_1,\alpha_2$.

ε	Parameter	MAE
0.3		
	θ	0.0980807
	α_0	0.021481
	α_1	0.055916
	α_2	0.058691
0.1		
	θ	0.0416584
	α_0	0.009259
	α_1	0.034566
	α_2	0.0368507
0.01		
	θ	0.021081
	α_0	0.005788
	α_1	0.033258
	α_2	0.0333328
0.001		
	θ	0.0205791
	α_0	0.005601
	α_1	0.03208
	α_2	0.03183

Table 1: Metrics of estimated parameters for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$.

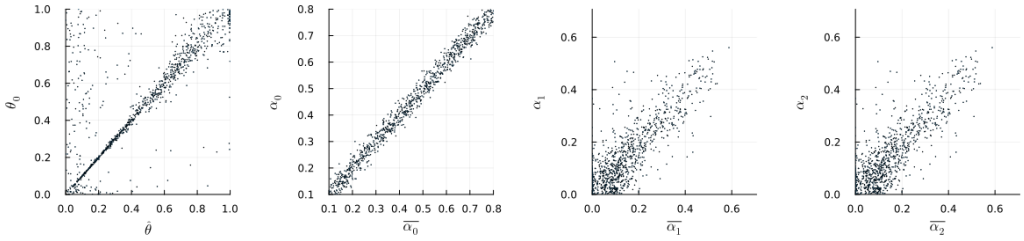

Figure 9: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.3$.

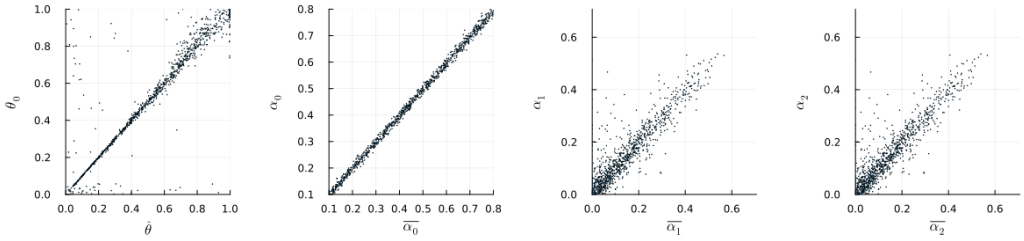

Figure 10: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.1$.

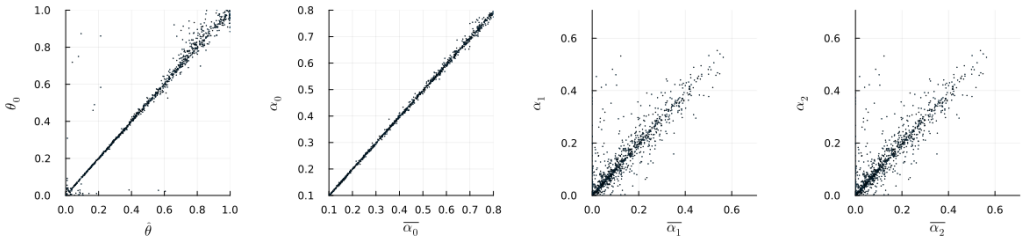

Figure 11: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.01$.

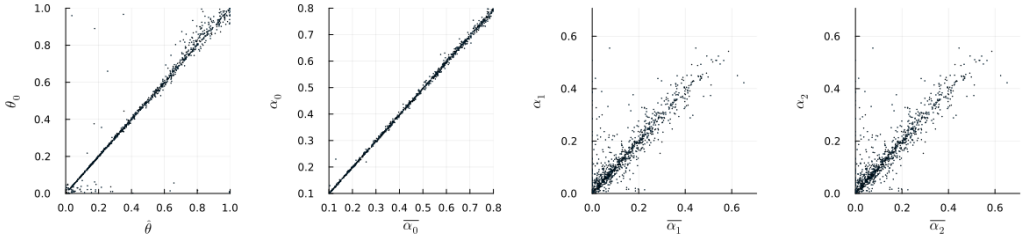

Figure 12: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.001$.

4.2.3 Forecasting

We now take the results from the previous section to complete a prediction simulation study of the model (4.7). We make use of a similar methodology and carry out our study as follows.

1. We generate a true parameter vector $(\theta, \alpha_0, \alpha_1, \alpha_2)$ and use this to create synthetic data for training.
2. We train the model using 100 observations on the time interval $[0, 1]$ by using the LS-GD algorithm with 500 test values of the period θ.
3. Once training is complete, we simulate model (4.7) on the time interval $[0, 3]$ with both the true and trained parameters vectors.

We generate the initial conditions as

$$(X_0, Y_0, Z_0) \sim \left(\text{Uniform}(0.4, 1),\ \frac{1 - X_0}{n},\ \frac{(1 - X_0)(n-1)}{n}\right),$$

where $n \sim \text{Uniform}\{2, 3, 4, 5, 6, 7\}$. The reason for this is the SIR model has the requirement that

$$X_t + Y_t + Z_t = 1, \qquad t \geq 0.$$

We use a component-wise form of mean-squared error as the metric of the trials. We run 1000 prediction trials, and for each trial, we have 100 observations/samples. Hence for each trial j, $j \in \{1, \ldots, 1000\}$, we compute the following sum.

$$T_j = \frac{1}{100} \sum_{k=1}^{100} \left[(X_{t_k} - \hat{X}_{t_k})^2,\ \ (Y_{t_k} - \hat{Y}_{t_k})^2,\ \ (Z_{t_k} - \hat{Z}_{t_k})^2 \right].$$

Then, we compute $\frac{1}{1000} \sum_{j=1}^{1000} T_j$. This metric allows us to check each component error individually while providing insight into the error of the trials of the whole model.

Remark 4.27 *There is nothing particularly special about our choice to randomly generate X_0 and then use its value to assign values to Y_0 and Z_0. Additionally, nothing special about the assignment of the factors $1/n$ and $(n-1)/n$ – merely a convenience for computation.*

The following table includes the true values against the trained (estimated) values for the proportional model.

$(\theta, \alpha_0, \alpha_1, \alpha_2)$	$(0.26836304, 0.15114833, 0.0621514, 0.096762)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.3$	$(0.27069755, 0.1707693, 0.11016354, 0.082015210)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.1$	$(0.261749799, 0.1717002782, 0.025364225, 0.1017260206)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.01$	$(0.2668796604, 0.146115022, 0.04899088, 0.0966312804)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.001$	$(0.26743667816, 0.151314344, 0.058359562, 0.0989769703)$

Table 2: True parameter values versus trained parameter values for proportional model.

The subsequent figures give simulation results using the true values and the trained values for the parameters. The true parameter observations are marked by a +, while the trained parameter observations are marked by a dot.

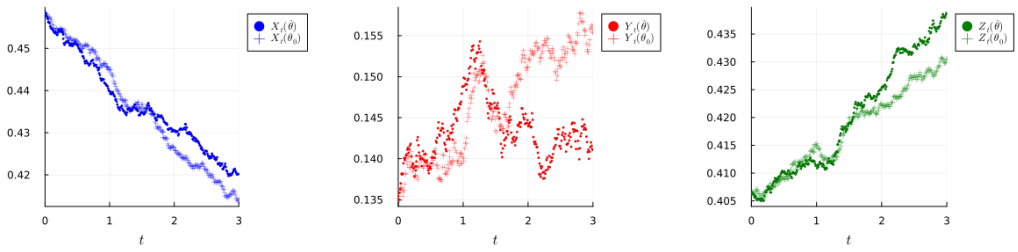

Figure 13: Simulation of model (4.7) for $\varepsilon = 0.3$.

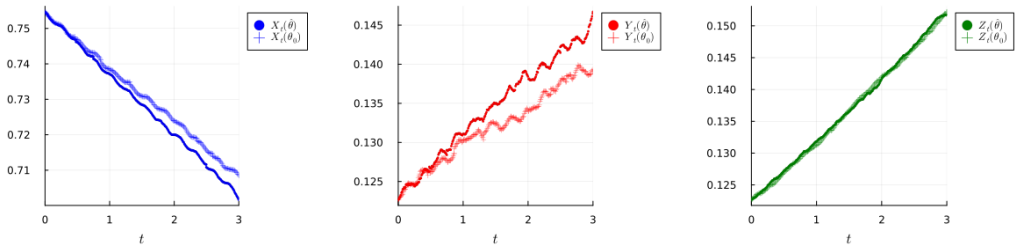

Figure 14: Simulation of model (4.7) for $\varepsilon = 0.1$.

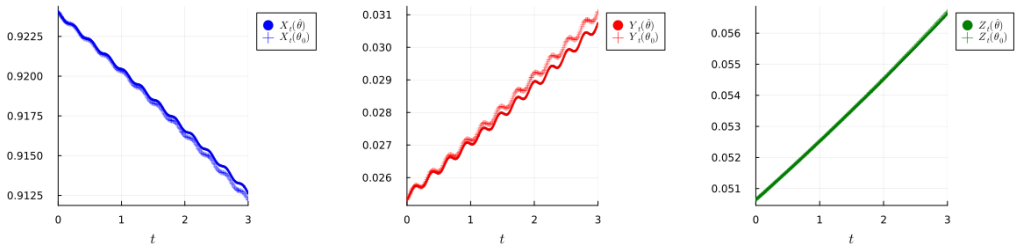

Figure 15: Simulation of model (4.7) for $\varepsilon = 0.01$.

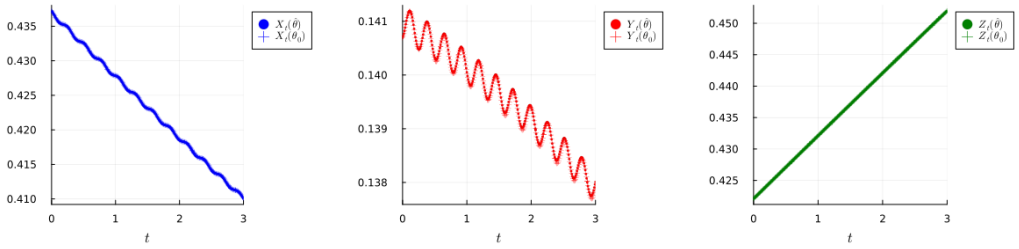

Figure 16: Simulation of model (4.7) for $\varepsilon = 0.001$.

ε	Component-wise MSE: $(\hat{X}_t, \hat{Y}_t, \hat{Z}_t)$
0.3	(0.000024732 0.00006754 0.00001291)
0.1	(0.000009734 0.00001537 0.000002157)
0.01	(0.000002564 0.000003539 0.0000006851)
0.001	(0.0000001432 0.00000003593 0.00000007774)

Table 3: True parameter values versus trained parameter values for model (4.7).

4.3 Simulation study of SIR model for population numbers

4.3.1 Model for population numbers

In this subsection, we consider the following model for population numbers:

$$\begin{bmatrix} dX_t \\ dY_t \\ dZ_t \end{bmatrix} = \begin{bmatrix} \Lambda - \nu X_t - \beta(t) X_t Y_t \\ \beta(t) X_t Y_t - (\nu + \gamma) Y_t \\ \gamma Y_t - \nu Z_t \end{bmatrix} dt + \varepsilon \begin{bmatrix} \sigma X_{t-} Y_{t-} Z_{t-} & 0 & 0 \\ 0 & \sigma X_{t-} Y_{t-} Z_{t-} & 0 \\ 0 & 0 & \sigma X_{t-} Y_{t-} Z_{t-} \end{bmatrix} \begin{bmatrix} dL_t^{(1)} \\ dL_t^{(2)} \\ dL_t^{(3)} \end{bmatrix}, \tag{4.10}$$

where

$$\begin{bmatrix} L_t^{(1)} \\ L_t^{(2)} \\ L_t^{(3)} \end{bmatrix} = \begin{bmatrix} B_t^{(1)} + \int_0^t \int_{\{|u|>0.1\}} uN^{(1)}(ds, du) \\ B_t^{(2)} + \int_0^t \int_{\{|u|>0.1\}} uN^{(2)}(ds, du) \\ B_t^{(3)} + \int_0^t \int_{\{|u|>0.1\}} uN^{(3)}(ds, du) \end{bmatrix},$$

$$\beta(t) = \alpha_0 + \alpha_1 \cos\left(\frac{2\pi t}{\theta}\right) + \alpha_2 \sin\left(\frac{2\pi t}{\theta}\right), \qquad \theta \in [0,1], \alpha_0, \alpha_1, \alpha_2 > 0,$$

$\Lambda, \nu, \gamma, \sigma > 0$ are constants, and $N^{(1)}, N^{(2)}, N^{(3)}$ are i.i.d. copies of $N = N_1 + N_2$ as given in (4.8).

By [21, Theorem 3.1], the SDE (4.10) has a unique, strong solution taking values in $\mathbb{R}^3_+$. We utilize the contrast function (4.3), which is similar to that given in [46] and takes the explicit form:

$$\Psi_{\varepsilon,n}(\theta, \alpha_0, \alpha_1, \alpha_2) = 100 \left(\sum_{k=1}^{100} P^*_{t_k}(\theta, \alpha_0, \alpha_1, \alpha_2) \Lambda^{-1}_{k-1} P_{t_k}(\theta, \alpha_0, \alpha_1, \alpha_2) \right),$$

where

$$P_{t_k}(\theta, \alpha_0, \alpha_1, \alpha_2) = \begin{bmatrix} X_{t_k} - X_{t_{k-1}} - \frac{1}{100}\left[\Lambda - \nu X_{t_{k-1}} - \beta(t_{k-1}) X_{t_{k-1}} Y_{t_{k-1}}\right] \\ Y_{t_k} - Y_{t_{k-1}} - \frac{1}{100}\left[\beta(t_{k-1}) X_{t_{k-1}} Y_{t_{k-1}} - (\nu + \gamma) Y_{t_{k-1}}\right] \\ Z_{t_k} - Z_{t_{k-1}} - \frac{1}{100}\left(\gamma Y_{t_{k-1}} - \nu Z_{t_{k-1}}\right) \end{bmatrix},$$

and

$$\Lambda_{k-1} = \begin{bmatrix} [\sigma X_{t_{k-1}} Y_{t_{k-1}} Z_{t_{k-1}}]^2 & 0 & 0 \\ 0 & [\sigma X_{t_{k-1}} Y_{t_{k-1}} Z_{t_{k-1}}]^2 & 0 \\ 0 & 0 & [\sigma X_{t_{k-1}} Y_{t_{k-1}} Z_{t_{k-1}}]^2 \end{bmatrix}.$$

4.3.2 Parameter estimation

We generate the synthetic data 10000 times and randomly choose 1000 such data sets to perform estimation upon. Set $\Lambda = 0.018, \nu = 0.00042$, $\gamma = 0.07142$, $\sigma = 0.5$ and $(X_0, Y_0, Z_0) = (2.3, 0.19, 0.25)$ (taken to be in millions). Again, for this study, we keep the jumps simple; namely, there is no compensated jump portion. We also use the same metric MAE to test our results, and the true values for the synthetic data are generated as given in (4.9).

ε	Parameter	MAE
0.3		
	θ	0.0944492
	α_0	0.023988
	α_1	0.0563110
	α_2	0.058243
0.1		
	θ	0.03130246
	α_0	0.00960035
	α_1	0.0365961
	α_2	0.0369104
0.01		
	θ	0.01966768
	α_0	0.0057932
	α_1	0.0283396
	α_2	0.0291711
0.001		
	θ	0.01657112
	α_0	0.0056
	α_1	0.02781
	α_2	0.0294392

Table 4: Metrics of estimated parameters where 1000 estimations are completed given 100 observations and we used 200 test values of $\hat{\theta}$.

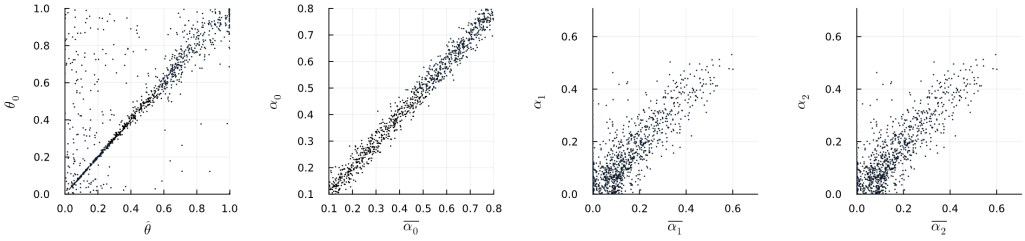

Figure 17: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.3$.

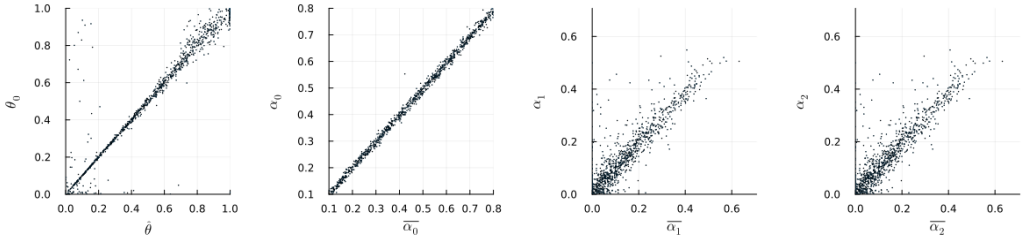

Figure 18: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.1$.

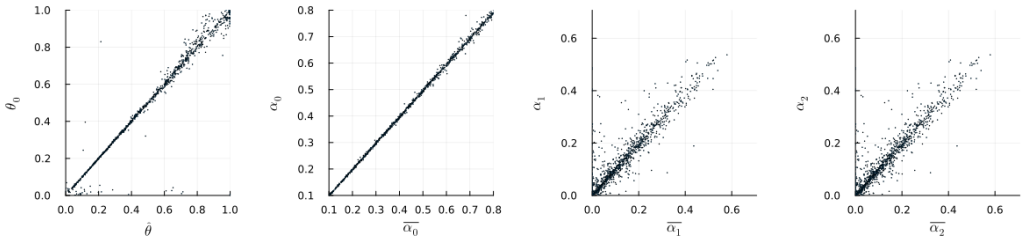

Figure 19: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.01$.

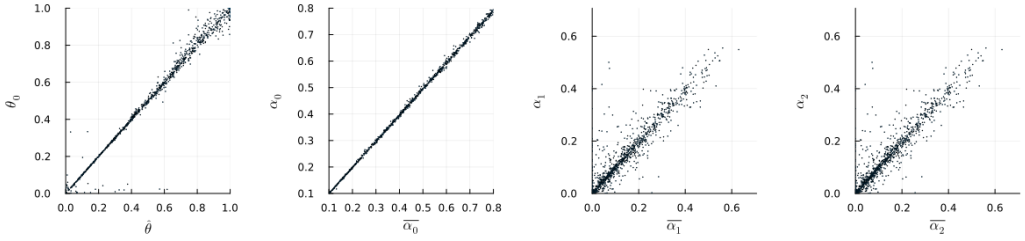

Figure 20: Estimates against true values for 1000 estimations of 100 observations using 200 values of $\hat{\theta}$ while $\varepsilon = 0.001$.

4.3.3 Forecasting

In a similar fashion to the previous subsection, we perform a predictive simulation study. The primary difference being the initial conditions are randomized such that

$$(X_0, Y_0, Z_0) \sim (\text{Uniform}(1,4),\ \text{Uniform}(0.1,1),\ \text{Uniform}(0.1,1)).$$

Otherwise, the methodology is the same as before, including using the component-wise MSE. The following table includes the randomly-generated true parameter values against those we obtained in the estimation. The subsequent figures below display the simulation of the model (4.10) using the true parameters and the values of the trained parameters. The true parameter observations are marked by a +, while the trained parameter observations are marked by a dot.

$(\theta, \alpha_0, \alpha_1, \alpha_2)$	$(0.26836304, 0.15114833, 0.0621514, 0.096762)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.3$	$(0.2759104, 0.15106501, 0.09268711, 0.0418802)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.1$	$(0.268413479, 0.14653753, 0.058398519, 0.116306536)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.01$	$(0.26945644, 0.1511415, 0.0667543397, 0.09347593)$
$(\hat{\theta}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$, $\varepsilon = 0.001$	$(0.26857489, 0.151209307, 0.06245031, 0.0967899)$

Table 5: True parameter values versus trained parameter values for model (4.10).

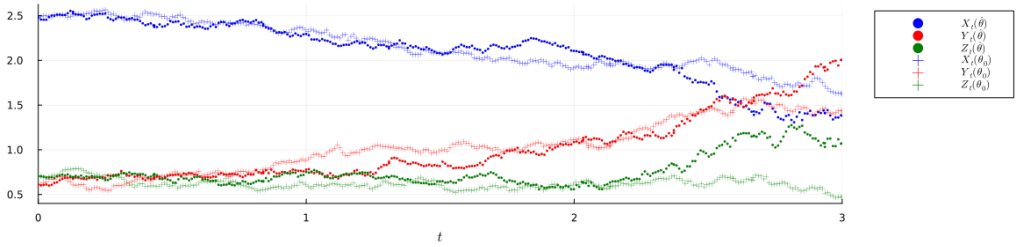

Figure 21: Simulation of model (4.10) for $\varepsilon = 0.3$.

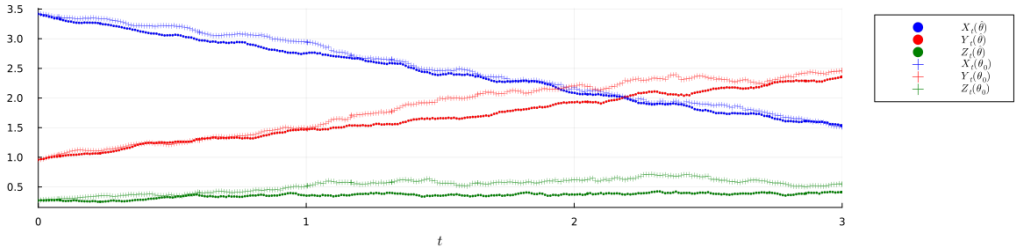

Figure 22: Simulation of model (4.10) for $\varepsilon = 0.1$.

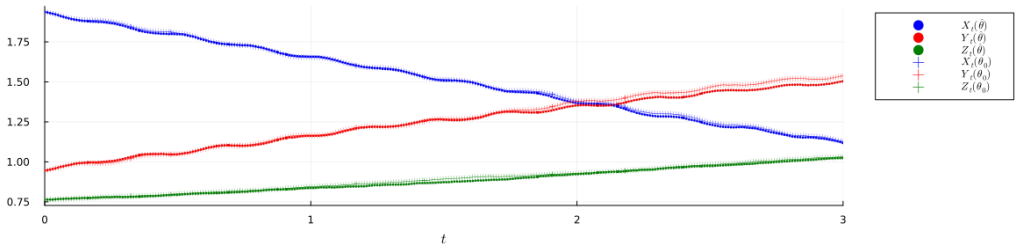

Figure 23: Simulation of model (4.10) for $\varepsilon = 0.01$.

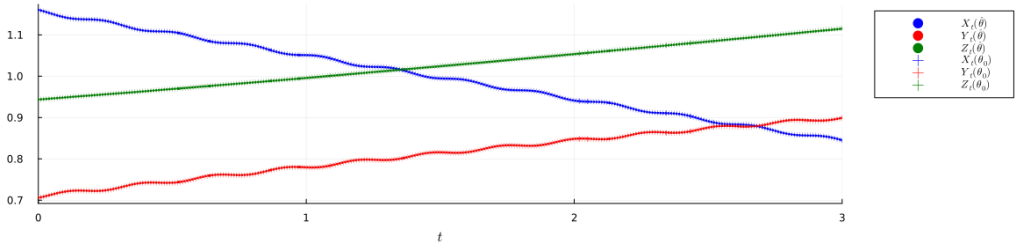

Figure 24: Simulation of model (4.10) for $\varepsilon = 0.001$.

ε	Component-wise MSE: $(\hat{X}_t, \hat{Y}_t, \hat{Z}_t)$
0.3	(0.119862755 0.4229034 0.1641841226)
0.1	(0.01354954 0.0286335 0.01803421)
0.01	(0.00032701 0.0003369 0.00015043)
0.001	(0.000221016 0.000165453 0.0000109)

Table 6: True parameter values versus trained parameter values for model (4.10).

Remark 4.28 *Mean-squared error is scale-dependent; hence, comparing the metrics, we have for the two model versions is of no practical use. The former model only takes values in the interval* $[0, 1]$ *whereas the latter model can take values larger than* 1.

4.4 Proofs of main results

In this section, we only give the proofs of Theorem 4.21 and 4.22. Following the similar arguments of [46] with suitable modifications presented in this section, we can prove Corollary 4.24. The details are omitted here.

Set $Y_t^{n,\varepsilon} = S^{\varepsilon}_{[nt]/n}$. For convenience, we will use C to denote a generic constant whose value may vary from place to place, and use E and P to denote E_{θ_0} and P_{θ_0}, respectively.

4.4.1 Proof of Theorem 4.21

Lemma 4.29 *Suppose* **(A1)** *and* **(A2)** *hold. Then,* $\{Y_t^{n,\varepsilon}\}$ *converges uniformly on compacts in probability to the deterministic process* $\{S_t^0\}$ *as* $\varepsilon \to 0$ *and* $n \to \infty$.

Proof. We follow the argument of [46, Lemma 4.1]. Let

$$J_t := N([0,t] \times \{|u| > 1\}),$$

which is a Poisson process with intensity $\lambda = \mu(\{|u| > 1\})$. Denote by $\tau_1 < \tau_2 < \cdots < \tau_n < \cdots$ the jump times of $\{J_t\}$. We have $\lim_{n\to\infty} \tau_n = \infty$ a.s.. We use the interlacing technique (cf. [3]) to construct the solution $(S_t^\varepsilon)_{t\geq 0}$ of (4.2). Set $\tau_0 = 0$. Let $\{Z_t^\varepsilon(i) : t \geq 0\}$ be the unique, strong solution to the SDE:

$$\begin{aligned} Z_t^\varepsilon(i) &= Z_0^\varepsilon(i) + \int_0^t b(s, Z_s^\varepsilon(i), \theta)ds + \varepsilon \int_0^t \left\{ \sigma(s, Z_s^\varepsilon(i))dB_s(i) + \int_{\{|u|\leq 1\}} H(s, Z_{s-}^\varepsilon(i), u)\widetilde{N}_i(ds, du) \right\}, \\ Z_0^\varepsilon(i) &= S_{\tau_{i-1}}^\varepsilon, \end{aligned}$$

where $B_t(i) = B_{\tau_{i-1}+t} - B_{\tau_{i-1}}$ and $N_i([0,t] \times A) = N([\tau_{i-1}, \tau_{i-1} + t] \times A)$ for any $A \in \mathcal{B}(\mathbb{R}^l - \{0\})$. Furthermore, let $Z_t^0 : t \geq 0$ be the unique, strong solution to the underlying deterministic differential equation:

$$Z_t^0(i) = Z_0^0(i) + \int_0^t b(s, Z_s^0(i), \theta)ds, \quad Z_0^0(i) = S_{\tau_{i-1}}^0.$$

By **(A1)**, we find that

$$Z_t^\varepsilon(i) = \begin{cases} S_{\tau_{i-1}+t}^\varepsilon, & 0 \leq t < \tau_i - \tau_{i-1}, \\ S_{\tau_i-}^\varepsilon, & t = \tau_i - \tau_{i-1}, \end{cases}$$

and $S_{\tau_i}^\varepsilon = S_{\tau_i-}^\varepsilon + \varepsilon G(\tau_i-, S_{\tau_i-}^\varepsilon, \xi_i)$, where $\{\xi_i : i \in \mathbb{N}\}$ are i.i.d. $\mathbb{R}^l$-valued random variables with common probability distribution $\frac{\mu(\cdot \cap \{|u|>1\})}{\lambda}$. First, we take $i = 1$. Then, $Z_0^\varepsilon(1) = x \in \mathbb{R}^d$. Let $\varepsilon > 0$ be fixed,

$$\overline{\tau}_M^\varepsilon := \inf\{t : |Z_t^\varepsilon(1)| \vee |Z_{t-}^\varepsilon(1)| > M\},$$

and

$$f \in C_b^2(\mathbb{R}^d) \text{ such that } f(x) = |x|^2 \text{ if } |x| \leq M.$$

By Itô's formula (cf. [52]), we find that

$$f(Z_{t\wedge\overline{\tau}_M^\varepsilon}^\varepsilon(1)) - f(x) - \int_0^{t\wedge\overline{\tau}_M^\varepsilon} A_s f(Z_s^\varepsilon(1))ds$$

is a martingale, where

$$\begin{aligned} A_s f(x) &= \sum_{k=1}^d b_i(s,x,\theta)\frac{\partial f}{\partial x_k}(x) + \frac{1}{2}\varepsilon^2 \sum_{k,j=1}^d \sum_{l=1}^r \sigma_l^k(s,x)\sigma_l^j(s,x)\frac{\partial^2 f}{\partial x_k \partial x_j}(x) \\ &\quad + \int_{\{|u|\leq 1\}} \left[f(x + \varepsilon H(s,x,u)) - f(x) - \varepsilon \sum_{k=1}^d H_k(s,x,u)\frac{\partial f}{\partial x_k}(x) \right] \mu(du). \end{aligned}$$

Now we consider

$$Z_t^\varepsilon(1) - Z_t^0(1) = \int_0^t [b(s, Z_s^\varepsilon(1), \theta) - b(s, Z_s^0(1), \theta)]ds + \varepsilon \int_0^t \sigma(s, Z_s^\varepsilon(1))dB_s(1) + \varepsilon \int_0^t \int_{\{|u|\le 1\}} H(s, Z_{s-}^\varepsilon(1), u)\tilde{N}_1(ds, du). \quad (4.11)$$

By Doob's inequality and **(A2)** given Equation (4.11), we get

$$\begin{aligned}
& E\left[\sup_{0\le s\le t} |Z_s^\varepsilon(1) - Z_s^0(1)|^2\right] \\
\le\ & 4E[|Z_t^\varepsilon(1) - Z_t^0(1)|^2] \\
\le\ & 12E\Bigg\{\left|\int_0^t b(s, Z_s^\varepsilon(1), \theta) - b(s, Z_s^0(1), \theta)ds\right|^2 \\
& \quad + \varepsilon^2\left|\int_0^t \sigma(s, Z_s^\varepsilon(1))dB_s(1)\right|^2 + \varepsilon^2\left|\int_0^t \int_{\{|u|\le 1\}} H(s, Z_{s-}^\varepsilon(1), u)\tilde{N}_1(ds, du)\right|^2\Bigg\} \\
\le\ & 12K^2\Bigg\{tE\left[\int_0^t |Z_s^\varepsilon(1) - Z_s^0(1)|^2 ds\right] + \varepsilon^2 E\left[\int_0^t (1 + |Z_s^\varepsilon(1)|)^2 ds\right] \\
& \quad + \varepsilon^2 E\left[\int_0^t \int_{\{|u|\le 1\}} \eta^2(u)(1 + |Z_{s-}^\varepsilon(1)|)^2 \mu(du)ds\right]\Bigg\} \\
\le\ & 12K^2\Bigg\{\left(t + 2\varepsilon^2 + 2\varepsilon^2 \int_{\{|u|\le 1\}} \eta^2(u)\mu(du)\right) E\left[\int_0^t |Z_s^\varepsilon(1) - Z_s^0(1)|^2 ds\right] \\
& \quad + 2t\varepsilon^2(1 + |x|)^2\left[1 + \int_{\{|u|\le 1\}} \eta^2(u)\mu(du)\right]\Bigg\}.
\end{aligned}$$

Then, by Gronwall's inequality, we obtain that

$$E\left[\sup_{0\le t\le 1} |Z_t^\varepsilon(1) - Z_t^0(1)|^2\right] \to 0 \text{ as } \varepsilon \to 0. \quad (4.12)$$

For any small $\delta > 0$, there is a T sufficiently large enough such that $P(\tau_1 > T) < \delta$. Recall that the processes $(S_t^\varepsilon)_{t\ge 0}$ and $(S_t^0)_{t\ge 0}$ are càdlàg and continuous, respectively. Then, by (4.12), we get

$$\begin{aligned}
P\left(\sup_{0\le t\le \tau_1} |S_{t-}^\varepsilon - S_t^0| > \gamma\right) &\le P\left(\sup_{0\le t\le \tau_1} |S_{t-}^\varepsilon - S_t^0| > \gamma; \tau_1 \le T\right) + P(\tau_1 > T) \\
&\le P\left(\sup_{0\le t\le T} |Z_t^\varepsilon(1) - Z_t^0(1)| > \gamma\right) + \delta \\
&\to 0 \text{ as } \quad \varepsilon,\ \delta \to 0.
\end{aligned}$$

Additionally, note that $S_{\tau_1}^\varepsilon = S_{\tau_1-}^\varepsilon + \varepsilon G(\tau_{1-}, S_{\tau_1-}^\varepsilon, \xi_1)$. Since $S_{\tau_1-}^\varepsilon \xrightarrow{P} S_{\tau_1}^0$, we get

$$|S_{\tau_1}^\varepsilon - S_{\tau_1}^0| \le |S_{\tau_1-}^\varepsilon - S_{\tau_1}^0| + \varepsilon K(1 + |S_{\tau_1-}^\varepsilon|)|\xi_1| \xrightarrow{P} 0.$$

Thus,

$$P\left(\sup_{0\le t\le\tau_1}|S_t^\varepsilon - S_t^0| > \gamma\right) \to 0 \text{ as } \varepsilon \to 0.$$

Next, we consider $\{Z_t^\varepsilon(2) : t \ge 0\}$, where $Z_0^\varepsilon(2) = S_{\tau_1}^\varepsilon$. We have

$$\begin{aligned} Z_t^\varepsilon(2) - Z_t^0(2) &= S_{\tau_1}^\varepsilon - S_{\tau_1}^0 + \int_0^t [b(s, Z_s^\varepsilon(2), \theta) - b(s, Z_s^0(2), \theta)]ds \\ &\quad + \varepsilon\int_0^t \sigma(s, Z_s^\varepsilon(2))dB_s(2) + \varepsilon\int_0^t\int_{\{|u|\le 1\}} H(s, Z_{s-}^\varepsilon(2), u)\tilde{N}_2(ds, du). \end{aligned}$$

Using the arguments above here will yield the result

$$\lim_{\varepsilon\to 0} E\left[\sup_{0\le t\le T}|Z_t^\varepsilon(2) - Z_t^0(2)|1_{\{|S_{\tau_1}^\varepsilon|+|S_{\tau_1}^0|\le M\}}\right] = 0.$$

For any small $\delta > 0$, there exists $M > 0$ such that $P(|S_{\tau_1}^0| > \frac{M}{4}) < \delta$. Then, we have that

$$\begin{aligned} &\limsup_{\varepsilon\to 0} P\left(\sup_{0\le t\le T\wedge\tau_2}|Z_t^\varepsilon(2) - Z_t^0(2)| > \gamma\right) \\ \le\ &\limsup_{\varepsilon\to 0} P\left(\sup_{0\le t\le T}|Z_t^\varepsilon(2) - Z_t^0(2)| > \gamma; |S_{\tau_1}^\varepsilon| + |S_{\tau_1}^0| \le M\right) \\ &+ \limsup_{\varepsilon\to 0} P\left(|S_{\tau_1}^\varepsilon - S_{\tau_1}^0| > \frac{M}{2}\right) + P\left(|S_{\tau_1}^0| > \frac{M}{4}\right) \\ <\ &\delta. \end{aligned}$$

Recall $S_{\tau_2}^\varepsilon = S_{\tau_{2-}}^\varepsilon + \varepsilon G(\tau_{2-}, S_{\tau_{2-}}^\varepsilon, \xi_2)$ and thus as before we have

$$P\left(\sup_{\tau_1\le t\le\tau_2}|S_t^\varepsilon - S_t^0| > \gamma\right) \to 0 \text{ as } \varepsilon \to 0.$$

By induction on $i \ge 1$ we now have

$$P\left(\sup_{\tau_{i-1}\le t\le\tau_i}|S_t^\varepsilon - S_t^0| > \gamma\right) \to 0 \text{ as } \varepsilon \to 0.$$

As $i \to \infty$, it happens to be $\tau_i \to \infty$ a.s., thus for any given $T > 0, \delta > 0$ there exists $i_0 \in \mathbb{N}$ such that when $i \ge i_0$ implies $P(\tau_i < T) < \frac{\delta}{2}$. Moreover, we have for every $i > 0, \delta > 0$ there exists ε_0 such that for all $\varepsilon < \varepsilon_0$

$$P\left(\sup_{\tau_{i-1}\le t\le\tau_i}|S_t^\varepsilon - S_t^0| > \gamma\right) < \frac{\delta}{2^{i+1}i_0}.$$

Then, for $\varepsilon < \varepsilon_0$,

$$P\left(\sup_{0\le t\le 1}|S_t^\varepsilon - S_t^0| > \gamma\right) \le \sum_{k=1}^{i} P\left(\sup_{\tau_{k-1}\le t\le\tau_k}|S_t^\varepsilon - S_t^0| > \gamma\right) < \frac{\delta}{2i_0}.$$

Thus,

$$\begin{aligned}
& P\left(\sup_{0\le t\le 1}|S_t^\varepsilon - S_t^0| > \gamma\right) \\
= \;& P\left(\sup_{0\le t\le 1}|S_t^\varepsilon - S_t^0| > \gamma; \tau_{i_0} < 1\right) + P\left(\sup_{0\le t\le 1}|S_t^\varepsilon - S_t^0| > \gamma; \tau_{i_0} \ge 1\right) \\
\le \;& P\left(\tau_{i_0} < 1\right) + \sum_{k=1}^{i_0} P\left(\sup_{0\le t\le 1}|S_t^\varepsilon - S_t^0|; \tau_{k-1} \le t < \tau_k\right) \\
< \;& \delta.
\end{aligned}$$

To wrap it up, since S_t^ε is a semimartingale, by [52, Theorem II.11], we find that for any fixed $\varepsilon > 0$,

$$\sup_{0\le t\le 1}|Y_t^{n,\varepsilon} - S_t^\varepsilon| \xrightarrow{P} 0 \;\text{ as } n\to\infty.$$

Therefore, we can ascertain

$$P\left(\sup_{0\le t\le 1}|Y_t^{n,\varepsilon} - S_t^0| > \gamma\right) \to 0 \text{ as } \varepsilon\to 0,\, n\to\infty. \qquad \square$$

Lemma 4.30 *Suppose* **(A1)** *and* **(A2)** *hold. Let* $f \in C_\uparrow^{1,1,1}\left([0,1]\times\mathbb{R}^d\times\Theta\right)$. *Then,*

$$\frac{1}{n}\sum_{k=1}^n f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) \xrightarrow{P_{\theta_0}} \int_0^1 f(s, S_s^0, \theta)ds.$$

as $\varepsilon\to 0$ *and* $n\to\infty$, *uniformly in* $\theta\in\Theta$.

Proof. We follow the argument of [45, Lemma 3.3]. Using the condition imposed on f and Lemma 4.29, we get

$$\begin{aligned}
& \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{k=1}^n f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) - \int_0^1 f(s, S_s^0, \theta)ds\right| \\
\le \;& \sup_{\theta\in\Theta}\sum_{k=1}^n\int_{t_{k-1}}^{t_k}|f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) - f(s, Y_s^{n,\varepsilon}, \theta)|ds + \sup_{\theta\in\Theta}\int_0^1|f(s, Y_s^{n,\varepsilon}, \theta) - f(s, S_s^0, \theta)|ds \\
\le \;& \sum_{k=1}^n\int_{t_{k-1}}^{t_k}\left(\int_0^1\sup_{\theta\in\Theta}|(\nabla_t f)(t_{k-1} + v(s-t_{k-1}), Y_s^{n,\varepsilon}, \theta)|dv\right)|s - t_{k-1}|ds \\
& + \int_0^1\left(\int_0^1\sup_{\theta\in\Theta}|(\nabla_x f)\left(s, S_s^0 + u\left(Y_s^{n,\varepsilon} - S_s^0\right), \theta\right)|du\right)|Y_s^{n,\varepsilon} - S_s^0|ds \\
\le \;& C\left(1 + \sup_{s\in[0,1]}|S_s^0| + \sup_{s\in[0,1]}|S_s^\varepsilon|\right)^\lambda\left[\frac{1}{n} + \sup_{s\in[0,1]}|Y_s^{n,\varepsilon} - S_s^0|\right] \\
\xrightarrow{P_{\theta_0}} \;& 0 \text{ as } \varepsilon\to 0,\, n\to\infty. \qquad \square
\end{aligned}$$

Define

$$\tau_m^\varepsilon = \inf\{t \geq 0 : |S_t^\varepsilon| \vee |S_{t-}^\varepsilon| \geq m\}, \qquad \tau_m^0 = \inf\{t \geq 0 : |S_t^0| \geq m\}.$$

Similar to [46, Lemma 4.3], we can prove the following lemma.

Lemma 4.31 *Suppose* **(A1)** *and* **(A2)** *hold. Then, for any* $m > 0$, $\tau_m^\varepsilon \xrightarrow{P_{\theta_0}} \tau_m^0$ *as* $\varepsilon \to 0$.

Lemma 4.32 *Suppose* **(A1)**–**(A3)** *hold. Let* $f \in C_\uparrow^{1,1,1}\left([0,1] \times \mathbb{R}^d \times \Theta\right)$. *Then, for* $1 \leq i \leq d$,

$$\sum_{k=1}^{n} f(t_{k-1}, S_{t_{k-1}}^\varepsilon, \theta)\left(S_{t_k}^{\varepsilon,i} - S_{t_{k-1}}^{\varepsilon,i} - b_i(t_{k-1}, S_{t_{k-1}}^\varepsilon, \theta_0)\Delta_{t_{k-1}}\right) \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon \to 0$ *and* $n \to \infty$, *uniformly in* $\theta \in \Theta$, *where* $S_t^{\varepsilon,i}$ *and* b_i *are the* i^{th} *components of* S_t^ε *and* b, *respectively, and* $\Delta_{t_{k-1}} := t_k - t_{k-1} = \frac{1}{n}$.

Proof. Recall that

$$\begin{aligned} S_{t_k}^{\varepsilon,i} &= S_{t_{k-1}}^{\varepsilon,i} + \int_{t_{k-1}}^{t_k} b_i(s, S_s^\varepsilon, \theta_0)ds + \varepsilon \int_{t_{k-1}}^{t_k} \Bigg(\sigma_i(s, S_s^\varepsilon)dB_s \\ &\quad + \int_{\{|u| \leq 1\}} H_i(s, S_{s-}^\varepsilon, u)\widetilde{N}(ds, du) + \int_{\{|u|>1\}} G_i(s, S_{s-}^\varepsilon, u)N(ds, du)\Bigg). \end{aligned}$$

Then, we have

$$
\begin{aligned}
&\sum_{k=1}^{n} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\left(S^{\varepsilon,i}_{t_k} - S^{\varepsilon,i}_{t_{k-1}} - b_i(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta_0)\Delta_{t_{k-1}}\right)\\
=\;& \sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\left(b_i(s, S^{\varepsilon}_{s}, \theta_0) - b_i(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta_0)\right) ds\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\sigma_i(s, S^{\varepsilon}_{s}) dB_s\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\int_{\{|u|\le 1\}} H_i(s, S^{\varepsilon}_{s-}, u)\widetilde{N}(ds, du)\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\int_{\{|u|> 1\}} G_i(s, S^{\varepsilon}_{s-}, u)N(ds, du)\\
=\;& \int_0^1 f(s, Y^{n,\varepsilon}_{s}, \theta)\left(b_i(s, S^{\varepsilon}_{s}, \theta_0) - b_i(s, Y^{n,\varepsilon}_{s}, \theta_0)\right) ds\\
&+\varepsilon\int_0^1 f(s, Y^{n,\varepsilon}_{s}, \theta)\sigma_i(s, S^{\varepsilon}_{s}) dB_s\\
&+\varepsilon\int_0^1 f(s, Y^{n,\varepsilon}_{s}, \theta)\int_{\{|u|\le 1\}} H_i(s, S^{\varepsilon}_{s-}, u)\widetilde{N}(ds, du)\\
&+\varepsilon\int_0^1 f(s, Y^{n,\varepsilon}_{s}, \theta)\int_{\{|u|> 1\}} G_i(s, S^{\varepsilon}_{s-}, u)N(ds, du)\\
&+\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left[f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_{s}, \theta)\right]\left(b_i(s, S^{\varepsilon}_{s}, \theta_0) - b_i(s, Y^{n,\varepsilon}_{s}, \theta_0)\right) ds\\
&+\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta)\left(b_i(s, Y^{n,\varepsilon}_{s}, \theta_0) - b_i(t_{k-1}, Y^{n,\varepsilon}_{t_{k-1}}, \theta_0)\right) ds\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left[f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_{s}, \theta)\right]\sigma_i(s, S^{\varepsilon}_{s}) dB_s\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left[f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_{s}, \theta)\right]\int_{\{|u|\le 1\}} H_i(s, S^{\varepsilon}_{s-}, u)\widetilde{N}(ds, du)\\
&+\varepsilon\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left[f(t_{k-1}, S^{\varepsilon}_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_{s}, \theta)\right]\int_{\{|u|> 1\}} G_i(s, S^{\varepsilon}_{s-}, u)N(ds, du).
\end{aligned}
$$

Using the condition imposed on f, (A2), (A3), Lemma 4.29, Lemma 4.31 and following the argument as in the proof of [45, Lemma 3.5], we can show that all of the above terms converge to zero in probability as $\varepsilon \to 0$ and $n \to \infty$, uniformly in $\theta \in \Theta$. □

Proof of Theorem 4.21. Define

$$\Phi_{n,\varepsilon}(\theta) := \Psi_{n,\varepsilon}(\theta) - \Psi_{n,\varepsilon}(\theta_0).$$

We have

$$\hat{\theta}_{n,\varepsilon} = \arg\min_{\theta\in\Theta}\Phi_{n,\varepsilon}(\theta),$$

and

$$\begin{aligned}\Phi_{n,\varepsilon}(\theta) &= -2\sum_{k=1}^{n}(b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0))^*\left(S^{\varepsilon}_{t_k}-S^{\varepsilon}_{t_{k-1}}-\frac{1}{n}b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)\right)\\ &\quad+\frac{1}{n}\sum_{k=1}^{n}|b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)|^2\\ &:= \Phi^{(1)}_{n,\varepsilon}(\theta)+\Phi^{(2)}_{n,\varepsilon}(\theta).\end{aligned}$$

By Lemma 4.32 and letting $f(t,x,\theta)=b_i(t,x,\theta)-b_i(t,x,\theta_0)$, $1\le i\le d$, we get $\sup_{\theta\in\Theta}|\Phi^{(1)}_{n,\varepsilon}(\theta)|\xrightarrow{P_{\theta_0}}0$ as $\varepsilon\to 0$ and $n\to\infty$. By Lemma 4.30 and letting $f(t,x,\theta)=|b(t,x,\theta)-b(t,x,\theta_0)|^2$, we get $\sup_{\theta\in\Theta}|\Phi^{(2)}_{n,\varepsilon}(\theta)-F(\theta)|\xrightarrow{P_{\theta_0}}0$ as $\varepsilon\to 0$ and $n\to\infty$, where

$$F(\theta) := \int_0^1 |b(t,S^0_t,\theta)-b(t,S^0_t,\theta_0)|^2dt.$$

Then, we have that

$$\sup_{\theta\in\Theta}|\Phi_{n,\varepsilon}(\theta)-F(\theta)|\xrightarrow{P_{\theta_0}}0 \quad \text{as } \varepsilon\to 0,\ n\to\infty.$$

By **(A4)** and the continuity of S^0, we get

$$\sup_{|\theta-\theta_0|>\delta} -F(\theta) < -F(\theta_0)=0, \qquad \forall\delta>0.$$

Therefore, the proof is complete by [61, Theorem 5.7]. □

4.4.2 Proof of Theorem 4.22

Note that

$$\nabla_\theta\Phi_{n,\varepsilon} = -2\sum_{k=1}^{n}(\nabla_\theta b)^*(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)(S^{\varepsilon}_{t_k}-S^{\varepsilon}_{t_{k-1}}-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)\Delta_{t_{k-1}}).$$

Define $G_{n,\varepsilon}(\theta)=(G^i_{n,\varepsilon})^*_{1\le i\le p}$, where

$$G^i_{n,\varepsilon}(\theta) := \sum_{k=1}^{n}(\partial_{\theta_i}b)^*(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)(S^{\varepsilon}_{t_k}-S^{\varepsilon}_{t_{k-1}}-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)\Delta_{t_{k-1}}), \qquad 1\le i\le p,$$

and define

$$K_{n,\varepsilon}(\theta) := \nabla_\theta G_{n,\varepsilon}(\theta),$$

namely, a $p \times p$ matrix comprised of $(\partial_{\theta_j} G^i_{n,\varepsilon})_{1\le i,j\le p}$. Lastly, define

$$K^{ij}(\theta) = \int_0^1 (\partial_{\theta_j}\partial_{\theta_i} b)^*(s, S^0_s, \theta)(b(s, S^0_s, \theta_0) - b(s, S^0_s, \theta))ds - I^{ij}(\theta),$$

and

$$K(\theta) = (K^{ij}(\theta))_{1\le i,j\le p}.$$

Lemma 4.33 *Suppose* **(A1)**, **(A2)** *and* **(A5)** *hold. Let* $f \in C^{1,1,1}_{\uparrow}\left([0,1]\times\mathbb{R}^d\times\Theta\right)$. *Then, for* $1 \le i \le d$ *and each* $\theta \in \Theta$,

$$\begin{aligned}
&\sum_{k=1}^n f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)\bigg(\int_{t_{k-1}}^{t_k} \sigma(s, S^\varepsilon_s)dB_s + \int_{t_{k-1}}^{t_k}\int_{\{|u|\le 1\}} H_i(s, S^\varepsilon_{s-}, u)\widetilde{N}(ds, du)\\
&\qquad + \int_{t_{k-1}}^{t_k}\int_{\{|u|>1\}} G_i(s, S^\varepsilon_{s-}, u)N(ds, du)\bigg)\\
\xrightarrow{P_{\theta_0}}\ &\int_0^1 f(s, S^0_s, \theta)\bigg(\sigma_i(s, S^0_s)dB_s + \int_{\{|u|\le 1\}} H_i(s, S^0_s, u)\widetilde{N}(ds, du)\\
&\qquad + \int_{\{|u|>1\}} G_i(s, S^0_s, u)N(ds, du)\bigg),
\end{aligned}$$

as $\varepsilon \to 0$ *and* $n \to \infty$.

Proof. We follow the arguments of [45, Lemma 3.4] and [46, Lemma 4.6]. Define

$$\begin{aligned}
A_{n,\varepsilon} &:= \sum_{k=1}^n f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)\bigg(\int_{t_{k-1}}^{t_k} \sigma_i(s, S^\varepsilon_s)dB_s + \int_{t_{k-1}}^{t_k}\int_{\{|u|\le 1\}} H_i(s, S^\varepsilon_{s-}, u)\widetilde{N}(ds, du)\\
&\qquad + \int_{t_{k-1}}^{t_k}\int_{\{|u|>1\}} G_i(s, S^\varepsilon_{s-}, u)N(ds, du)\bigg),\\
B_{n,\varepsilon} &:= \int_0^1 f(s, Y^{n,\varepsilon}_s, \theta)\bigg(\sigma_i(s, S^\varepsilon_s)dB_s + \int_{\{|u|\le 1\}} H_i(s, S^\varepsilon_{s-}, u)\widetilde{N}(ds, du)\\
&\qquad + \int_{\{|u|>1\}} G_i(s, S^\varepsilon_{s-}, u)N(ds, du)\bigg).
\end{aligned}$$

We can show that $A_{n,\varepsilon} = B_{n,\varepsilon} + o_P(1)$. In fact, we have

$$\begin{aligned}
&|A_{n,\varepsilon} - B_{n,\varepsilon}|\\
\le\ &\sum_{k=1}^n \int_{t_{k-1}}^{t_k} |f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_s, \theta)| \int_{\{|u|>1\}} |G_i(s, S^\varepsilon_{s-}, u)|N(ds, du)\\
&+\bigg|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} [f(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) - f(s, Y^{n,\varepsilon}_s, \theta)]\bigg(\sigma_i(s, S^\varepsilon_s)dB_s + \int_{\{|u|\le 1\}} H_i(s, S^\varepsilon_{s-}, u)\widetilde{N}(ds, du)\bigg)\bigg|\\
:=\ &I_{n,\varepsilon} + J_{n,\varepsilon},
\end{aligned}$$

$$
\begin{aligned}
I_{n,\varepsilon} &\leq \sup_{\theta\in\Theta}\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(\int_0^1 |(\nabla_t f)(t_{k-1}+v(s-t_{k-1}),Y_s^{n,\varepsilon},\theta)|dv\right)|s-t_{k-1}| \\
&\quad \cdot\int_{\{|u|>1\}}|G_i(s,S_{s-}^{\varepsilon},u)|N(ds,du) \\
&\leq \frac{CK}{n}\left(1+\sup_{t\in[0,1]}|S_t^{\varepsilon}|\right)^{\lambda+1}\int_{\{|u|>1\}}\xi(u)N(ds,du) \\
&\xrightarrow{P} 0 \text{ as } \varepsilon\to 0,\ n\to\infty.
\end{aligned}
$$

For $J_{n,\varepsilon}$ and $\eta>0$, using the stopping time τ_m^{ε}, Lemma 4.29, Markov's inequality, and dominated convergence we have

$$
\begin{aligned}
&P\left(\left|\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}[f(t_{k-1},S_{t_{k-1}}^{\varepsilon},\theta)-f(s,Y_s^{n,\varepsilon},\theta)]\left(\sigma_i(s,S_s^{\varepsilon})1_{\{s\leq\tau_m^{\varepsilon}\}}dB_s\right.\right.\right. \\
&\qquad\left.\left.\left.+\int_{\{|u|\leq 1\}}H_i(s,S_{s-}^{\varepsilon},u)1_{\{s\leq\tau_m^{\varepsilon}\}}\widetilde{N}(ds,du)\right)\right|>\eta\right) \\
\leq\ & \frac{1}{\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}|f(t_{k-1},S_{t_{k-1}}^{\varepsilon},\theta)-f(s,Y_s^{n,\varepsilon},\theta)|^2\sigma_i^2(s,S_s^{\varepsilon})1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&+\frac{1}{\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}|f(t_{k-1},S_{t_{k-1}}^{\varepsilon},\theta)-f(s,Y_s^{n,\varepsilon},\theta)|^2 1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&\quad\cdot\left(E\left\{\int_0^1\int_{\{|u|\leq 1\}}|H_i(s,S_{s-}^{\varepsilon},u)|^2\mu(du)1_{\{s\leq\tau_m^{\varepsilon}\}}\right\}ds\right)^{1/2} \\
\leq\ & \frac{1}{\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(\int_0^1|(\nabla_t f)(t_{k-1}+v(s-t_{k-1}),Y_s^{n,\varepsilon},\theta)|dv|s-t_{k-1}|\right)^2\sigma_i^2(s,S_s^{\varepsilon})1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&+\frac{1}{\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(\int_0^1|(\nabla_t f)(t_{k-1}+v(s-t_{k-1}),Y_s^{n,\varepsilon},\theta)|dv|s-t_{k-1}|\right)^2 1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&\quad\cdot\left(E\left\{\int_0^1\int_{\{|u|\leq 1\}}|H_i(s,S_{s-}^{\varepsilon},u)|^2\mu(du)1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
\leq\ & \frac{1}{n\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(C(1+|Y_s^{n,\varepsilon}|)^{\lambda}\right)^2\sigma_i^2(s,S_s^{\varepsilon})1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&+\frac{K}{n\eta}\left(E\left\{\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(C(1+|Y_s^{n,\varepsilon}|)^{\lambda}\right)^2 1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2} \\
&\quad\cdot\left(E\left\{\int_0^1(1+|Y_s^{n,\varepsilon}|)^2 1_{\{s\leq\tau_m^{\varepsilon}\}}ds\right\}\right)^{1/2}\left(\int_{\{|u|\leq 1\}}\eta^2(u)\mu(du)\right)^{1/2} \\
\to\ & 0 \text{ as } \varepsilon\to 0,\ n\to\infty,
\end{aligned}
$$

and

$$
\begin{aligned}
& P\Bigg(\Bigg|\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}[f(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)-f(s,Y^{n,\varepsilon}_s,\theta)]\Bigg(\sigma_i(s,S^{\varepsilon}_s)dB_s \\
& \qquad + \int_{\{|u|\le 1\}} H_i(s,S^{\varepsilon}_{s-},u)\widetilde{N}(ds,du)\Bigg)\Bigg| > \eta\Bigg) \\
\le\ & P\Bigg(\Bigg|\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}[f(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta)-f(s,Y^{n,\varepsilon}_s,\theta)]\Bigg(\sigma_i(s,S^{\varepsilon}_s)1_{\{s\le\tau^{\varepsilon}_m\}}dB_s \\
& \qquad + \int_{\{|u|\le 1\}} H_i(s,S^{\varepsilon}_{s-},u)1_{\{s\le\tau^{\varepsilon}_m\}}\widetilde{N}(ds,du)\Bigg)\Bigg| > \eta\Bigg) + P(\tau^{\varepsilon}_m < 1) \\
\to\ & 0 \text{ as } \varepsilon\to 0,\ n\to\infty,
\end{aligned}
$$

by Lemma 4.31.

Finally, using the condition imposed on f, (A5), [25, Problem 13, page 151], the continuous mapping theorem, Lemma 4.29 and similar arguments as above, we can show that

$$
\begin{aligned}
B_{n,\varepsilon} \xrightarrow{P_{\theta_0}} & \int_0^1 f(s,S^0_s,\theta)\Bigg(\sigma_i(s,S^0_s)dB_s + \int_{\{|u|\le 1\}} H_i(s,S^0_s,u)\widetilde{N}(ds,du) \\
& + \int_{\{|u|>1\}} G_i(s,S^0_s,u)N(ds,du)\Bigg)
\end{aligned}
$$

as $\varepsilon\to 0$ and $n\to\infty$. □

Lemma 4.34 *Suppose* **(A1)**–**(A6)** *hold. Then, for* $1\le i\le p$,

$$
\begin{aligned}
\varepsilon^{-1}G^i_{n,\varepsilon}(\theta_0) \xrightarrow{P_{\theta_0}} & \int_0^1 (\partial_{\theta_i}b)^*(s,S^0_s,\theta_0)\Bigg(\sigma(s,S^0_s)dB_s + \int_{\{|u|\le 1\}} H(s,S^0_s,u)\widetilde{N}(ds,du) \\
& + \int_{\{|u|>1\}} G(s,S^0_s,u)N(ds,du)\Bigg)
\end{aligned}
$$

as $\varepsilon\to 0$ *and* $n\to\infty$.

Proof. We follow the arguments of [45, Lemma 3.6] and [46, Lemma 4.7]. For $1\le i\le p$, we have

$$
\begin{aligned}
& \varepsilon^{-1}G^i_{n,\varepsilon}(\theta_0) \\
=\ & \varepsilon^{-1}\sum_{k=1}^{n}(\partial_{\theta_i}b)^*(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)(S^{\varepsilon}_{t_k}-S^{\varepsilon}_{t_{k-1}}-b(t_{k-1},S^{\varepsilon}_{t_k},\theta_0)\Delta_{k-1})
\end{aligned}
$$

$$
\begin{aligned}
&= \varepsilon^{-1}\sum_{k=1}^{n}(\partial_{\theta_i}b)^*(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)\int_{t_{k-1}}^{t_k}(b(s,S^{\varepsilon}_s,\theta_0)-b(t_{k-1},S^{\varepsilon}_{t_k},\theta_0))ds\\
&\quad+\sum_{k=1}^{n}(\partial_{\theta_i}b)^*(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)\int_{t_{k-1}}^{t_k}\Big(\sigma(s,S^{\varepsilon}_s)dB_s+\int_{\{|u|\le 1\}}H(s,S^{\varepsilon}_{s-},u)\widetilde{N}(ds,du)\\
&\qquad+\int_{\{|u|>1\}}G(s,X^{\varepsilon}_{s-},u)N(ds,du)\Big)\\
&:= H^{(1)}_{n,\varepsilon}(\theta_0)+H^{(2)}_{n,\varepsilon}(\theta_0).
\end{aligned}
$$

Using Lemma 4.33 with $f(t,x,\theta)=(\partial_{\theta_i}b_j(t,x,\theta))^*(t,x,\theta)$ and $\theta=\theta_0$ for $1\le i\le p$, $1\le j\le d$, we get

$$
\begin{aligned}
H^{(2)}_{n,\varepsilon}(\theta_0) \xrightarrow{P_{\theta_0}} &\int_0^1(\partial_{\theta_i}b)^*(s,S^0_s,\theta_0)\Big(\sigma(s,S^0_s)dB_s+\int_{\{|u|\le 1\}}H(s,S^0_s,u)\widetilde{N}(ds,du)\\
&+\int_{\{|u|>1\}}G(s,S^0_s,u)N(ds,du)\Big)
\end{aligned}
$$

as $\varepsilon\to 0$ and $n\to\infty$.

For $H^{(1)}_{n,\varepsilon}(\theta_0)$, given $s\in[t_{k-1},t_k]$, we have that

$$
\begin{aligned}
&S^{\varepsilon}_s-S^{\varepsilon}_{t_{k-1}}\\
=\ &\int_{t_{k-1}}^{s}(b(r,S^{\varepsilon}_r,\theta_0)-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0))dr+b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)(s-t_{k-1})\\
&+\varepsilon\int_{t_{k-1}}^{s}\Big(\sigma(s,S^{\varepsilon}_s)dB_s+\int_{\{|u|\le 1\}}H(s,S^{\varepsilon}_{s-},u)\widetilde{N}(ds,du)+\int_{\{|u|>1\}}G(s,S^{\varepsilon}_{s-},u)N(ds,du)\Big).
\end{aligned}
$$

Using the Lipschitz condition on b and the Cauchy-Schwarz, we get

$$
\begin{aligned}
|S^{\varepsilon}_s-S^{\varepsilon}_{t_{k-1}}|^2 \le\ &2\Big|\int_{t_{k-1}}^{s}(b(r,S^{\varepsilon}_r,\theta_0)-b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0))dr\Big|^2\\
&+2\Big\{|b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)|(s-t_{k-1})+\varepsilon\int_{t_{k-1}}^{s}\Big(\sigma(r,S^{\varepsilon}_r)dB_r\\
&+\int_{\{|u|\le 1\}}H(r,S^{\varepsilon}_{r-},u)\widetilde{N}(dr,du)+\int_{\{|u|>1\}}G(r,S^{\varepsilon}_{r-},u)N(dr,du)\Big)\Big\}^2\\
\le\ &2K^2n^{-1}\Big(\int_{t_{k-1}}^{s}|S^{\varepsilon}_r-S^{\varepsilon}_{t_{k-1}}|dr\Big)^2+2\Big\{n^{-1}|b(t_{k-1},S^{\varepsilon}_{t_{k-1}},\theta_0)|\\
&+\varepsilon\sup_{t_{k-1}\le s\le t_k}\Big|\int_{t_{k-1}}^{s}\Big(\sigma(r,S^{\varepsilon}_r)dB_r+\int_{\{|u|\le 1\}}H(r,S^{\varepsilon}_{r-},u)\widetilde{N}(dr,du)\\
&+\int_{\{|u|>1\}}G(r,S^{\varepsilon}_{r-},u)N(dr,du)\Big)\Big|\Big\}^2.
\end{aligned}
$$

Now applying Gronwall's inequality, we get

$$
\begin{aligned}
&\sup_{t_{k-1}\le s\le t_k}|S^\varepsilon_s - S^\varepsilon_{t_{k-1}}| \\
\le\ & \sqrt{2}e^{K^2n^{-2}}\Bigg(n^{-1}|b(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)| \\
&+\varepsilon\sup_{t_{k-1}\le s\le t_k}\left|\int_{t_{k-1}}^s \sigma(r,S^\varepsilon_r)dB_r + \int_{\{|u|\le 1\}} H(r,S^\varepsilon_{r-},u)\widetilde{N}(dr,du) + \int_{\{|u|>1\}} G(r,S^\varepsilon_{r-},u)N(dr,du)\right|\Bigg).
\end{aligned}
$$

Further, by the Lipschitz condition on b and (A3), we obtain that

$$
\begin{aligned}
&|H^{(1)}_{n,\varepsilon}(\theta_0)|1_{\{1\le\tau^\varepsilon_m\}} \\
\le\ & \varepsilon^{-1}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)|\cdot\int_{t_{k-1}}^{t_k}|(b(s,S^\varepsilon_s,\theta_0)-b(t_{k-1},S^\varepsilon_{t_k},\theta_0))|1_{\{s\le\tau^\varepsilon_m\}}ds \\
\le\ & \varepsilon^{-1}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)|\cdot\left[\int_{t_{k-1}}^{t_k} K|S^\varepsilon_s - S^\varepsilon_{t_{k-1}}|ds + Cn^{-1}(1+m)^\lambda\right] \\
\le\ & (n\varepsilon)^{-1}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)|\left[K|\sup_{t_{k-1}\le s\le t_k}|S^\varepsilon_s - S^\varepsilon_{t_{k-1}}| + C(1+m)^\lambda\right] \\
\le\ & \sqrt{2}(n\varepsilon)^{-1}Ke^{K^2n^{-2}}n^{-1}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)|\cdot|b(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)| \\
&+\sqrt{2}n^{-1}Ke^{K^2n^{-2}}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)| \\
&\cdot\sup_{t_{k-1}\le s\le t_k}\left|\int_{t_{k-1}}^s \sigma(r,S^\varepsilon_r)dB_r + \int_{\{|u|\le 1\}} H(r,S^\varepsilon_{r-},u)\widetilde{N}(dr,du) + \int_{\{|u|>1\}} G(r,S^\varepsilon_{r-},u)N(dr,du)\right| \\
&+C(1+m)^\lambda(n\varepsilon)^{-1}\sum_{k=1}^n |(\partial_{\theta_i}b)^*(t_{k-1},S^\varepsilon_{t_{k-1}},\theta_0)| \\
:=\ & H^{(1,1)}_{n,\varepsilon}(\theta_0) + H^{(1,2)}_{n,\varepsilon}(\theta_0) + H^{(1,3)}_{n,\varepsilon}(\theta_0).
\end{aligned}
$$

By (A2) and (A3), we can show that all of the above terms converge to zero in probability as $\varepsilon\to 0$ and $n\to\infty$. Finally, the proof is completed by Lemma 4.31. □

Lemma 4.35 *Assume that conditions (A1)-(A6) hold and $I(\theta_0)$ is positive definite. Then,*

$$\sup_{\theta\in\Theta}|K_{n,\varepsilon}(\theta) - K(\theta)| \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon\to 0$ and $n\to\infty$.

Proof. We follow the argument of [45, Lemma 3.7]. For $1 \leq i, j \leq p$, we have

$$\begin{aligned} K^{ij}_{n,\varepsilon}(\theta) &= \partial_{\theta_j} G^i_{n,\varepsilon}(\theta) \\ &= \sum_{k=1}^{n} (\partial_{\theta_j}\partial_{\theta_i} b)^*(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)(S^\varepsilon_{t_k} - S^\varepsilon_{t_{k-1}} - b(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta_0)\Delta_{t_{k-1}}) \\ &\quad + n^{-1} \sum_{k=1}^{n} \Big((\partial_{\theta_j}\partial_{\theta_i} b)^*(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)(b(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta_0) - b(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)) \\ &\qquad\qquad - (\partial_{\theta_i} b)^*(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta)\partial_{\theta_j} b(t_{k-1}, S^\varepsilon_{t_{k-1}}, \theta) \Big) \\ &:= K^{(1)}_{n,\varepsilon}(\theta) + K^{(2)}_{n,\varepsilon}(\theta). \end{aligned}$$

By application of Lemma 4.32 and setting $f(t, x, \theta) = \partial_{\theta_j}\partial_{\theta_i} b_l(t, x, \theta)$, $1 \leq i, j \leq p, 1 \leq l \leq d$, we get

$$\sup_{\theta\in\Theta} |K^{ij,(1)}_{n,\varepsilon}(\theta)| \xrightarrow{P_{\theta_0}} 0 \text{ as } \varepsilon \to 0, n \to \infty.$$

By application of Lemma 4.30 and setting $f(t, x, \theta) = (\partial_{\theta_j}\partial_{\theta_i} b)^*(t, x, \theta)(b(t, x, \theta_0) - b(t, x, \theta)) - (\partial_{\theta_i} b)^*(t, x, \theta)\partial_{\theta_j} b(t, x, \theta)$, we get

$$\sup_{\theta\in\Theta} |K^{ij,(2)}_{n,\varepsilon}(\theta) - K^{ij}(\theta)| \xrightarrow{P_{\theta_0}} 0 \text{ as } \varepsilon \to 0, n \to \infty.$$

Therefore, the proof is complete. □

Proof of Theorem 4.22. The proof is very similar to that given in [45]. For the sake of completeness, we still include it here. We follow the arguments of [60] and [45]. Consider the closed ball

$$B(\theta_0; \rho) := \{\theta : |\theta - \theta_0| \leq \rho\}, \qquad \rho > 0.$$

Using the established consistency of estimator $\hat{\theta}_{n,\varepsilon}$ from Theorem 4.21, we can find a sequence $\eta_{n,\varepsilon} \to 0$ as $\varepsilon \to 0$ and $n \to \infty$ such that

$$B(\theta_0; \eta_{n,\varepsilon}) \subset \Theta \text{ and } P_{\theta_0}(\hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})) \to 1.$$

By application of Taylor's theorem when $\hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})$, we get

$$D_{n,\varepsilon} T_{n,\varepsilon} = \varepsilon^{-1}(G_{n,\varepsilon}(\hat{\theta}_{n,\varepsilon}) - G_{n,\varepsilon}(\theta_0)),$$

where $D_{n,\varepsilon} := \int_0^1 K_{n,\varepsilon}(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0))du$ and $T_{n,\varepsilon} := \varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0)$ as $B(\theta_0; \eta_{n,\varepsilon})$ is a convex subset of the parameter space Θ. We have

$$\begin{aligned} & |D_{n,\varepsilon} - K_{n,\varepsilon}(\theta_0)| 1_{\{\hat{\theta}_{n,\varepsilon} \in B(\theta_0;\eta_{n,\varepsilon})\}} \\ \leq & \sup_{\theta\in B(\theta_0;\eta_{n,\varepsilon})} |K_{n,\varepsilon}(\theta) - K_{n,\varepsilon}(\theta_0)| \\ \leq & \sup_{\theta\in B(\theta_0;\eta_{n,\varepsilon})} |K_{n,\varepsilon}(\theta) - K(\theta)| + \sup_{\theta\in B(\theta_0;\eta_{n,\varepsilon})} |K(\theta) - K(\theta_0)| + |K_{n,\varepsilon}(\theta_0) - K(\theta_0)|. \end{aligned}$$

Hence, by Lemma 4.35, we obtain that

$$D_{n,\varepsilon} \to K_{n,\varepsilon} \xrightarrow{P_{\theta_0}} 0 \text{ as } \varepsilon \to 0, n \to \infty.$$

Noting that $K(\theta)$ is continuous with respect to θ and $-K(\theta_0) = I(\theta_0)$ is positive definite; hence, there exists $\delta > 0$ such that

$$\inf_{|w|=1} |K(\theta_0)w| > 2\delta. \tag{4.13}$$

Choosing such a $\delta > 0$, there exist $\varepsilon(\delta) > 0$ and $N(\delta) > 0$ such that for all $\varepsilon \in (0, \varepsilon(\delta))$, $n > N(\delta)$ we have

$$B(\theta_0; \eta_{n,\varepsilon}) \subset \Theta \text{ and } |K(\theta) - K(\theta_0)| < \frac{\delta}{2} \text{ for } \theta \in B(\theta_0; \eta_{n,\varepsilon}).$$

Choose a $\delta > 0$ satisfying (4.13) and set

$$\Gamma_{n,\varepsilon} = \left\{ \sup_{|\theta-\theta_0| \le \eta_{n,\varepsilon}} |K_{n,\varepsilon}(\theta) - K(\theta_0)| < \frac{\delta}{2}; \hat{\theta}_{n,\varepsilon} \in B(\theta_0;\, \eta_{n,\varepsilon}) \right\}.$$

Now, for $\varepsilon \in (0, \varepsilon(\delta))$ and $n > N(\delta)$, on the set $\Gamma_{n,\varepsilon}$ we have

$$\begin{aligned}
& \sup_{|w|=1} |(D_{n,\varepsilon} - K(\theta_0))w| \\
\le\ & \sup_{|w|=1} \left| \left(D_{n,\varepsilon} - \int_0^1 K_{n,\varepsilon}(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0))du \right) w \right| \\
& + \sup_{|w|=1} \left| \left(\int_0^1 K_{n,\varepsilon}(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0))du - K(\theta_0) \right) w \right| \\
\le\ & \sup_{|\theta-\theta_0| \le \eta_{n,\varepsilon}} |K_{n,\varepsilon}(\theta) - K(\theta)| + \frac{\delta}{2} \\
<\ & \delta.
\end{aligned}$$

Namely, on the set $\Gamma_{n,\varepsilon}$ and the result of (4.13), we have

$$\inf_{|w|=1} |D_{n,\varepsilon} w| \ge \inf_{|w|=1} |K(\theta_0)w| - \inf_{|w|=1} |(D_{n,\varepsilon} - K(\theta_0))w| > \delta > 0.$$

Set $\mathfrak{D}_{n,\varepsilon} = \{D_{n,\varepsilon} \text{ is invertible}, \hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})\}$. Then, by Lemma 4.35, we get

$$P_{\theta_0}(\mathfrak{D}_{n,\varepsilon}) \ge P_{\theta_0}(\Gamma_{n,\varepsilon}) \to 1 \text{ as } \varepsilon \to 0, n \to \infty.$$

Set

$$U_{n,\varepsilon} = D_{n,\varepsilon} 1_{\mathfrak{D}_{n,\varepsilon}} + I_{p\times p} 1_{\mathfrak{D}^c_{n,\varepsilon}},$$

where $I_{p\times p}$ is the identity matrix. Now,

$$|U_{n,\varepsilon} - K(\theta_0)| \le |D_{n,\varepsilon} - K(\theta_0)| 1_{\mathfrak{D}_{n,\varepsilon}} + |I_{p\times p} - K(\theta_0)| 1_{\mathfrak{D}^c_{n,\varepsilon}} \xrightarrow{P_{\theta_0}} 0$$

since $P_{\theta_0}(\mathfrak{D}_{n,\varepsilon}) \to 1$. Thus, by Lemma 4.34, we obtain that

$$\begin{aligned} T_{n,\varepsilon} &= U_{n,\varepsilon}^{-1} D_{n,\varepsilon} T_{n,\varepsilon} 1_{\mathfrak{D}_{n,\varepsilon}} + T_{n,\varepsilon} 1_{\mathfrak{D}_{n,\varepsilon}^c} \\ &= U_{n,\varepsilon}^{-1}(-\varepsilon^{-1}(G_{n,\varepsilon}(\theta_0))) 1_{\mathfrak{D}_{n,\varepsilon}} + T_{n,\varepsilon} 1_{\mathfrak{D}_{n,\varepsilon}^c} \\ &\xrightarrow{P_{\theta_0}} I^{-1}(\theta_0) \left(\int_0^1 (\partial_{\theta_i} b)^*(r, S_r^0, \theta) \Bigg\{ \sigma(r, S_r^0) dB_r \right. \\ &\qquad \left. + \int_{\{|u| \le 1\}} H(r, S_r^0, u) \widetilde{N}(dr, du) + \int_{\{|u|>1\}} G(r, S_r^0, u) N(dr, du) \Bigg\} \right)^*_{1 \le i \le p} \end{aligned}$$

as $\varepsilon \to 0$ and $n \to \infty$. The proof is complete. □

4.5 Conclusion

The novelty of this chapter is:

- Estimating the unknown period of a time-dependent periodic transmission function as well as its unknown coefficients.
- The introduction and use of a Linear-Search Gradient-Descent algorithm to iteratively solve for suitable approximations to the LSEs.
- Parameter estimation results for when the noise coefficient (diffusion matrix) is a non-square or non-invertible matrix.
- Extending asymptotic results for consistency and rate of convergence given by Long et al. in [46] to include SDEs driven by general Lévy noises with time-dependent coefficients.

We aimed to study the ability to estimate periodic transmission parameters for a stochastic SIR model. The model comes in one of two forms, population proportions or population numbers, both of which we have given simulation studies on which we are able to estimate the period of a periodic transmission function effectively.

The theoretical results show that these estimations efforts hold in general. These results aid in generalizing existing efforts in the literature on parameter estimation for SDEs driven by Lévy noises with small coefficient ε. Moreover, despite our use case, the given results on asymptotics are not limited to the study of epidemiological models.

Chapter 5

Parameter estimation of stochastic SIR model with COVID-19 data from New York City

5.1 Introduction

The COVID-19 pandemic erupted in early 2020, with the disease spreading at an alarming rate towards the end of 2019 and the beginning of the new year. Several factors contributed to this, including widespread travel, insufficient testing in the early stages, and a lack of preventative measures to control the spread of the infection. Indeed, events such as COVID-19 will spur research using SDEs and parameter estimation to understand the event better. Some notable recent works are available in the literature (cf. [1, 2, 4, 10, 48, 33, 50, 57, 58, 63]).

Over three years since the pandemic began, we have gained valuable insights into its behaviour. With this knowledge, we are better equipped to face future pandemics caused by similarly infectious diseases. One of the defining features of the pandemic was the recurring waves of positive cases that surged and subsided as preventative measures were implemented or relaxed. Moreover, the abundance of infectious waves raises the question of how we can study these seemingly periodic occurrences and in fact, the study of the various waves occurring in the pandemic has been studied as well (cf. [2, 48, 50]). We also observed different variants of the disease, each with varying transmissibility levels and specific symptoms. Another notable driving force of the pandemic was that worldwide travel is commonplace; hence it is necessary to consider the importation and exportation of disease as another defining feature of the pandemic.

The classical SIR model can provide a basic understanding of pandemics, but it falls short when capturing the reality of COVID-19 as it does not contain the complexity possible in the USSIR. Our studies not only confirmed the existence of unique, strong solutions for USSIR but also delved into the long-term behaviours of the disease. However, as the COVID-19 pandemic is a recent event, it will take time to study the dynamics of disease persistence and extinction thoroughly. Hence, in this chapter, the work is an application of the theoretical results presented in Chapter 4 (cf. [22, E. and Sun]).

Many epidemiological models include birth and natural mortality rate; however, we omit these features here as this study was not concerned with a long period of time, and over a matter of a few years, birth and death rates remain relatively stable. Moreover, the population we consider is in the millions; hence, minor birth and natural death variations will not make a noticeable difference. Also noteworthy is that the distribution of vaccinations in the United States commenced in the latter part of 2020, approximately one year following the onset of the pandemic; moreover, it's important to remember that the vaccine's widespread administration to prevent infections was not an immediate process. Thus, in the studies herein, we do not yet consider the impact of the vaccines.

5.2 Model and methodology

As before, we make no assumptions about the explicit form of the noise, simply that it is a Lévy noise with a small dispersion rate given by ε. Consider the following small-noise proportional stochastic SIR model, which generalizes equation (1) given in [57] by N. Stollenwerk et al.

$$\begin{bmatrix} dX_t \\ dY_t \\ dZ_t \end{bmatrix} = \begin{bmatrix} -\beta(t)X_t(Y_t+\varrho) \\ \beta(t)X_t(Y_t+\varrho)-\gamma Y_t \\ \gamma Y_t \end{bmatrix} dt + \varepsilon\sigma(t,X_{t-},Y_{t-},Z_{t-}) \begin{bmatrix} dL_t^{(1)} \\ dL_t^{(2)} \\ dL_t^{(3)} \end{bmatrix}, \tag{5.1}$$

where in the drift portion $\beta(t)$ is a non-negative periodic transmission function, $\varrho \in \mathbb{R}$ is the import/export rate and $\gamma > 0$ is the removal rate. As for the noise portion, $\varepsilon \in (0,1)$ is the small dispersion rate and $\sigma(t, X_{t-}, Y_{t-}, Z_{t-})$ is an unknown time-dependant noise coefficient of the present Lévy noise $L_t = [L_t^{(1)}, L_t^{(2)}, L_t^{(3)}]$.

We consider two versions of $\beta(t)$:

$$\textbf{(V1)}\ \beta(t) = \alpha_0 + \alpha_1 \cos\left(\frac{2\pi}{\theta}t\right) + \alpha_2 \sin\left(\frac{2\pi}{\theta}t\right)$$

or

$$\textbf{(V2)}\ \beta(t) = \alpha_0 + \alpha_1 \cos\left(\frac{2\pi}{\theta}t\right) + \alpha_2 \sin\left(\frac{2\pi}{\theta}t\right) + \alpha_3 \cos\left(\frac{4\pi}{\theta}t\right) + \alpha_4 \sin\left(\frac{4\pi}{\theta}t\right),$$

with $\theta \in (0,1), \alpha_i > 0; i \in \{0,1,2,3,4\}$.

Stochastic SIR model (5.1) generalizes model (1) given in [57] by N. Stollenwerk et al. by the inclusion of small Lévy noise and a periodic transmission. Moreover, model (5.1) is an explicit form that falls under our work on the USSIR in [21]. Furthermore, we can ascertain the existence of a unique, strong solution by [21, Theorem 2.1, E. and Sun].

In order to complete our parameter estimation, we recall the contrast function given in Chapter 4 as

$$\Psi_{n,\varepsilon}(\vartheta) = n\sum_{k=1}^{n} P_k^*(\vartheta)P_k(\vartheta),$$

where

$$P_k(\vartheta) = \begin{bmatrix} X_{t_k}^{data} - X_{t_{k-1}}^{data} - \frac{1}{n}(-\beta(t_{k-1})X_{t_{k-1}}^{data}(Y_{t_{k-1}}^{data}+\varrho)) \\ Y_{t_k}^{data} - Y_{t_{k-1}}^{data} - \frac{1}{n}(\beta(t_{k-1})X_{t_{k-1}}^{data}(Y_{t_{k-1}}^{data}+\varrho) - \gamma Y_{t_{k-1}}^{data}) \\ Z_{t_k}^{data} - Z_{t_{k-1}}^{data} - \frac{1}{n}\gamma Y_{t_{k-1}}^{data} \end{bmatrix}$$

arising from the real-world data for $k \in \{1,\dots,n\}$, where n is the number of observations. Let $\hat{\vartheta}_{n,\varepsilon}$ be a minimum contrast estimator, i.e., a random variable satisfying

$$\hat{\vartheta}_{n,\varepsilon} := \arg\min_{\theta\in\Theta} \Psi_{n,\varepsilon}(\vartheta).$$

As we pointed out in Chapter 4, finding a closed form of $\hat{\vartheta}_{n,\varepsilon}$ is rather difficult; hence, in the same fashion as our previous work, we find suitable approximations $\hat{\vartheta}^*_{n,\varepsilon}$ of $\hat{\vartheta}_{n,\varepsilon}$. We do so by the same methodology in that work, using the LS-GD algorithm. In order to make use of the LS-GD algorithm, we need to set the test values of the unknown period θ. As before, we perform our estimation from the time $t = 0$ to $t = 1$ thus we chose 100 test values of θ, denoted by $\hat{\theta}$ such that $\hat{\theta} \sim \text{Uniform}(\frac{i-1}{100}, \frac{i}{100})$ (cf. [21, Algorithm 1, page 7, E. and Sun]).

The unknown period parameter θ and the import/export parameter ϱ are central to our interests here. Given our estimation from time $t = 0$ to 1 and the correspondence to the data being daily, each test value $\hat{\theta}$ taking some value $s \in (0, 1)$ then corresponds to a proportion (percentage) of time for the unknown transmission period; hence the found value is translatable to the number of days corresponding to the same proportion for the total observed days.

As the values for susceptible, infected and removed/recovered are represented proportionally, we interpret the import/export rate ϱ value to correspond to a proportional quantity of incoming or outgoing infectives. Moreover, if the value obtained is negative, then we take this to signal that the number of infected being exported is greater than the number imported and should that value be positive, then we take it to mean the inverse.

Upon finishing the estimation, we continue following our previous study regime, where we take the real-world data and compare it to the estimated outcomes using the returned parameter values. The real-world data is used to specify the three compartments: susceptible, infected and recovered/removed. This conversion gives us the data in a form where it is possible to compare with the estimated results of each compartment quickly.

5.3 Data for NYC

For our data analyses, we use data available for New York City, which was made publicly available at JHU CSSE COVID-19 Data (cf. [19]). More specifically, the data we utilize is the daily confirmed cases from which we can obtain the cumulative confirmed cases and the daily deaths which are COVID-19 related. Although this data is incomplete, as is often the case with real-world data, it is sufficient to accomplish our estimation efforts. In Figure 25 below, the confirmed cases for the first 650 days of the pandemic are displayed. The recommendation to stay-at-home began on 14 March 2020 in New York, and on 20 March 2020, the governor of New York implemented the PAUSE order mandating the first lockdown for the state–including New York City.

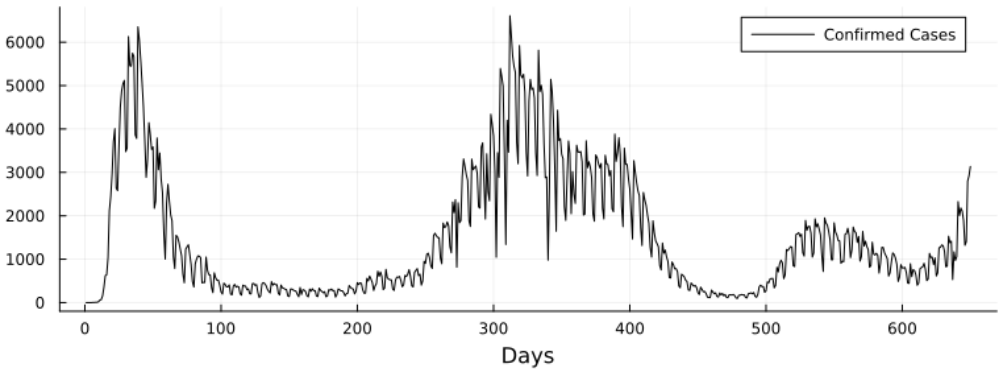

Figure 25: NYC confirmed cases for the first 650 days beginning 1 March 2020.

In the work we completed in [21, 22], we consider a model in which each compartment has a non-zero initial state, and this is true of our application here. The first cases of COVID-19 appeared in NYC at the start of March 2020, and by the lockdown mandate on 20 March 2020, there had already been both infections and deaths, thus giving us suitable data to study the model with non-zero initial values in each compartment: susceptible, infected and removed. This gives us a sample size of 630 days to work with for our estimation regime. NYC is an ideal candidate for our study as there is readily available data, and it is also a densely populated and well-travelled geographical location (i.e., the importance of the inclusion of import/export in the model (5.1)).

Vaccination in NYC began at the start of 2021; however, according to NYC Health vaccination data only 24.2% of the city had received at least the first dose of the vaccine by 20 March 2021 (cf. [49]). This is a full year (365) days after the day we utilized as the start of our studies; namely, there remained millions of susceptible people in NYC. Moreover, as seen in Figure 25, the second significant wave of confirmed cases peaked not much after 300 days since 1 March 2020. The importance of this information is that in our study, as mentioned, it took much time to ensure enough of the population was vaccinated to impact the spread of COVID-19; hence we are justified in the omission of vaccination in the work presented here. Moreover, it was not for another 7-14 days after the second dose that protection from infection was in effect, where the second dose would be administered four weeks after the first dose–this further elaborates the time needed for widespread vaccination to become effective in infection prevention and future work will include the consideration of vaccines.

5.4 Available data conversion to susceptible, infected and removed compartments

John Hopkins University (JHU) mortality analyses study of COVID-19 gave a case fatality between $0.1\% - 4.9\%$ depending on country with the United States having a rate at 1.1%. We use an average of 14 days for recovery from COVID-19 since many sources have estimated the recovery period to be between 7 and 21 days.

Since we make use of available data on confirmed cases and deaths, it makes the most sense to

begin to construct the infected compartment of the model. Along with the available data, we form the infected compartment for a given day n by taking the total cumulative amount of confirmed cases up to that day and removing 98.9% of the confirmed cases for the preceding 14 days prior to day n when $n > 14$ and also subtracting the number of deaths which occurred that day–the 98.9% originates from 1.1% case fatality rate given by JHU. The susceptible compartment is calculated by starting with a population of 8 million and subtracting the infected amount. Finally, the removed amount is calculated as the sum of the deaths and those removed from the infected compartment after the 14-day recovery period has passed.

Lastly, each compartment is normalized by the total population. Per the obtained results in Chapters 3 and 4, we find no detriment to our studies by choosing the proportional model form for this data analysis. Moreover, the proportional model has an advantage in that it quickly provides the percentage of the total population that each compartment contributes; moreover, percentages are more easily understood than raw numbers, especially given the population of NYC is in the multi-millions.

5.5 Results

Below we include plots, values of the estimated parameters and a metric of the estimation. The metric we use is the same as given in [22, Section 3.3, E. and Sun], that is, the component-wise MSE; the component-wise MSE allows us to measure our results for each component. To recall its form, it is given as

$$MSE_{(\mathbf{Vi})} = \frac{1}{630}\sum_{k=1}^{630}\left[(X_{t_k}^{data} - X_{t_k}^{estimated})^2, \quad (Y_{t_k}^{data} - Y_{t_k}^{estimated})^2, \quad (Z_{t_k}^{data} - Z_{t_k}^{estimated})^2\right], i = 1, 2.$$

The results for model (5.1) with **(V1)** and **(V2)** of $\beta(t)$ will appear on the next two pages, respectively.

Remark 5.36 *Recall the number of days we use is* 630*; hence, the appearance of* 630 *in the above sum.*

First, we begin with the results of the estimation for model (5.1) with **(V1)** of $\beta(t)$ as follows

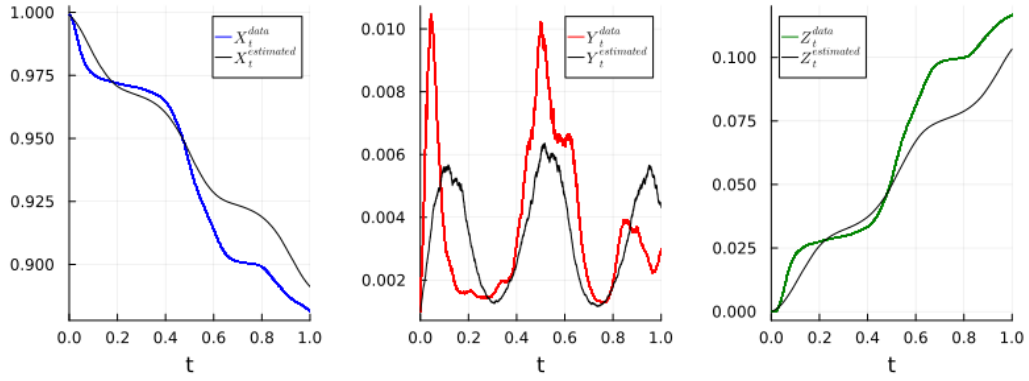

Figure 26: Data and estimation results compared using **(V1)** $\beta(t)$.

θ	0.4253264872306153
α_0	16.512145245907206
α_1	6.650423177688102
α_2	5.353434244651672
ϱ	0.0027282688418273957
γ	29.11209168217865

Table 7: Estimated parameters of model (5.1) for **(V1)** of $\beta(t)$.

These parameters yield

$$MSE_{(\mathbf{V1})} = \begin{bmatrix} 0.0016169192753345636 \\ 4.4186570235619574 \cdot 10^{-6} \\ 0.0016169192753345636 \end{bmatrix}.$$

We present here the results when utilizing **(V2)** of $\beta(t)$:

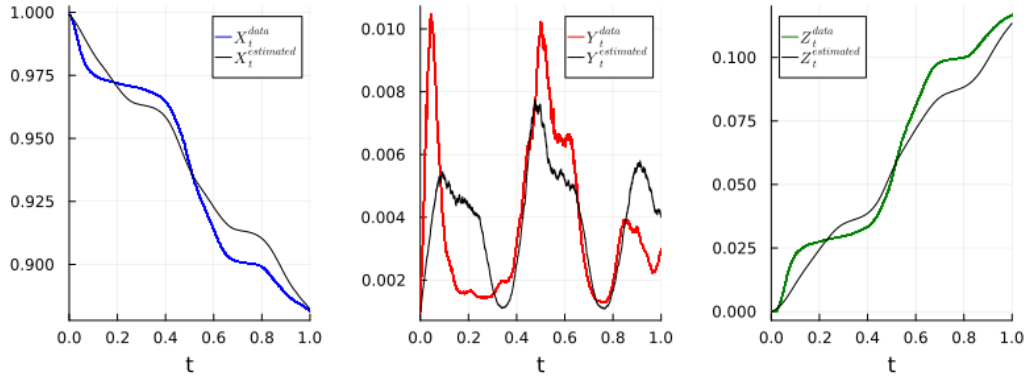

Figure 27: Data and estimation results compared using **(V2)** $\beta(t)$.

θ	0.4149391426488074
α_0	20.46707762062233
α_1	5.852151299221485
α_2	7.654681611402113
α_3	6.85271727216665
α_4	$3.542677117346328 \cdot 10^{-13}$
ϱ	0.0016373669036948662
γ	28.98752767851761

Table 8: Estimated parameters of model (5.1) for **(V2)** of $\beta(t)$.

These parameters yield

$$MSE_{(\mathbf{V2})} = \begin{bmatrix} 6.190004147478986 \cdot 10^{-5} \\ 4.837478808802567 \cdot 10^{-6} \\ 6.190004147478986 \cdot 10^{-5} \end{bmatrix}.$$

As we can see in the metrics for **(V1)** and **(V2)**, the accuracy with respect to the infected compartment of **(V1)** is marginally better than **(V2)**; however, given how close they are this is not of concern. Another finding of interest is that in the plots, we can see **(V2)** does yield more similarity to the path of the data than that of **(V1)**; regardless, at present, there seems little to

no benefit in increasing the number of $\alpha_i, i > 4$ in $\beta(t)$–this would only increase the time cost of computation.

Given the results above, periodicity is clearly a feature of the COVID-19 pandemic. Moreover, our findings yield a period in real-world days to be in the vicinity of 260 days. There is still much to learn about what forces impact this periodicity, but we are now better equipped to study and understand this phenomenon. Additionally, the inclusion of import and export is a logical topic of study, given the disease would not have grown to be a pandemic had it not been for the importation and exportation of the disease occurring due to travelling infected people. Both estimations presented here (i.e., the use of (**V1**) and (**V2**)) yielded a positive import/export rate ϱ which signals there was a net increase of infected coming into NYC-from a real-world perspective, this is reasonable as NYC is a very popular travel destination.

With the estimation results in hand, predictive analysis is performed on 315 days beyond the 630 days utilized in the estimation regime. As (**V1**) yielded the smaller metric (with respect to the infected compartment), we use the resultant parameters from (**V1**) to complete our prediction. Recall that time $t = 0$ to 1 corresponds to 630 days beginning on 20 March 2020; thus, in the following for time $t > 1$, the correspondence is with days 651 to 965. Namely, the following plot contains the simulation using the found parameters against the data for time $t = 0$ until $t = 1.5$ corresponding to 945 days commencing from 20 March 2020.

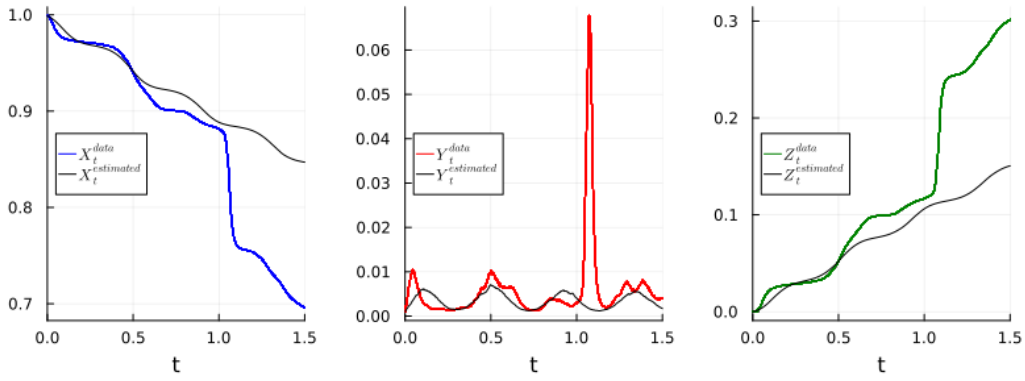

Figure 28: Data and predictive analysis simulation using **(V1)** $\beta(t)$.

Given this plot, there are two key characteristics to take away: the first is the significant spike in the infected compartment shortly after time $t = 1$; the second is that despite the significant spike in infected, the estimated infected $Y_t^{estimated}$ is reasonably close to the data infected Y_t^{data} after the significant spike, that is the section between approximately time $t = 1.15$ to 1.5. At initial thought, the significant spike and seemingly inaccuracy of the prediction for the time approximately $t = 1$ to 1.15 appear concerning, but if we revisit the data, we find an explanation. Namely, if we consider the number of confirmed cases and the number of COVID-19 tests over the first 965 days, we see a clear explanation–see the following plot.

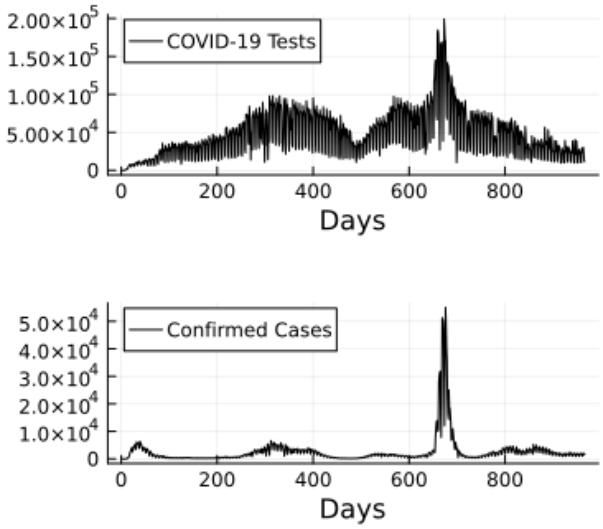

Figure 29: NYC daily tests and daily confirmed cases for first 965 days.

As can be seen in the plot, there is a sharp increase in both cases and COVID-19 tests simultaneously. This is no accident; these spikes began with the start of the holiday season in December 2021; namely, a time of year with an abundance of travel and at a time there was fear of another potential lockdown; hence many more tests were being administered, which indeed would correspond to an increase in confirmed cases. Recall that confirmed cases are the most crucial data we have for the modelling and estimation completed herein, and this signals the importance of data reporting at all times but especially during a pandemic. Moreover, it would be unwise to think that the amount of data available captures the actual number of infected; that is, we should expect the data to give an underestimation of the actual number of infected ultimately.

5.6 Closing remarks and data availability

These results are novel in the realm of studying epidemiological phenomena by use of SDEs. In the study and parameter estimation of a (stochastic) SIR model, the number of parameters is often limited to a transmission rate β and recovery rate γ; however, we established results interpretable in the real-world via a much more complex system than that of the classical formulation. Furthermore, the presence of time-dependency and periodicity contributed to a much more realistic scenario. Additionally, these results validate the theoretical results given in Chapter 4 (cf. [22]).

All data and sources used herein are cited and available for use by the public. The data from John Hopkins University is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) by John Hopkins University on behalf of its Centre for Systems Science in Engineering. ©John Hopkins University 2020.

Chapter 6

Discussion

This work stems from the novel results we accomplished across [21, 22, E. and Sun] and ultimately adds to the available literature for future studies of computational epidemiology. Here we give a discussion of these contributions and paths forward for those who wish to continue this work.

6.1 Contributions of thesis

The contributions of the thesis are:

- **Novel framework (the USSIR) generalizing the study of stochastic SIR models**
  This work establishes a framework to more readily handle the complexity of disease outbreaks as they happen in the real-world. For us, COVID-19 was the disease studied, but the results are not isolated to this particular disease. Moreover, the robust model includes two forms (proportional and population numbers), time-dependency, periodicity and driving noise that can be made rather general and complex. Results were provided to understand a disease's critical extinction and persistence dynamics without relying on an explicit form of coefficients determining these behaviours.

- **Parameter estimation of a time-dependent stochastic SIR model with periodic transmission**
  Periodic behaviour is a natural occurrence in the world; thus, questions about the existence of periodicity in disease dynamics are a logical inquiry. As seen with COVID-19, there have been waves, and where there are waves, there is periodicity.

- **Contrast function $\Psi_{n,\varepsilon}(\theta)$ for obtaining least-squares estimators for time-dependent models without explicit knowledge of noise**
  Many models require time-dependency; for us, that necessity was the notion of periodicity of transmission. The presence of unavoidable noise is a fact regarding data collection. Fortunately, Lévy noises are general enough that it is reasonable to assume the noise may be modelled by a Lévy process but beyond this, we cannot always ensure we can precisely give the explicit form of the process. Our contrast function $\Psi_{n,\varepsilon}(\theta)$ and the consistency and rate of convergence results show that we can still obtain relevant results when estimating parameters without abundant information on the noise–including time-dependent models.

- **The linear-search gradient-descent algorithm**
  Given an estimation regime, obtaining estimators in closed-form is not always a convenient task; furthermore, it simply may not be possible, as is often the case in theory and practice, it suffices to find suitable approximations, specifically for us, approximations to LSE estimators. Estimation is no easy task; if there is interaction among parameters, the task becomes even more difficult. Recall the definition of our periodic transmission function $\beta(t)$,

we have

$$\alpha_0 + \sum_{k=1}^{K} \alpha_{k,1} \cos\left(\frac{2\pi}{\theta} kt\right) + \alpha_{k,2} \sin\left(\frac{2\pi}{\theta} kt\right)$$

and so it is easy to see that given the multiplication of an unknown scalar parameter and a trigonometric function with an unknown period parameter, inter-play occurs between unknown parameters. Thus, the power of LS-GD was the ability to iteratively circumvent this problem by fixing an unknown period and then carrying out the estimation of the scalar factor by established methods.

- **Application to real-world COVID-19 data**
  The work would have been incomplete without some validation of the theoretical work herein. In a world where data collection is at an all-time high, applying the theoretical results we have achieved to some real data was a reasonable task. Namely, the data for New York City is plentiful; however, data collection is far from perfect. The use of a periodic transmission function was validated and provided insight into the different waves of COVID-19. The importation and exportation of COVID-19 is a reality, one which did not require the work here to ascertain, but nonetheless, our work shows it is possible to quantify this feature, thus expanding our ability to understand pandemics.

6.2 Future work

Work of this sort answers some questions but ultimately opens up new questions. Some potential follow-ups we wish to explore are:

- **Continuing studies of the USSIR**
  The USSIR model is very general, but this can be made more general. Namely, since our work focused on strong solutions to the coupled SDEs it merits an investigation into weak solutions under weaker conditions placed upon the coefficient functions.

  Another direction worth pursuing is, if we consider the COVID-19 pandemic, some events impacted disease spread, such as the widespread use of lockdowns or the development of vaccinations. For such scenarios, it could be advantageous to have a model which considers these (time-dependent) events. An investigation into a piece-wise USSIR model would allow for the inclusion of time-dependent events, which can significantly impact the form of the model. Namely, a model in which there are no features for lockdown or vaccination prior to some time T but after this time, such features are present (e.g., the widespread administration of vaccines is an event that can shape the dynamics of disease spread, thus making sense to include two different transmission functions triggered by some time T, say $\beta_1(t), t < T$ and $\beta_2(t), T \geq T$, which are pre-vaccine and post-vaccine, respectively). This is especially important when considering estimation, as the parameter values can differ between the absence of vaccinations and the abundance of vaccinated (immune) people). Additionally, using real-world data, we acknowledge that the reported quantities are not the true values of the quantities we wish to study. Hence, in the future, we will consider a filtering approach to data analysis with real-world data to account for the disparities

between the observed quantities and the true quantities. To further investigate the gaps between observed and true states, we wish to include the introduction of hidden states in subsequent studies and applications of the USSIR model–particularly when considering real-world data.

Another area of interest is the inclusion of random coefficients in the SDEs defining the USSIR. Higa et al. [34] have previously considered and obtained results on the existence and uniqueness of solutions to Stratonovich SDEs with random coefficients. The use of Lévy noise to drive our SDEs dramatically increases the realism of our model; that is, this captures the environmental fluctuations. Now consider the scenario in which the transmission or recovery from the infection is modelable by a random process; namely, they have their own "noise" disjointed from the environmental noises explained by the driving Lévy processes. Gourieroux and Lu [28] considered a remarkable deterministic SIR model with functional transmission; they then extended it to a stochastic transmission; in other words, they considered a random differential equation SIR model. Diseases and how they spread are indeed very complex, and to understand them better, we should consider all the possible ways in which noise (randomness) impacts the disease-spread dynamics.

- **Optimization of LS-GD**
  The LS-GD was immensely useful in our work but was also computationally costly; hence, optimizing the algorithm would significantly improve its applicability. The fastest way to optimize it would be to reduce the time of the linear-search portion, that is, the part of the algorithm utilized for fixing the unknown period parameter in our periodic transmission.

We wish to thank Professor Michael Kouritzin for invaluable discussion, insights and suggestions regarding future work we aim to pursue.

References

[1] Alenezi, M.N., Al-Anzi, F. S. and Alabdulrazzaq, H. (2021). Building a sensible SIR estimation model for COVID-19 outspread in Kuwait. Alexandria Engineering Journal 60 3161–75.

[2] AlQadi, H. and Bani-Yaghoub, M. (2022). Incorporating global dynamics to improve the accuracy of disease models: Example of a COVID-19 SIR model. Y. E. Khudyakov, ed PLOS ONE 17.

[3] Applebaum, D. (2009). Lévy Processes and Stochastic Calculus. 2nd ed. Cambridge University Press.

[4] Bagal, D. K., Rath, A., Barua, A. and Patnaik, D. (2020). Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods. Chaos, Solitons & Fractals 140 110154

[5] Bailey, N. (1975). The Mathematical Theory of Infectious Diseases and Its Applications, 2nd ed. London: Griffin.

[6] Bao, J. and Yuan, C. (2012). Stochastic population dynamics driven by Lévy noise. Journal of Mathematical Analysis and Applications 391 363–75.

[7] Beck, A. (2017). First-Order Methods in Optimization. SIAM-Society for Industrial and Applied Mathematics.

[8] Bezanson, J., Edelman, A., Karpinski, S. and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. SIAM Review 59 65–98.

[9] Bishwal, J.P.N. (2008). Parameter Estimation in Stochastic Differential Equations. Lecture Notes in Mathematics. Springer.

[10] Bodini, A., Pasquali, S., Pievatolo, A. and Ruggeri, F. (2021). Underdetection in a stochastic SIR model for the analysis of the COVID-19 Italian epidemic. Stochastic Environmental Research and Risk Assessment 36 137–55.

[11] Brauer, F., van den Driessche, P., Wu, J. (2008). Mathematical Epidemiology, 1st ed. Mathematical Biosciences Subseries 1945. Springer.

[12] Brauer, F., Castillo-Chavez, C. and Feng, Z. (2019). Mathematical Models in Epidemiology. Vol 69 Springer.

[13] Caraballo, T., El Fatini, M., El Khalifi, M. and Rathinasamy, A. (2021). Analysis of a stochastic coronavirus (COVID-19) Lévy jump model with protective measures. Stochastic Analysis and Applications 41 45–59.

[14] Chen, G., Li, T. and Liu, C. (2013). Lyapunov exponent of a stochastic SIRS model. Comptes Rendus Mathematique 351 33–5.

[15] Chowdhury, S., Roychowdhury, S. and Chaudhuri, I. (2021). Universality and herd immunity threshold: Revisiting the SIR model for COVID-19. International Journal of Modern Physics C 32 2150128.

[16] Clémençon, S., Chi Tran, V. and de Arazoza, H. (2008). A stochastic SIR model with contact-tracing: large population limits and statistical inference. Journal of Biological Dynamics 2 392–414.

[17] Clum, C., Mixon, D. (2020). Parameter estimation in the SIR model from early infections. arXiv:2008.04286.

[18] Dalal, N., Greenhalgh, D. and Mao, X. (2007). A stochastic model of AIDS and condom use. Journal of Mathematical Analysis and Applications 325 36–53.

[19] Dong, E., Du, H. and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 20 533–4.

[20] Dorogovtsev, A. J. (1976). The consistency of an estimate of a parameter of a stochastic differential equation. Theory of Probability and Mathematical Statistics 10 73–82.

[21] Easlick, T., Sun, W. (2022). A unified stochastic SIR model driven by Lévy noise with time-dependency. arXiv:2201.03406v2.

[22] Easlick, T., Sun, W. (2023). Parameter estimation of stochastic SIR model driven by small Lévy noises with time-dependent periodic transmission. arXiv:2303.07983v1.

[23] El Koufi, A., Adnani, J., Bennar, A. and Yousfi, N. (2019). Analysis of a stochastic SIR model with vaccination and nonlinear incidence rate. International Journal of Differential Equations 2019 1–9.

[24] EL Koufi, A., Adnani, J., Bennar, A. and Yousfi, N. (2021). Dynamics of a stochastic SIR epidemic model driven by Lévy jumps with saturated incidence rate and saturated treatment function. Stochastic Analysis and Applications 40 1048–66.

[25] Ethier, S.N., Kurtz, T.G. (1986). Markov Processes: Characterization and Convergence. John Wiley and Sons.

[26] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. Statistics 21 99–116.

[27] Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. Stochastic Processes and their Applications 119 679–99.

[28] Gourieroux, C., Lu, Y. (2020). SIR model with stochastic transmission. arXiv:2011.07816.

[29] Gray, A., Greenhalgh, D., Hu, L., Mao, X. and Pan, J. (2011). A stochastic differential equation SIS epidemic model. SIAM Journal on Applied Mathematics 71 876–902.

[30] Greenhalgh, S. and Day, T. (2017). Time-varying and state-dependent recovery rates in epidemiological models. Infectious Disease Modelling 2 419–30.

[31] Guo, X.-X. and Sun, W. (2021). Periodic solutions of stochastic differential equations driven by Lévy noises. Journal of Nonlinear Science 31:32.

[32] Hager, W. W. and Zhang, H. (2006). Algorithm 851 : CG-DESCENT, a conjugate gradient method with guaranteed descent. ACM Transactions on Mathematical Software 32 113–37.

[33] Hespanha, J. P., Chinchilla, R., Costa, R. R., Erdal, M. K. and Yang, G. (2021). Forecasting COVID-19 cases based on a parameter-varying stochastic SIR model. Annual Reviews in Control 51 460–76.

[34] Kohatsu-Higa, A., León, J. A., Nualart, D. and Leon, J. A. (1997). Stochastic differential equations with random coefficients. Bernoulli 3 233.

[35] Ji, C., Jiang, D. and Shi, N. (2011). Multigroup SIR epidemic model with stochastic perturbation. Physica A: Statistical Mechanics and its Applications 390 1747–62.

[36] Ji, C. and Jiang, D. (2014). Threshold behaviour of a stochastic SIR model. Applied Mathematical Modelling 38 5067–79.

[37] Kasonga, R. A. (1988). The consistency of a non-linear least squares estimator from diffusion processes. Stochastic Processes and their Applications 30 263–75.

[38] Kermack, W., McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 115 700–21.

[39] Kobayashi, M. and Shimizu, Y. (2022). Least-squares estimators based on the Adams method for stochastic differential equations with small Lévy noise. Japanese Journal of Statistics and Data Science 5 217–40.

[40] Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. The Annals of Statistics 18:1158–1171.

[41] Le Breton, A. (1976). On continuous and discrete sampling for parameter estimation in diffusion type processes. Stochastic Systems Vol. 1 124-144.

[42] Liptser, R.S., Shiryayev, A.N. (1986). Theory of Martingales. Kluwer Academic Publishers.

[43] Liu, Q., Jiang, D., Shi, N. and Hayat, T. (2018). Dynamics of a stochastic delayed SIR epidemic model with vaccination and double diseases driven by Lévy jumps. Physica A: Statistical Mechanics and its Applications 492 2010–2018.

[44] Liu, Y., Zhang, Y. and Wang, Q. (2020). A stochastic SIR epidemic model with Lévy jump and media coverage. Advances in Difference Equations 2020.

[45] Long, H., Shimizu, Y. and Sun, W. (2013). Least squares estimators for discretely observed stochastic processes driven by small Lévy noises. Journal of Multivariate Analysis 116 422–39.

[46] Long, H., Ma, C. and Shimizu, Y. (2017). Least squares estimators for stochastic differential equations driven by small Lévy noises. Stochastic Processes and their Applications 127 1475–95.

[47] Mogensen, P.K., Riseth, A.N. (2018). Optim: A mathematical optimization package for Julia. Journal of Open Source Software vol. 3-24 615.

[48] Muñoz-Fernández, G. A., Seoane, J. M. and Seoane-Sepúlveda, J. B. (2021). A SIR-type model describing the successive waves of COVID-19. Chaos, Solitons & Fractals 144 110682.

[49] NYC Health. Covid-19: Data. In: COVID-19: Latest Data - NYC Health. https://www.nyc.gov/site/doh/covid/covid-19-data.page. Accessed 1 April 2023

[50] Perakis, G., Singhvi, D., Skali Lami, O. and Thayaparan, L. (2022). COVID-19: A multiwave SIR-based model for learning waves. Production and Operations Management 32 1471–89.

[51] Privault, N. and Wang, L. (2021). Stochastic SIR Lévy jump model with heavy-tailed increments. Journal of Nonlinear Science 31:15.

[52] Protter, P. (1990). Stochastic Integration and Differential Equations: A New Approach. Springer-Verlag.

[53] Rackauckas, C. and Nie, Q. (2017). DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. Journal of Open Research Software vol. 5-1 15.

[54] Schlickeiser, R. and Kröger, M. (2021). Analytical modelling of the temporal evolution of epidemics outbreaks accounting for vaccinations. Physics 3 386–426.

[55] Shimizu, Y. (2009). Quadratic type contrast functions for discretely observed Non-Ergodic diffusion processes. Osaka University, Mathematical Science Research Report No. 09-04.

[56] Sørensen, M. and Uchida, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations. Bernoulli 9 1051-1069.

[57] Stollenwerk, N., Bidaurrazaga Van-Dierdonck, J., Mar, J., Eguiguren Arrizabalaga, I., Cusimano, N., Knopoff, D., Anam, V., Aguiar, M. (2021). The interplay between subcritical fluctuations and import: understanding COVID-19 epidemiology dynamics medRxiv 2020.12.25.20248840.

[58] Tesfay, A., Saeed, T., Zeb, A., Tesfay, D., Khalaf, A. and Brannan, J. (2021). Dynamics of a stochastic COVID-19 epidemic model with jump-diffusion. Advances in Difference Equations 2021:228.

[59] Tornatore, E., Maria Buccellato, S. and Vetro, P. (2005). Stability of a stochastic SIR system. Physica A: Statistical Mechanics and its Applications 354 111–26.

[60] Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. Scandinavian Journal of Statistics 31 553–66.

[61] van der Vaart, A. W. (1998). Asymptotic Statistics. Vol Series Number 3. Cambridge University Press.

[62] Xu, P. and Shimada, S. (2000). Least squares parameter estimation in multiplicative noise models. Communications in Statistics - Simulation and Computation 29 83–96.

[63] Zhang, Z., Zeb, A., Hussain, S. and Alzahrani, E. (2020). Dynamics of COVID-19 mathematical model with stochastic perturbation. Advances in Difference Equations 2020:451.

[64] Zhou, Y. and Zhang, W. (2016). Threshold of a stochastic SIR epidemic model with Lévy jumps. Physica A: Statistical Mechanics and its Applications 446 204–16.